BDAI 7팀 · 데이터 분석 보고서

# 장바구니 앞에서 멈춘 고객을 찾다

## 행동 로그 1,958만 건 기반 구매 여정 병목 진단과 단계별 마케팅 처방

| | |
|---|---|
| 분석 대상 | D2C 뷰티 이커머스 행동 로그 |
| 분석 기간 | 2019-10-01 ~ 2020-02-29 (5개월, UTC) |
| 데이터 규모 | 19,583,742행 · 활동 고객 1,639,358명 |
| 작성일 | 2026-10-19 |



| 팀원 | 역할 |
|---|---|
| 소혜 | 팀장 · 기획/발표 |
| 마로 | 분석 |
| 병철 | 데이터 엔지니어링 · 전처리/EDA |
| 보경 | 시각화 · Tableau |

# 목차


## Chapter.1 분석 목적 & 배경

## Chapter.2 활용 데이터

2-1. 데이터 개요와 정제
2-2. 분석 기준

## Chapter.3 분석 과정

3-1. 분석 절차와 방법
3-2. 지표 체계

## Chapter.4 분석 결과

4-1. 같은 상품 퍼널
4-2. 단계별 도달·전환·이탈 표
4-3. 가격대 · 첫 방문
4-4. 행동 패턴 · 브랜드 · 시간대
4-5. 장바구니 미구매 고객
4-6. 구매 시점 · 놓친 금액
4-7. 재구매 · 코호트
4-8. 리텐션 매트릭스
4-9. 고가치 고객
4-10. 고객 유형 · 매출 집중
4-11. 상품 ABC 등급
4-12. RFM × ABC 교차
4-13. 행동 × 미구매 × 고객 가치
4-14. 세그먼트 · 미구매 4유형 × 고객 특성
4-15. 세그먼트 · 방치형 세부
4-16. 세그먼트 · RFM 7유형 프로필
4-17. 세그먼트 · RFM 7유형 × 미구매 4유형

## Chapter.5 해석과 마케팅 제언

5-1. 가설 검증
5-2. 분석 질문에 대한 답
5-3. 전략 1 · 돌아오게 하고 구매를 돕는다
5-4. 전략 2 · 네 그룹별 다음 행동
5-5. 기대 효과 · 실행 순서 · 검증

## Chapter.6 한계점 및 결론



**요약** — 방문 고객의 6.74%만 구매합니다. 가장 큰 병목은 상품을 보고 담는 단계(같은 상품 조회→담기 11.10%)이고, 담은 고객의 77.65%(29.8만 명)는 7일 안에 사지 않았으며 그중 72.6%는 장바구니를 그대로 둔 방치형입니다. 산 고객의 12.93%만 30일 안에 다시 삽니다. 단계별로 **장바구니 리마인드 · 두 번째 구매 혜택 · 고가 상품 담기 장벽 완화**를 우선 제안합니다.


팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

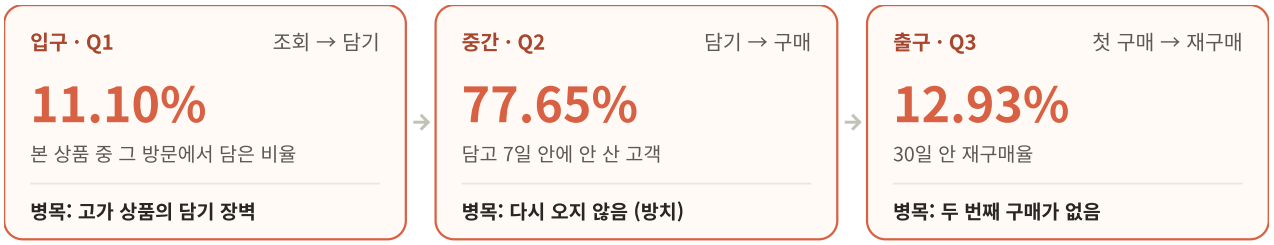
# Chapter.1 분석 목적 & 배경

## “분석 배경”

뷰티 이커머스 고객은 상품을 여러 번 비교하고 장바구니에 담아 둔 뒤 나중에 사는 경우가 많습니다. 이 쇼핑몰은 5개월 동안 164만 명이 방문했지만 구매한 고객은 6.74%에 그쳤습니다. 어느 단계에서, 어떤 고객이 빠지는지 모르면 마케팅 예산을 모든 고객에게 똑같이 쓰게 됩니다.

**[그림 1-1] 구매 여정의 세 병목**

입구는 같은 방문·같은 상품 기준, 중간·출구는 고객 기준




## “분석 목적”

구매 여정(조회 → 담기 → 구매 → 재구매)의 병목과 놓친 고객·금액을 찾아, **단계별로 다른 마케팅 처방과 우선순위**를 제안합니다. 이 보고서는 마케팅 리더가 다음 달 리소스를 어느 단계·어느 고객군에 먼저 쓸지 결정하는 데 쓰입니다.

