

APANG

AmericaPang



아팡팀 (Retention → Revenue 발표자료)

<https://colab.research.google.com/drive/1MM13NF91Z2hjDqGRfCK03TF1PRDMgiJw#scrollTo=VkECC2RlotZI>

🛒 Instacart AARRR & Retention → Revenue 분석
▼ 팀원 및 데이터셋 소개



팀원

- 윤준성
- 박정우
- 고명희

📌 정식 명칭

Instacart Online Grocery Shopping Dataset 2017

📌 성격

- 온라인 식료품 쇼핑 서비스의 실제 주문 이력 데이터
- 고객이 어떤 상품을, 어떤 조합으로, 언제, 얼마나 자주 구매했는지 기록
- 마켓 바스켓 분석, 리텐션 분석, 추천 시스템 연습에 최적화된 데이터셋



Instacart 마켓 바스켓 데이터셋은 **고객 × 주문 × 상품 × 시간**이 모두 연결된
실제 온라인 장보기 행동 데이터

▼ 1. 분석 배경 & 핵심 질문(가설)

- Retention은 어떤 경로를 통해 Revenue를 증가시키는가?
- 이 분석의 핵심은 재구매가 실제 매출 증가로 어떻게 이어지는지 구조적으로 보는 것에 있음
- 예를 들어 재구매가 높아지면, 고객은 구매 타이밍 안정화, 주문 1회당 장바구니 크기 증가, 고객 성숙도 증가 등으로
자신의 행동이 변화할 가능성이 있어 어떤 경로를 통해 매출을 증가시키는지 보고자 하는 것에 목적이 있음

가설

- 재구매 ↑ → 구매 주기 안정화
- 재구매 ↑ → 주문 1회당 장바구니 크기 ↑
- 고객 성숙도 ↑ → Revenue ↑

▼ 2. 데이터 개요

- 데이터: Instacart 공개 데이터
- 분석 범위: Train set
- 분석 단위
 - 고객 단위 (user-level)
 - 주문 단위 (order-level)

📌 본 분석은 '구매까지 도달한 사용자 집합'을 대상으로 함

📋 기준 컬럼

컬럼명	의미
order_id	주문 ID (각 주문을 구분하는 고유 값)
product_id	상품 ID (각 상품을 구분하는 고유 값)
add_to_cart_order	해당 주문에서 장바구니에 담긴 순서 (1부터 시작)
reordered	재구매 여부• 1 = 이전에 구매한 적 있는 상품• 0 = 처음 구매한 상품
user_id	사용자(고객) ID

컬럼명	의미
eval_set	데이터셋 구분용 컬럼• prior / train / test
order_number	해당 사용자의 누적 주문 횟수 (첫 주문 = 1)
order_dow	주문한 요일 (day of week)보통 0~6 (0=일요일, 6=토요일)
order_hour_of_day	주문한 시간 (0~23시)
days_since_prior_order	이전 주문 이후 경과 일수• 첫 주문은 보통 NULL
product_name	상품 이름
aisle_id	진열대(aisle) ID
department_id	상품이 속한 부서(department) ID
aisle	진열대 이름 (예: fresh fruits)
department	부서 이름 (예: produce)

주문 단위 파생 컬럼

컬럼명	의미
basket_size	해당 주문의 전체 상품 수
unique_products	주문에 포함된 서로 다른 상품 개수
unique_aisles	주문에 포함된 서로 다른 aisle 개수
unique_departments	주문에 포함된 서로 다른 department 개수
reordered_items	주문 내 재구매 상품 개수
reorder_rate	재구매 비율 (<code>reordered_items / basket_size</code>)

집계 테이블용 컬럼

시간대별 / 요일별 / 단계별 컬럼

컬럼명	의미
orders	집계 기준에 해당하는 전체 주문 수
users	집계 기준에 해당하는 고유 사용자 수
avg_basket_size	평균 장바구니 크기
avg_reorder_rate	평균 재구매율

주문 단계 기준

컬럼명	의미
order_stage	주문 단계 (예: 신규/중기/충성 고객 구간 등)

부서 기준 집계

컬럼명	의미
items	해당 부서에서 판매된 전체 상품 수
orders	해당 부서 상품이 포함된 주문 수

컬럼명	의미
avg_reordered_rate	해당 부서 상품의 평균 재구매율

▼ 3. AARRR 프레임워크 정의

- **Acquisition(유입):** 전체 고유 사용자 수와 신규 고객 수(첫 주문을 수행한 고객)
- **Activation(활동):** 신규 고객 중에서 실제로 첫 구매를 완료한 비율(첫 구매를 완료한 고객 수를 신규 고객 수로 나눈 비율)
 - 👉 AARRR에서 Activation은 보통 회원가입, 앱 설치, 첫 클릭 등을 의미함
 - 👉 그러나 본 분석의 목적은 실질적인 비즈니스 가치 창출 지점이므로,
Activation을 첫 구매 완료로 정의하였음
 - 👉 “처음 온 사람들 중, 구매 경험까지 제대로 한 비율은 얼마인가?”
- **Retention(유지):** 한 번 산 사람이 다시 돌아오는가?
 - order_stage(주문단계)
 - order_hour_of_day(주문한 시간대)
 - order_dow(요일)
 - gap_bucket(이전 주문 이후 경과일 수 구간) 기준으로 분석
 - 👉 “언제 주문한 고객이, 얼마 만에 다시 주문하는가?”
 - 👉 “시간대/요일/주문 단계에 따라 재구매 성향이 달라지는가?”
- **Revenue(매출):** ‘평균 장바구니 크기(average basket size)’를 대체 지표로 사용
 - order_stage
 - order_hour_of_day
 - order_dow
 - gap_bucket 기준으로 시각화
 - 👉 “어떤 조건에서 고객은 더 많이 담아 구매하는가?”



- **Referral(추천) 단계는 분석에서 제외**

추천(Referral)은 보통

- 초대 코드
- 추천 링크
- 공유 기반 전파 구조

같은 **명확한 추천 메커니즘**이 있어야 분석 가능함

👉 본 데이터에서는 추천 경로를 식별할 수 있는 정보가 부족하였고, 추천이 핵심 성장 동인이 아니라 고 판단하여, **AARRR의 Referral 단계는 분석에서 제외**

▼ 4. Acquisition & Activation 결과

- 전체 고유 사용자 수: **206,209명**
- 신규 고객 수(첫 주문): **206,209명**
- Activation Rate: **100%**

Number of activated customers (first purchase): **206209**

Activation Rate: **100.00%**

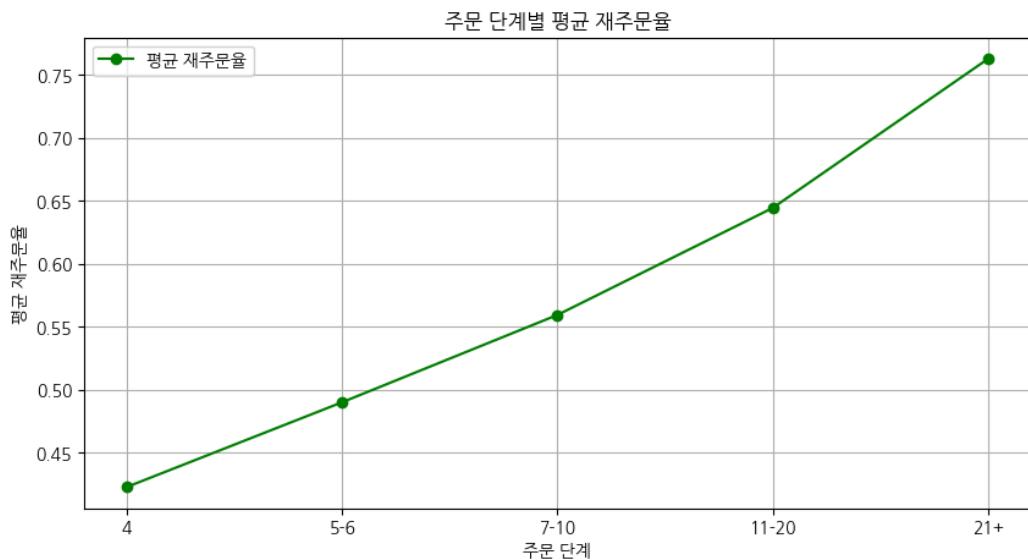
해석

- 첫 구매 진입 장벽 매우 낮음
- Activation은 '첫 구매 완료'로 정의되었으며, 신규 고객 또한 첫 구매를 수행한 고객으로 정의되었기 때문에 Activation Rate는 구조적으로 100%로 계산된다.
- Activation Rate 자체보다는 Activation 이후의 Retention과 Revenue 차이를 보는 것이 목적
- Activation은 필터링 단계 역할을 수행

▼ 5. Retention 분석 – 고객 성숙도

- 주문 횟수가 증가할수록 재구매율 상승
 - 초기 고객(≤ 4 회): **42%**
 - 성숙 고객(21회+): **76%**

📌 Retention은 단발 이벤트가 아니라 누적되는 고객 성숙도의 결과





시사점

- 재구매는 단발적 행동이 아니라 주문 경험 누적에 의해 강화됨
- 초기 고객(4회 이하)을 중기 단계로 전환시키는 전략이 장기 리텐션의 핵심

“고객 성숙도가 높아질수록 재구매 확률이 구조적으로 상승하므로, 초기 고객을 빠르게 중·고단계 주문 경험으로 전환시키는 것이 리텐션 성과를 좌우한다.”