## “분석 방향”

| 질문 | 분석 질문 | 구매 여정 |
|---|---|---|
| **Q1** | 구매 전환과 이탈 행동은 고객의 행동 패턴·시간대·가격대에 따라 어떻게 다른가? | 담기 전 |
| **Q2** | 구매 의도가 높지만 구매하지 않은 고객은 누구이며, 규모는 얼마인가? | 담은 후 |
| **Q3** | 지속적인 관리 가치가 높은 고객의 행동 특성은 무엇이며, 유형별로 어떤 전략이 효과적일까? | 산 후 |

| 가설 | 내용 |
|---|---|
| ① | View→Cart와 Cart→Purchase 전환율에 차이가 있으며, 이탈은 View→Cart 단계에서 가장 크고 가격이 높을수록 커질 것이다 |
| ② | 장바구니 도달 이후의 행동 유형에 따라 최종 구매전환율과 구매까지 소요시간에 차이가 있을 것이다 |
| ③ | 상위 N%의 고가치 고객은 첫 구매 후 재구매까지 걸리는 기간이 일반 고객보다 짧을 것이다 |

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

# Chapter.2 활용 데이터

## 2-1. "데이터 개요와 정제"

Kaggle에 공개된 화장품 쇼핑몰 행동 로그를 팀에서 정제해 사용했습니다. 조회·담기·삭제·구매 네 가지 이벤트가 시간 순서대로 기록되어 있어, 고객 한 명의 구매 여정을 따라갈 수 있습니다.

| 항목 | 내용 |
|---|---|
| 데이터 | Kaggle "eCommerce events history in cosmetics shop" 팀 정제본 · 19,583,742행 · 고객 1,639,358명 · 상품 54,571개 · 세션 4,536,272개 |
| 이벤트 | view 9,656,790 · cart 5,652,991 · remove_from_cart 2,987,859 · purchase 1,286,102 |
| 전처리 | 완전 중복 제거 · category_code 삭제(결측 98%) · brand 결측은 unknown(42.2%) · 가격 0·음수, 세션 결측은 행 유지 후 필요한 분석에서만 제외 · 시간 UTC |
| 분석 질문 | Q1 전환·이탈과 행동·시간·가격 · Q2 안 산 의도 고객과 규모 · Q3 관리 가치 고객과 전략 · (추가) Q4 행동·미구매·가치 교차 |

## 2-2. "분석 기준"

| 구분 | 기준 |
|---|---|
| **Q1 단위·계산** | 같은 고객·같은 세션·같은 상품 조합에서 조회 < 담기 < 구매(시간 순서, 같은 시각 불인정). 다른 세션은 연결하지 않음 |
| **F4 장바구니 이탈률** | 고객×상품, 담은 뒤 7일 안 같은 상품 미구매 → 77.33% (Q2 쪽 지표) |
| **가격대** | 경계: 상품 대표가격(양수 가격 중앙값)을 상품 수로 3등분, 저가 ≤ 3.32 < 중가 ≤ 6.35 < 고가. 배정: Q1은 첫 조회 가격, Q2는 첫 담기 가격 |
| **Q2 의도 고객** | 장바구니에 1번 이상 담은 고객. 첫 담기 후 7일 안에 아무것도 안 사면 미구매. 첫 담기 2/23 이후 제외 |
| **I3 · I4** | I3 = 담은 고객 ÷ 전체(24.30%). I4 = 미구매 고객이 7일 안 담은 상품의 담은 시점 가격 합(952만), 보조로 남은 상품 707만 · 상위 1% 제외 795만 |
| **Q3 주문·재구매** | 주문 = 고객의 구매일(UTC). 재구매율 = 첫 구매 후 30일 안(첫 구매 1/30 이전 고객). 5개월 값은 참고 |
| **고가치·세분화** | 가설 ③ 검증 = 첫 주문 금액 상위 20%. 관리 유형 = RFM 5점 7유형. 2×2(방문 4회 × 금액 상위 15%)는 Q4용 |
| **검증** | 같은 기간 비교 + 시간 분리 검증(이전 달까지 이력 → 다음 달 구매) 병행. 핵심 비교는 통계 검정 |

> **데이터의 한계** — 주문 번호·수량·통화 정보가 없어 주문은 고객의 구매일로, 금액은 가격 × 1개로 봅니다. 카테고리 이름이 없고(결측 98%) 2019년 10월 이전 기록이 없어 10월 코호트에는 기존 고객이 섞일 수 있습니다.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

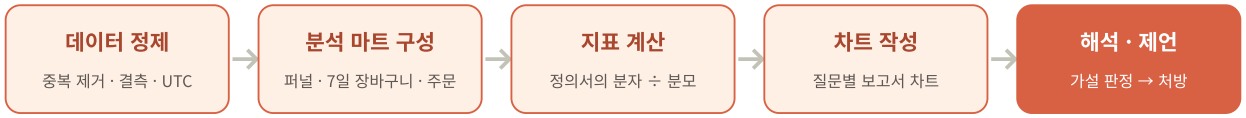
# Chapter.3 분석 과정

## 3-1. “분석 절차와 방법”

네 명이 나눠 분석한 결과를 같은 기준으로 다시 계산해 합쳤습니다. 질문마다 판정 단위를 다르게 두고(Q1 고객×세션×상품, Q2·Q3 고객), 핵심 비교는 통계 검정으로 확인했습니다.