▼ 6. Retention 분석 – 구매 주기

- 3~14일 구매 간격에서 재구매율 최고
- 주간·격주 단위의 습관적 구매 패턴 확인

☞ 재구매의 핵심은 빈도가 아니라 리듬

구매 주기	재주문율	평균 장바구니 크기	해석 포인트
1~2일	0.62	7.53(최소)	즉시 보충성 주문
3~7일	0.68 (최대)	10.87	주간 구매 리듬
8~14일	0.65	11.28 (최대)	격주 계획 구매
15~30일	0.54(최소)	10.49	이탈 위험 구간



시사점

☞ Retention 관점

- 재구매는 무작위 빈도가 아니라 3~14일의 안정적인 구매 리듬에서 가장 잘 발생

☞ Revenue 관점

- 8~14일 구매 주기 고객이 주문 1회당 가치 최대
- Retention이 Revenue로 전환되는 핵심 구간

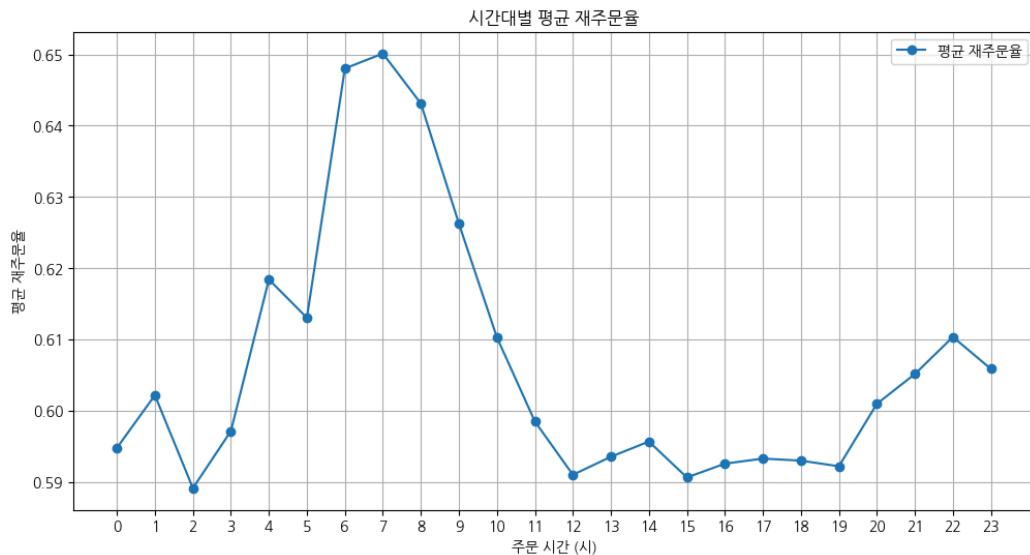
▼ 7. Retention 분석 – 시간대 & 요일

📊 주요 관찰 결과

① 시간대별 평균 재주문율 분석

- 가장 높은 재주문율
 - 06~08시 (특히 7시, 약 0.65)
- 가장 낮은 재주문율
 - 12~15시 구간 (약 0.59 전후)

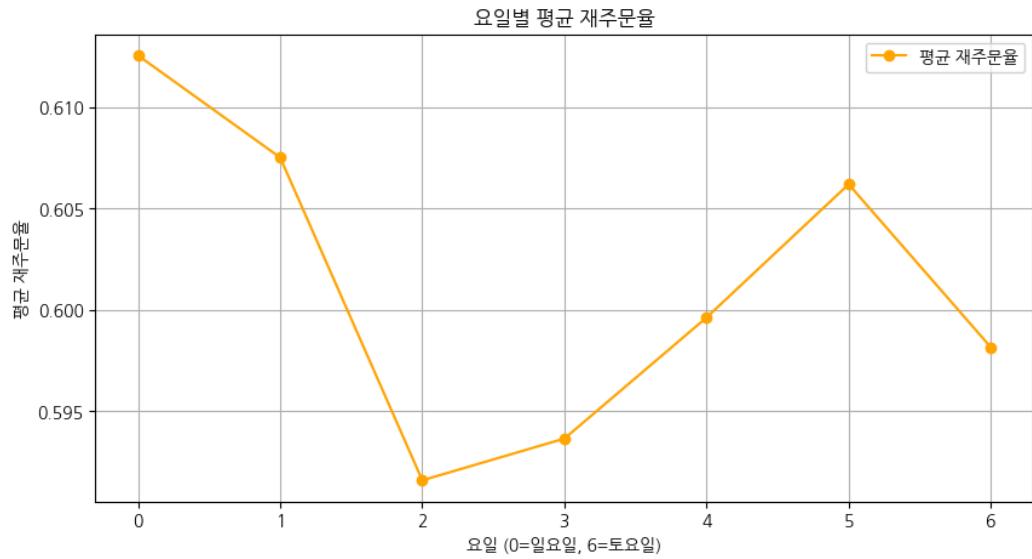
- 저녁 이후(20~23시)
 - 재주문율이 다시 완만하게 상승 (약 0.60~0.61)



② 요일별 평균 재주문율 분석

📊 주요 관찰 결과

- 재주문율이 가장 높은曜일
 - 일요일(0)
 - 금요일(5)
- 재주문율이 가장 낮은曜일
 - 화요일(2)
- 주중 종반(수·목)은 점진적 회복



시사점

① Retention 캠페인 최적 요일

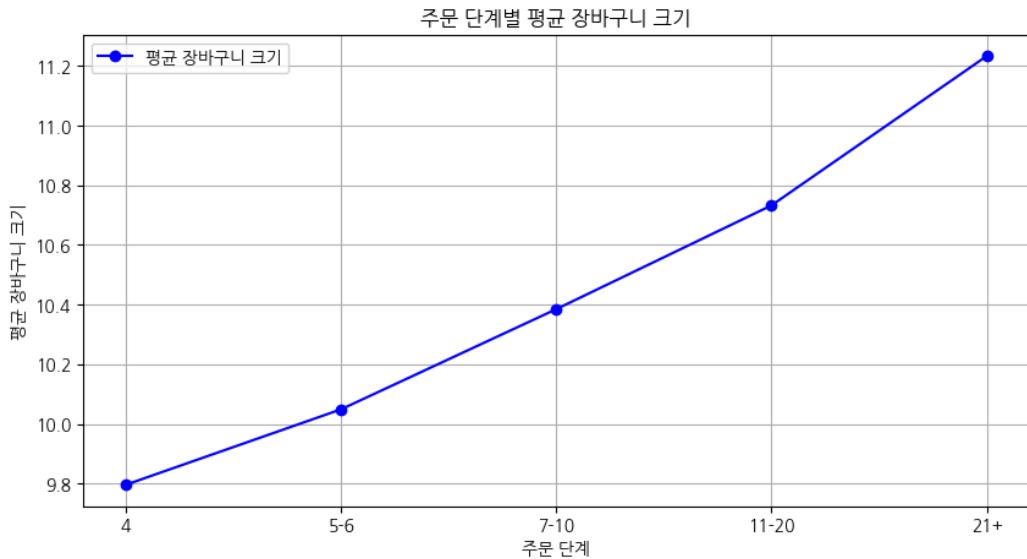
- 일요일, 금요일
- 화요일은 재구매 유도 전략이 필요한 취약 구간
 - 할인, 리마인드 푸시, 개인화 추천 적합

▼ 8. Revenue 분석 – 고객 성숙도별 구매 규모 변화

주문 단계별 평균 장바구니 크기 – 시사점

📊 핵심 관찰

- 주문 단계가 증가할수록 **평균 장바구니 크기**가 일관되게 증가
 - 초기(4회): **약 9.8개**
 - 중기(7-10회): **약 10.4개**
 - 성숙(21회 이상): **약 11.2개**
- 단발적 구매가 아닌, 경험 축적에 따른 구매 규모 확대가 확인됨💡 주요 시사점



시사점

① 재구매 고객은 '더 자주'가 아니라 '더 크게' 구매한다

- 고객 성숙도 상승은
 - 주문 빈도 증가뿐 아니라
 - 주문 1건당 가치(장바구니 크기) 증가로 연결
- Retention은 단순 유지 지표가 아니라 매출 증대 지표

② 초기 고객을 빠르게 '중기 단계'로 이동시키는 것이 중요

- 5~6회 → 7~10회 구간에서 장바구니 크기 상승이 뚜렷
- 이 구간 진입 여부가
 - 고객 LTV 분기점 역할

전략적 의미

“첫 구매 이후 몇 번 더 사게 만드느냐가 매출을 좌우”

③ 신규 고객 할인보다 성숙 고객 전환 시 Retention 투자 효율이 높음

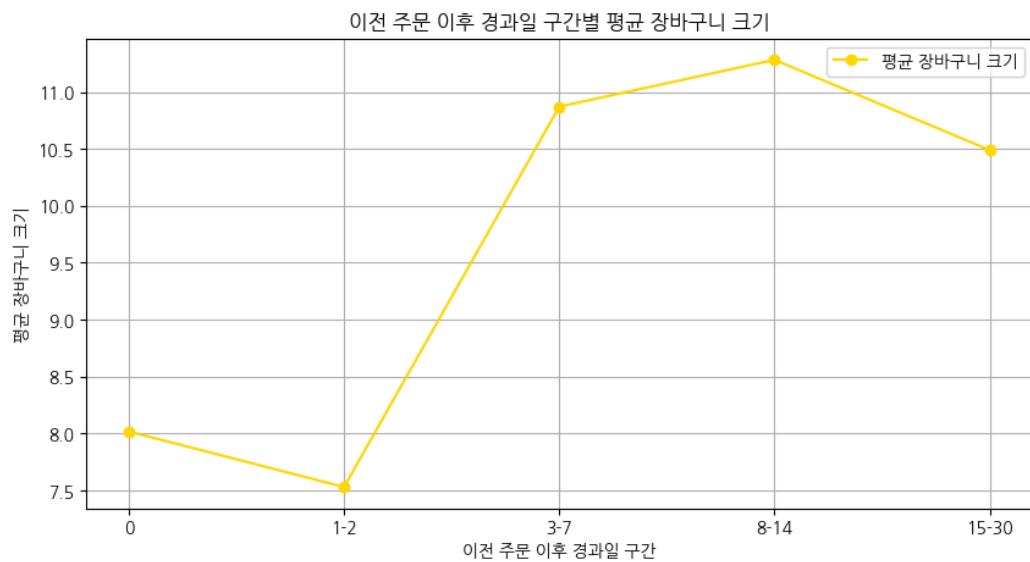
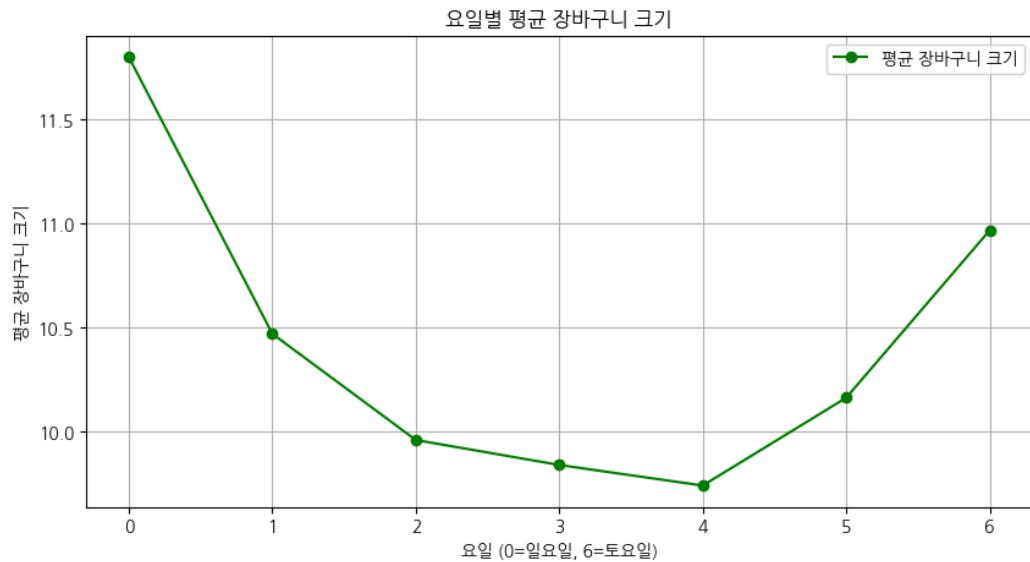
- 신규 → 성숙 고객 전환 시
 - 평균 장바구니 크기 약 +14~15% 증가
- CAC(신규 유입 비용) 대비
 - Retention 투자 ROI가 더 높을 가능성