**[그림 3-1] 분석 절차**

팀 정제 데이터 → 질문별 분석 마트 → 지표 → 차트 → 해석




| 질문 | 분석 방법 | 시각화 |
|---|---|---|
| **Q1** | 같은 방문·같은 상품 순차 퍼널, 가격대·브랜드·카테고리·조회 횟수 분해, 카테고리 내 비교 | 퍼널 막대, 그룹 막대, 요일×시간 하이라이트 테이블 |
| **Q2** | 첫 담기 후 7일 판정, 장바구니 4유형, 구매까지 시간, 놓친 금액(상위 1% 제외 범위) | 유형별 막대, 시간 분포, 가격대 누적 막대 |
| **Q3** | 첫 구매 코호트, 생존 분석(카플란–마이어 · 로그순위 검정), RFM 7유형, 파레토 | 코호트 히트맵, 생존 곡선, 비중 비교 막대 |
| **검증** | 비율 차이 신뢰구간 · $\chi^2$ 검정, 시간 분리 검증, 순서 조건 민감도 | 통계 확인 표 |

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 3-2. “K1 ≈ 담기율 × 담은 고객의 구매율, 그리고 다시 사는 고객”

**6.74%**
K1 사용자 구매전환율
월 6.45%(10월) ~ 8.56%(11월)

**24.30%**
IM1 담기율
1%p 오르면 K1 약 +0.28%p

**27.52%**
담은 고객의 기간 내 구매율
구매 고객의 99.2%가 담기 경험

**12.93%**
V1 30일 재구매율
다음 달 구매 고객을 만듦

북극성 지표는 세 질문이 모두 모이는 “더 많은 고객이 사게 하려면?”을 재는 **사용자 구매전환율(K1)**입니다. 한 고객이 여러 번 방문해도 한 명으로 셉니다. K1은 담기율(Q1이 움직임)과 담은 고객의 구매율(Q2가 움직임)의 곱이고, 재구매(Q3)는 다음 달 K1의 바탕이 됩니다. 비율 지표라 할인만으로 오를 수 있어 매출과 주문당 금액을 가드레일로 함께 봅니다.

| ID | 지표 | 기준 | 값 (5개월) |
|---|---|---|---|
| **K1** | 사용자 구매전환율 (북극성) | 구매 고객 ÷ 활동 고객 | **6.74%** |
| **IM1 · I3** | 담기율 | 담은 고객 ÷ 활동 고객 | **24.30%** |
| **IM2** | 담은 고객의 7일 구매율 | 7일 안 구매 ÷ 의도 고객 | **22.35%** |
| **F1 · F2 · F3** | 같은 상품 조회→담기 · 담기→구매 · 순차 전환 | 같은 방문·같은 상품, 시간 순서 | **11.10% · 15.31% · 1.70%** |
| **F5 · F6** | 담기 후 삭제율 · 반복 조회(3회+) 구매율 | 같은 방문·같은 상품 | **19.11% · 6.75%** |
| **B2** | 같은 방문 미구매 → 7일 안 다른 방문 구매 | 같은 상품 | **16.92%** |
| **F4** | 장바구니 7일 이탈률 (Q2 쪽) | 고객×상품 7일 | **77.33%** |
| **I1 · I2** | 의도 고객 · 미구매율 | 첫 담기 후 7일 | **384,380명 · 77.65%** |
| **I4** | 놓친 금액 | 담은 시점 가격, 상품당 1번 | **952만 (707만 · 795만)** |
| **V1** | 30일 재구매율 | 구매일 주문 | **12.93%** |
| **V3** | 코호트 재구매율 (다음 달) | 11월 이후 코호트 | **8.5~9.8%** |
| **V4** | 고객당 30일 결제 | 첫 구매 후 30일 | **46.82** |
| **G1 · G2** | 매출 · 주문당 금액 | 구매일 주문 | **6,348,268 · 41.13** |

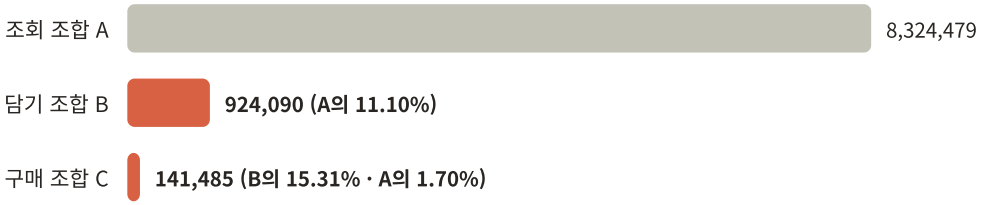
팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