“고객이 성숙해질수록 주문은 반복될 뿐 아니라, 한 번의 주문 자체가 더 커진다.”

▼ 9. Revenue 분석 – 구매 주기 & 요일

- 8~14일 구매 주기에서 장바구니 최대
- 주말 주문에서 장바구니 크기 최고

👉 Retention = 매출 레버리지





시사점

- 계획적인 재구매 주기(8~14일)를 가진 고객이 주문 1회당 구매 규모가 가장 커 매출 기여도가 가장 높음
- 주말 주문은 시간적 여유와 계획 소비가 결합되어 장바구니 확장(묶음·대용량 구매)이 가장 잘 발생
- 따라서 매출 증대는 신규 유입 확대보다 구매 리듬을 유지·강화하는 **Retention 전략**에서 더 효과적으로 발생

▼ 10. 구조적 결론(Why → How → So What)

① Retention → 고객 성숙도 상승

- 재구매가 누적될수록 고객은 주문 경험을 축적하며 성숙 고객으로 전환
- 주문 단계별 분석 결과,
 - 초기 고객(<4회)의 재구매율은 **42%**
 - 성숙 고객(21회+)의 재구매율은 **76%**
- 이는 Retention이 단발 이벤트가 아닌 누적 효과임을 의미

👉 초기 고객을 중·고단계 주문 경험으로 빠르게 전환하는 것이 리텐션 성과의 핵심

② Retention → 구매 주기 안정화 (리듬 형성)

- 재구매율은 무작위 빈도보다 3~14일의 안정적인 구매 주기에서 가장 높게 나타남
- 특히,
 - 3~7일: 주간 구매 리듬
 - 8~14일: 격주 계획 구매
- 구매 리듬이 형성된 고객일수록 이탈 위험이 낮고, 재구매가 지속됨

👉 재구매의 핵심은 '자주'가 아니라 '일정하게'

③ 고객 성숙도·구매 리듬 → 주문 1회당 가치 증가

- 고객이 성숙해질수록 평균 장바구니 크기 증가
 - 초기 고객: 약 **10.0개**
 - 성숙 고객: 약 **11.3개**
- 구매 주기 측면에서도,
 - 8~14일 구매 주기 고객이 장바구니 크기 최대
- 주말 주문 역시 장바구니 확장이 가장 크게 발생