# Chapter.4 분석 결과

## 4-1. “본 상품의 11.10%만 담기고, 담은 상품의 15.31%가 그 방문에서 팔립니다”

**[그림 4-1] 같은 방문 · 같은 상품 순차 퍼널 (5개월)**

고객×세션×상품 조합 · 조회 < 담기 < 구매 시간 순서 · 다른 세션은 연결하지 않음




| 월 | K1 | F1 담기 | F2 구매 | F3 순차 | 매출 |
|---|---|---|---|---|---|
| 10월 | 6.45% | 13.56% | 12.56% | 1.70% | 1,210,922 |
| 11월 | 8.56% | 11.19% | 16.85% | 1.89% | 1,530,831 |
| 12월 | 6.92% | 10.09% | 16.85% | 1.70% | 1,077,689 |
| 1월 | 6.88% | 10.48% | 15.93% | 1.67% | 1,321,825 |
| 2월 | 6.59% | 10.20% | 15.00% | 1.53% | 1,207,001 |

월 = 조합의 첫 조회 월(F1~F3), 구매일(매출). 2월은 데이터 끝이라 세션이 잘릴 수 있음.

| 계산 방식 | 조회→담기 | 담기→구매 |
|---|---|---|
| **같은 방문·같은 상품 (F1·F2)** | **11.10%** | **15.31%** |
| 같은 상품, 순서 무관 (민감도) | 14.99% | 14.84% |

같은 상품을 따라가므로 A를 보고 B를 산 경우는 전환으로 세지 않습니다. 순서 무관 값은 시간 순서 조건의 영향을 보여 주는 민감도입니다.

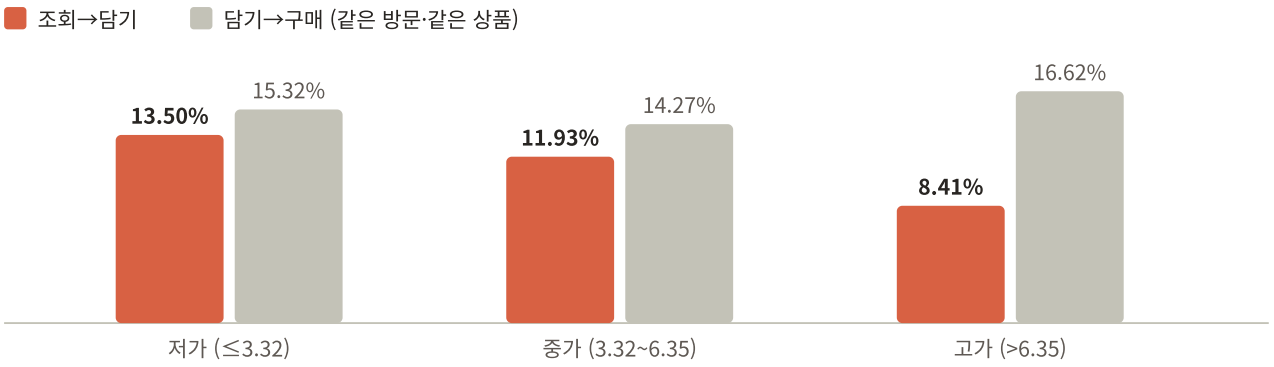
**같은 방문 기준을 읽을 때.** 담고도 그 방문에서 안 산 조합 782,605개(84.69%) 중 **16.92%**는 7일 안에 다른 방문에서 같은 상품을 샀습니다(B2). 같은 방문 미구매는 영구 이탈이 아니며, 담은 뒤의 고객 단위 이탈은 Q2(7일)에서 봅니다. 순서 조건을 빼면 조회→담기 14.99% · 담기→구매 14.84%로 두 단계가 비슷해집니다. 담기 조합의 25.9%는 담기가 조회보다 먼저이거나 같은 시각인 경우라, “조회 후 담기”만 셀 때 담기 단계가 더 좁아 보입니다.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-3. “가격은 ‘담을지’에만 작용하고, 첫 방문과 재방문의 차이는 작습니다”

**[그림 4-2] 가격대별 전환 (첫 조회 가격 배정, 같은 방문·같은 상품)**

비쌀수록 덜 담지만, 담은 뒤에는 고가가 오히려 약간 더 산다




### 가격 효과는 진짜인가

| 확인 | 저가 | 고가 |
|---|---|---|
| 조회→담기 (대표) | 13.50% | **8.41%** |
| 담기→구매 (대표) | 15.32% | 16.62% |
| 순서 무관 조회→담기 | 18.90% | 10.68% |
| 순서 무관 담기→구매 | 15.09% | 15.64% |

저가·고가 조회가 각각 100조합 이상인 카테고리 197개 중 **168개 (85.28%)**에서 고가 담기율이 낮았고, 격차 중앙값은 -5.13%p입니다. 카테고리 구성만으로 생긴 차이가 아닙니다(인과 검증은 아님). 장비류 카테고리(…255120)처럼 가격과 무관하게 거의 안 담기는 상품군도 있습니다 .

0가격 27,284조합은 별도 그룹으로 전체 퍼널에 포함. 통화 미확인.

### 첫 방문 vs 재방문 (관측 기준)

| | 조회 조합 | 조회→담기 | 담기→구매 |
|---|---|---|---|
| 관측 첫 방문 | 2,698,190 | 10.87% | **17.15%** |
| 관측 재방문 | 5,626,289 | 11.21% | 14.46% |

같은 상품으로 보면 첫 방문과 재방문의 담기율 차이는 0.3%p뿐이고, 담은 뒤에는 첫 방문이 더 삽니다.

### 가설 ① 판정

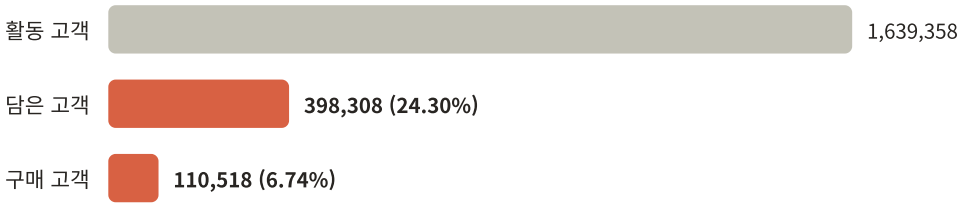
**부분 지지** 두 단계 전환율은 다르고(11.10% vs 15.31%), 이탈은 조회→담기에서 가장 큽니다. 가격이 오를수록 조회→담기는 13.50% → 8.41%로 떨어지지만, 담기→구매는 가격이 높다고 낮아지지 않습니다(14~17%).

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-2. “단계별 도달 · 전환율 · 이탈 표”

가장 많이 빠지는 곳은 담기 전 단계입니다. 고객 기준으로도, 같은 상품 기준으로도 조회→담기에서 이탈이 가장 큽니다.

### ① 고객 기준 퍼널 (기간 전체, 고객 1명은 1번만 집계)




| 단계 | 도달 인원 | 단계 전환율 | 누적 전환율 | 이탈 인원 | 이탈률 |
|---|---|---|---|---|---|
| 활동 (조회 이상) | 1,639,358 | — | 100% | — | — |
| 장바구니 담기 | 398,308 | 24.30% | 24.30% | **1,241,050** | **75.70%** |
| 구매 (담은 고객 중) | 109,607 | 27.52% | 6.69% | 288,701 | 72.48% |

구매 고객 전체는 110,518명(K1 6.74%), 그중 99.18%가 담기를 거쳐 구매. 담은 고객 구매율 27.52%는 기간 전체 기준이며, Q2의 7일 판정(I2 미구매율 77.65%)과는 관찰 기간이 다릅니다.

### ② 같은 상품 퍼널 (고객×세션×상품 조합, 조회 < 담기 < 구매 시간 순서)

| 단계 | 도달 조합 | 단계 전환율 | 누적 전환율 | 이탈 조합 | 이탈률 |
|---|---|---|---|---|---|
| 조회 | 8,324,479 | — | 100% | — | — |
| 담기 | 924,090 | F1 11.10% | 11.10% | **7,400,389** | **88.90%** |
| 구매 | 141,485 | F2 15.31% | F3 1.70% | 782,605 | 84.69% |

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-4. “여러 번 보고, 뺐다가 다시 담는 상품이 팔립니다”

### 같은 방문 안 행동 (같은 상품)

| 조회 횟수 | 조회→담기 | 담기→구매 |
|---|---|---|
| 1회 | 9.55% | 13.93% |
| 2회 | 20.62% | 18.83% |
| 3~4회 | 30.24% | 21.21% |
| 5회 이상 | 36.26% | 24.12% |

| 첫 담기 후 경로 (구매 전) | 담기 조합 | 담기→구매 |
|---|---|---|
| 삭제 없음 | 747,514 | 16.57% |
| 삭제 후 재담기 없음 | 156,679 | 7.16% |
| **삭제 후 재담기** | 19,897 | **32.17%** |

조회 1회 → 5회 이상이면 담기율 9.55% → 36.26%, 담은 뒤 구매율도 13.93% → 24.12%. 뺀 상품은 거의 안 팔리지만 다시 담으면 가장 잘 팔립니다(Q2 재담기형 56.5%와 같은 방향). 경로는 결과와 함께 생기는 기술통계입니다.