👉 Retention은 주문 빈도뿐 아니라 주문 1건의 질(Revenue)을 함께 키움



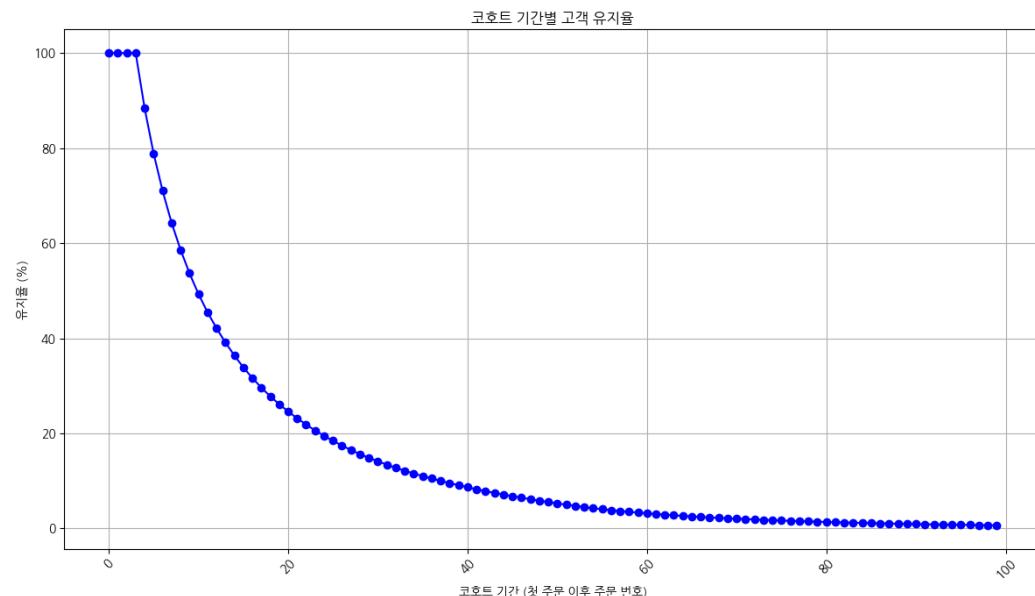
시사점

본 분석에서 리텐션은 단순히 다시 구매했는지를 보는 지표로써가 아니라, 고객의 구매 행동이 **일정한 주기 (3~14일)**와 **성숙한 주문 단계**로 발전하는 과정이다. 이러한 리텐션이 형성된 고객은 재구매율뿐 아니라 주문 1회당 장바구니 크기도 증가하며, 결과적으로 리텐션은 반복 구매를 넘어 **주문 단위 매출 가치를 키우는 구조적 성장 메커니즘**으로 작동하였다.

▼ 11. 코호트 [추가 분석] – 언제 이탈하는가?

- 각 사용자별 첫 주문 번호를 식별하기 위해 orders 데이터프레임을 user_id 기준으로 그룹화한 뒤, 각 그룹에서 order_number의 최솟값 계산
- 각 user_id와 해당 사용자의 첫 주문 번호(first_order_number)를 포함하는 데이터프레임을 생성

① 코호트 기간별 고객 유지를



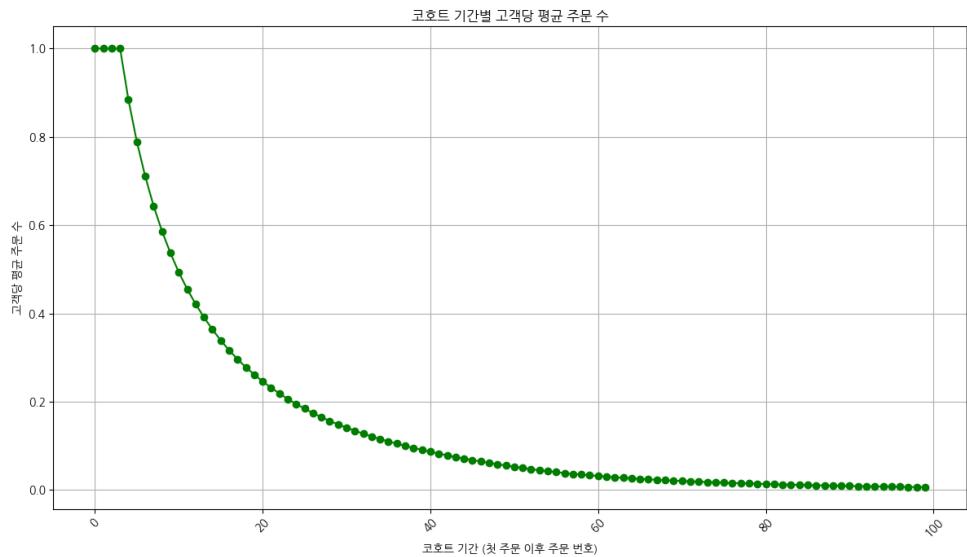
관찰

- 첫 주문 이후 **3~4회 주문까지는 높은 유지를 유지**
- **4회 주문 이후부터 유지를 하락 가속**
- 장기적으로는 유지를 완만한 꼬리 형태로 감소

해석

- 초기 몇 번의 주문은 비교적 자연스럽게 발생
- 그러나 **4번째 주문을 넘기지 못한 고객은 이탈 가능성 급증**
- 즉, 고객은 “몇 번 더 써볼까”의 단계 이후에 **계속 쓸지/떠날지를 결정**

시각화 (2) 코호트 기간별 고객당 평균 주문 수



관찰

- cohort_period 증가에 따라 고객당 평균 주문 수 지속 감소
- 일부 충성 고객은 반복 주문을 유지
- 다수 고객은 일정 시점 이후 주문 중단

해석(두 그래프를 합쳐서 보면 나오는 논리)

관찰	의미
유지율 감소	활동 고객 수가 줄어듦
평균 주문 수 감소	활동 고객 중 일부만 계속 주문
긴 꼬리	소수의 고빈도 고객 존재

- 평균값 감소는
 - 고객 전체가 덜 주문해서가 아니라
 - 주문을 멈춘 고객이 누적되기 때문
- 장기적으로는 소수의 충성 고객만 남아 평균을 지탱
- 평균 감소의 원인은 모든 고객의 행동 변화가 아니라, 주문을 중단한 고객의 누적 효과



시사점

1 이탈은 초기에 집중된다

- 첫 3~4회 주문이 생존 여부를 가르는 분기점
- 이 구간을 넘긴 고객은 장기 고객으로 전환될 확률 급증

2 리텐션은 '평균'이 아니라 '분포'의 문제

- 평균 유지율은 완만히 감소
- 하지만 실제로는
 - 다수 고객은 초기에 이탈
 - 소수 고객만 장기 생존

3 초기 재구매 경험이 장기 성과를 결정

- 첫 주문 이후
 - 빠른 재주문 경험**
 - 구매 성공 경험 누적
- 이 여부가 고객 수명을 좌우

⚠ 코호트 분석의 한계

- 집단 평균은 파악 가능
- 하지만
 - 개별 고객의 최대 이용 기간
 - 고객 수명 분포는 직접적으로 설명하지 못함

👉 이 한계를 보완하기 위해 **고객별 최대 주문 주기(수명)** 분석을 추가 수행

▼ 12. 리텐션 [추가 분석] – 얼마나 오래 남는가?

- 고객이 주문 횟수를 기준으로 서비스를 얼마나 오랫동안 사용했는지 이해하기 위해, 각 user_id별 cohort_period의 최댓값을 계산
- orders 데이터프레임을 user_id로 그룹화한 다음 cohort_period 열의 최댓값을 취하는 방식으로 진행

📊 핵심 지표 요약

- 전체 고객 수: **206,209명**
- 평균 고객 수명: **15.6회 주문**
- 중앙값(median): **9회 주문**
- 하위 25%: **5회 주문 이내 이탈**

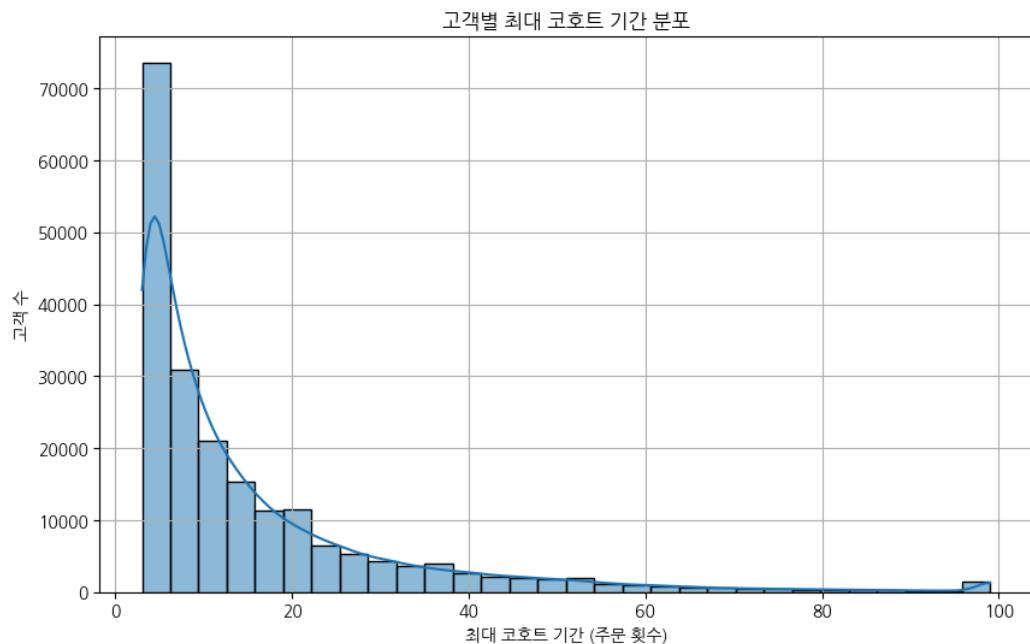
- 상위 25%: **19회** 이상 주문
- 최대: **99회** 주문 (극소수 고충성 고객)

```

count      206,209.00
mean       15.59
std        16.65
min        3.00
25%        5.00
50%        9.00
75%       19.00
max       99.00
Name: max_cohort_period, dtype: float64

```

시각화 해석 (고객 수명 분포)



관찰

- 분포는 오른쪽 꼬리가 긴 형태(right-skewed)
- 고객 수는 주문 횟수가 적을수록 급격히 많음
- 5~10회 주문 구간에 고객이 가장 집중
- 이후 주문 횟수가 증가할수록 고객 수 급감

해석

- 다수의 고객은 비교적 이른 시점에 이탈
- 소수의 고객만이 장기적으로 서비스를 이용
- 평균 수명(15.6회)은 소수 고충성 고객에 의해 끌어올려진 값