### 브랜드 (조회 규모 상위 8)

| 브랜드 | 조회 조합 | 담기 | 구매 |
|---|---|---|---|
| trind | 3,426,335 | 11.45% | 15.15% |
| roubloff | 520,800 | 12.47% | 16.07% |
| grace | 354,011 | 10.95% | 13.31% |
| invisibobble | 353,171 | 11.89% | 15.49% |
| marutaka-foot | 292,203 | 11.15% | 12.18% |
| essie | 180,494 | 10.57% | 18.50% |
| kamill | 172,619 | 8.22% | 17.94% |
| ikoo | 150,854 | 12.44% | 13.62% |

규모가 크고 담기율이 낮은 브랜드부터 상세페이지 개선 후보. 가격·상품 구성이 섞인 결과라 브랜드 경쟁력으로 단정하지 않습니다.

### 속도 · 시간대 · 요일 (UTC, 첫 조회 시각)

| 지표 | 결과 |
|---|---|
| 전환 속도 | 조회→담기 중앙값 0.48분(5분 안 89.5%) · 담기→구매 중앙값 13.87분(60분 안 86.17%) |
| 조회→담기 높은 시간 / 낮은 시간 | 1시 13.13% · 0시 12.76% / 15시 10.56% |
| 담기→구매 높은 시간 / 낮은 시간 | 11시 17.60% · 8시 17.28% / 2시 9.38% |
| 요일 순차 전환 (F3) | 월 1.70% · 화 1.73% · 수 1.66% · 목 1.76% · 금 1.78% · 토 1.62% · 일 1.63% |

> **Q1 정리.** 병목은 “보고 담는” 단계이고, 고가 상품에서 특히 좁습니다. 담은 뒤에는 가격과 무관하게 15% 안팎이 그 방문에서 사며, 사는 경우 대부분 1시간 안입니다. 반복 조회와 삭제 후 재담기는 구매 관심의 신호입니다. 시간대 결과는 현지 시간이 확인되지 않아 UTC로만 해석합니다.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-5. “담고도 7일 안에 안 산 고객 29.8만 명, 그중 3분의 2는 다시 오지도 않았습니다”

**384,380**
I1 의도 고객 (담기 1회+)
전체 담은 고객 398,308명

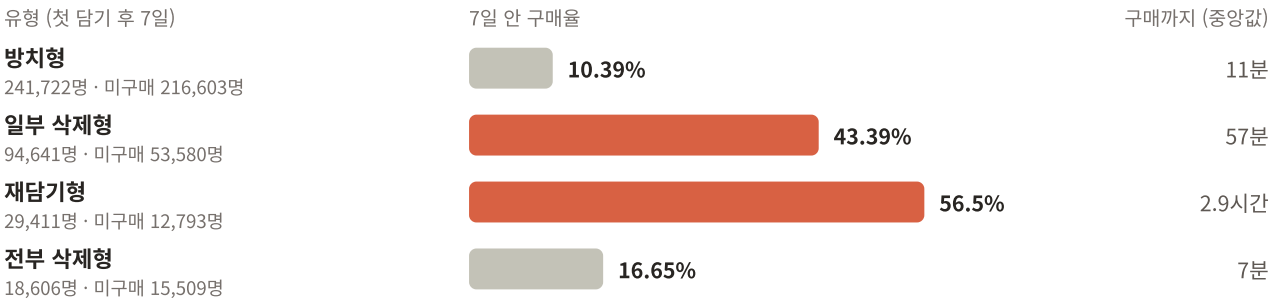
**77.65%**
I2 미구매율 · 298,485명

**24.30%**
I3 담기율
미구매 의도 고객 비중
18.21%

**952만**
I4 놓친 금액
남은 상품 707만 · 상위 1% 제외 795만

### [그림 4-3] 유형별 7일 구매율과 구매까지 시간 (4유형)

고객 기준 · 첫 담기 후 7일 · 삭제 없음 = 방치 / 남은 상품 없음 = 전부 삭제 / 남은 상품 중 뺐다 다시 담은 것이 있으면 재담기 / 나머지 = 일부 삭제




### 다시 왔는지

| 첫 담기 후 7일 | 의도 고객 | 7일 구매율 | 미구매 |
|---|---|---|---|
| **다시 안 옴** | 237,907 | 14.07% | **204,433 (68.5%)** |
| 다시 옴 | 146,473 | 35.79% | 94,052 (31.5%) |

다시 안 온 20.4만 명은 리마인드로 ‘다시 오게’, 다시 와서도 안 산 9.4만 명은 쿠폰·대안 상품처럼 ‘결정을 돕는’ 대응이 맞습니다.

### 첫 1시간 행동

| 1시간 안에 안 산 고객의 초기 행동 | 이후 구매율 |
|---|---|
| 재조회·삭제 없음 (미구매의 66.9%) | 6.79% |
| 재조회만 | 11.71% |
| 재조회·삭제 (재담기 없음) | 24.64% |
| 재조회·삭제·재담기 | **29.22%** |

1개만 담은 고객은 재담기를 해도 구매율이 0.1~0.5%p만 올라, 재담기는 여러 개 담은 고객에게만 강한 신호입니다.

**담은 상품 수:** 1개 6.48% · 2~3개 22.29% · 4개 이상 37.62%. 1개만 담고 안 산 고객이 미구매의 48.2%입니다.

---

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-6. “사는 고객은 1시간 안에 사고, 놓친 금액의 절반은 고가 상품입니다”

### 첫 담기 후 누적 구매율

| 기간 | 누적 구매율 | 30일 구매자 중 |
|---|---|---|
| 1시간 | 13.15% | 50.7% |
| 1일 | 18.79% | 72.9% |
| 3일 | 20.73% | 80.7% |
| 7일 | 22.35% | 87.2% |
| 14일 | 23.75% | 92.9% |
| 30일 | 25.48% | 100.0% |

구매의 58.7%가 1시간 안, 83.9%가 24시간 안입니다. 30일 구매자의 87%가 7일 안에 사서 7일 판정이 타당합니다.

### 7일 뒤에 돌아와 산 경우

| 7일 미구매 시도의 30일 안 같은 상품 구매 | 비율 |
|---|---|
| 전체 | 2.99% (중앙 13.1일째) |
| 핵심 고가치형 고객 | **8.44%** |
| 일반형 고객 | 2.96% |
| 삭제 후 재담기 시도 / 삭제만 | 5.78% / 0.83% |

7일 미구매가 영구 이탈은 아닙니다. 늦은 복귀는 고가치·재담기 고객에 몰려 있어 7~14일 재알림을 실험해 볼 근거가 됩니다.

### 첫 담은 상품 가격대별 (상품 수 3등분)

| 가격대 | 의도 고객 | 7일 구매율 | 미구매 | 놓친 금액 |
|---|---|---|---|---|
| 저가 | 122,819 | 19.19% | 99,250 | **226만** |
| 중가 | 122,774 | 20.73% | 97,328 | **248만** |
| 고가 | 138,395 | 26.59% | 101,593 | **477만** |

고가 상품을 처음 담은 고객이 오히려 더 많이 사지만(26.59%), 놓친 금액은 전체의 절반입니다. 방치형이 남긴 상품 수는 저가가 44%지만 금액은 고가가 65%입니다.

### 놓친 금액을 읽을 때

- 미구매 고객이 많이 담은 상위 20개 상품 중 16개는 상품 미구매율이 평균(77.33%)보다 낮습니다. **많이 버려진 상품 ≠ 안 팔리는 상품**이라 규모와 전환율을 함께 봐야 합니다.
- 놓친 금액이 큰 고객은 비싼 상품 하나가 아니라 **서로 다른 상품을 많이 담은** 고객입니다(금액 차이의 67.8%가 상품 수, 31.7%가 가격).
- 상위 1% 고객이 놓친 금액의 16.5%를 차지합니다. 규모는 952만(전체)과 795만(상위 1% 제외) 범위로 보고합니다.

---

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-7. “첫 구매 고객의 12.93%만 30일 안에 다시 삽니다. 첫 달이 고비입니다”

**12.93%**
V1 30일 재구매율
91,979명 기준

**12.3일**
재구매까지 중앙값
(30일 안 재구매자)

**20.57%**
참고: 5개월 동안 2번 이상 산 고객
79.4%는 한 번만 삼

**46.82**
V4 고객당 30일 결제 주문당 금액 41.13

| 첫 구매 월 | 구매 고객 | 30일 재구매 | 간격 | 14일 재구매 | +1월 | +2월 | +3월 | +4월 |
|---|---|---|---|---|---|---|---|---|
| 10월 | 25,762 | **16.73%** | 13.7일 | 8.59% | 18.49% | 12.78% | 13.19% | 10.45% |
| 11월 | 26,761 | **13.44%** | 12.0일 | 7.66% | 9.78% | 10.15% | 7.92% | — |
| 12월 | 19,702 | **9.03%** | 10.3일 | 5.46% | 8.52% | 6.46% | — | — |
| 1월 | 20,427 | **11.19%** | 12.0일 | 6.33% | 8.86% | — | — | — |
| 2월 | 17,866 | — | — | 6.19% | — | — | — | — |

주문 = 구매일. +N월은 첫 구매 월 M 고객 중 M+N월에 산 비율(V3). 2월 코호트는 30일을 다 볼 수 없어 14일만 계산.

> **읽는 법.** 10월 코호트가 모든 지표에서 가장 높은 것은 데이터가 10월부터라 9월 이전부터 사던 고객이 섞였기 때문으로 보입니다. 비교는 11월 이후 코호트로 합니다. 11월 이후 코호트의 다음 달 재구매율은 8.5~9.8%이고, 한 번 남은 고객은 그 뒤에도 비슷하게 남습니다(11월: +1월 9.8% → +2월 10.1% → +3월 7.9%). 12월 코호트가 가장 낮아 연말 일회성 구매가 섞였을 수 있습니다.