시사점

초기 10회 주문 이전이 리텐션의 골든 타임

- 고객의 절반은 10회 주문 이전에 이탈
- 하위 25%는 6회 이전에 이탈
- 초기 주문 경험이 장기 고객 전환을 결정

소수의 장기 고객이 전체 성과를 견인

- 극소수 고객이 20회 이상, 일부는 100회 가까이 주문
- 이들이
 - 장기 리텐션
 - 높은 재구매율
 - 큰 장바구니

를 통해 전체 매출에 큰 기여 가능성

▼ 13. 고객 세그먼트 -3 세그먼트 구조

고객 세그먼트 3가지 구조

1. 신규 고객 (New Customers)

주문 1~3회

특징

- 첫 구매 완료 → 진입 장벽 낮음
- 재구매율 낮음 ($\approx 40\%$ 대)
- 구매 주기 불안정
- 장바구니 크기 작음
- 이탈 위험 가장 높은 구간

행동 해석

- 서비스 탐색 단계

- 필요할 때만 사용
- 아직 “여기서 계속 사야겠다”는 판단 전

핵심 인사이트

| 신규 고객은 아직 고객이 아니라 체험자에 가깝다.(이탈 위험 최고)

◆ 2. 습관 형성 고객 (Habit-forming Customers)

주문 4~10회

특징

- 재구매율 뚜렷한 상승 구간
- 3~14일 구매 주기 형성
- 장바구니 크기 증가 시작 (≈ 10.5 개 내외)
- 고객 수 가장 많이 분포
- 장기 고객 전환의 골든 타임

행동 해석

- 주간·격주 장보기 루틴 형성
- 구매가 “선택”에서 “습관”으로 전환
- 다음 주문을 예상하고 계획함

핵심 인사이트

| 리텐션이 행동 패턴으로 고착되는 단계다. (전환 골든 타임)

◆ 3. 충성 고객 (Loyal Customers)

주문 11회 이상

특징

- 재구매율 최고 수준 ($\approx 70\%$ 이상)
- 구매 주기 매우 안정
- 평균 장바구니 크기 최대 (≈ 11.2 개+)
- 고객 비중은 작지만 매출 기여도 큼

행동 해석

- 서비스에 대한 신뢰 확립
- 묶음·대용량 구매 수용
- 가격 민감도 낮음

핵심 인사이트

| 충성 고객은 매출 구조를 떠받치는 핵심 자산이다. (매출 핵심 자산)

▼ 14. 세그먼트별 이벤트 추진 전략

이벤트의 목적은 할인을 주는 것이 아니라, 고객을 다음 단계로 이동시키는 것이다.

세그먼트별 이벤트 전략 한눈에 요약

세그먼트	핵심 목표	이벤트 핵심
신규 고객	첫 재구매 유도	3번째 주문 미션
습관 형성 고객	구매 리듬 고정	7·14일 리마인드
충성 고객	이탈 방지·매출 극대화	VIP·장바구니 확장

◆ 1. 신규 고객 이벤트 전략

(주문 1~3회 | 이탈 방지 구간)

대상

- 첫 구매를 완료한 고객
- 주문 1~3회 고객
- 최근 주문 이후 재주문 이력이 없는 고객

기획 의도

- 초기 이탈 방지
- 첫 재구매 경험을 빠르게 만들어 구매 진입 장벽 제거
- 고객을 '체험자'에서 '사용자'로 전환

이벤트 내용

- “3번째 주문 달성 미션”
 - 3번째 주문 완료 시 할인 쿠폰 / 무료 배송 제공
- 첫 주문 상품 기반 개인화 추천
- 첫 주문 후 3~7일 이내 리마인드 푸시
- 소액 혜택 중심 (진입 부담 최소화)

◆ 2. 습관 형성 고객 이벤트 전략

(주문 4~10회 | 전환 핵심 구간)

대상

- 주문 4~10회 고객

- 최근 구매 주기 3~14일 고객
 - 반복 주문 이력이 있는 고객
-

기획 의도

- 구매 리듬 고정 (주간·격주)
 - 구매를 '선택'이 아닌 '습관'으로 전환
 - 장기 고객(충성 고객)으로의 전환 가속
-

이벤트 내용

- “7일 · 14일 장보기 리마인드”
 - 지난 주문 기반 자동 추천 장바구니
 - 세트/묶음 상품 제안
 - 일정 주기 내 재주문 시 소액 보상(적립금, 쿠폰)
 - 주말 중심 노출 (장바구니 확장 타이밍)
-

◆ 3. 충성 고객 이벤트 전략

(주문 11회 이상 | 매출 핵심 구간)

대상

- 주문 11회 이상 고객
 - 평균 장바구니 크기 상위 고객
 - 구매 주기 안정 고객
-

기획 의도

- 이탈 방지 및 감정적 락인(lock-in)
 - 소수 고가치 고객의 장기 유지
 - 매출 구조 안정화
-

이벤트 내용

- VIP 등급 프로그램
 - 상시 무료 배송
 - 전용 할인
- 장바구니 확장 인센티브
 - n개 이상 구매 시 추가 할인
- 정기 구매/추천 큐레이션
- 혜택보다는 ‘대우받는 경험’ 중심

Keep

Problem

Try