### 행사 날 들어온 고객

급등일(10일)에 처음 산 고객은 첫 주문이 더 크지만(32.01 vs 27.29) 30일 재구매율이 더 낮습니다(11.41% vs 13.19%, $p<0.001$). 특별 기간 코호트 분석에서도 할로윈·블랙프라이데이·연말에 처음 관찰된 고객의 구매 비중과 고객당 결제액이 일반 기간과 같거나 낮았습니다. 행사 성과는 당일 매출이 아니라 30일 재구매까지 봐야 합니다.

---

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-8. “리텐션 매트릭스: 첫 구매 다음 달에 돌아오는 고객은 10명 중 1명 안팎입니다”

첫 구매 월(코호트)별로, 첫 구매 후 N개월째 달에 다시 구매한 고객 비율입니다. 주문 = 고객의 구매일(UTC).

**[그림 4-4] 첫 구매 월 코호트 리텐션 히트맵**

M+0 = 첫 구매 달(100%) · 칸 = 그 달에 한 번 이상 다시 구매한 고객 비율 · 회색 = 관찰 기간 밖

| 첫 구매 월 | 코호트 고객 | M+0 | M+1 | M+2 | M+3 | M+4 |
|---|---|---|---|---|---|---|
| **10월** | 25,762 | **100%** | **18.49%** | **12.78%** | **13.19%** | **10.45%** |
| **11월** | 26,761 | **100%** | **9.78%** | **10.15%** | **7.92%** | |
| **12월** | 19,702 | **100%** | **8.52%** | **6.46%** | | |
| **1월** | 20,427 | **100%** | **8.86%** | | | |
| **2월** | 17,866 | **100%** | | | | |

| 첫 구매 월 | 10월 | 11월 | 12월 | 1월 |
|---|---|---|---|---|
| 30일 재구매율 (V1) | 16.73% | 13.44% | 9.03% | 11.19% |
| 14일 재구매율 | 8.59% | 7.66% | 5.46% | 6.33% |

**읽는 법** — 11월 이후 코호트는 다음 달 재구매율이 8.5~9.8%로 첫 달 이탈이 큽니다. 10월 코호트가 높은 것(18.49%)은 10월 이전부터 사던 기존 고객이 섞였기 때문이라(데이터 시작 시점) 비교는 11월 이후 코호트로 합니다.

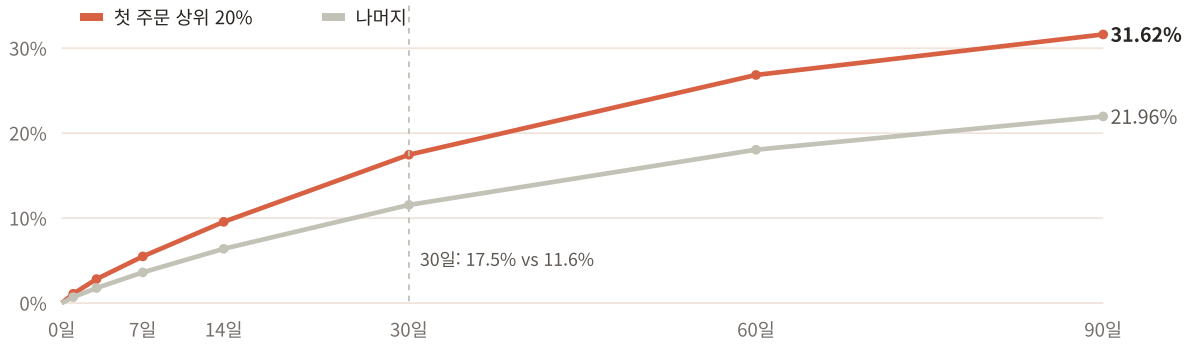
**활용** — 첫 구매 후 첫 달이 고비 → 첫 구매 후 2주 안 두 번째 구매 혜택.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-9. “첫 주문이 큰 고객은 재구매에 더 빨리 이릅니다 (더 많이 돌아와서)”

**[그림 4-5] 첫 구매 후 누적 재구매율 (생존 분석)**

재구매하지 않은 고객까지 포함, 2/29 관찰 종료 시점에서 중도 절단 · 1일부터 90일까지 모든 시점에서 약 1.5배




| 첫 주문 금액 기준 | 상위 20% | 나머지 | 읽는 법 |
|---|---|---|---|
| 7일 · 30일 · 90일 누적 재구매율 | **5.5% · 17.46% · 31.62%** | 3.6% · 11.55% · 21.96% | 모든 시점에서 약 1.5배 |
| 고객 25%가 재구매하기까지 | **52일** | 123일 | 같은 비율에 2.4배 빨리 도달 |
| 재구매 속도 (위험비) | **1.54배** · 로그순위 검정 $\chi^2$=830 | | 상위 10% 1.55배 · 11월 이후 1.51배 |
| 재구매한 고객끼리의 간격 중앙값 | 12.48일 (전체 24일) | 12.28일 (전체 25일) | 30일 안 재구매자 기준 · 차이 없음 |
| 30일 재구매율 (V1 분모 기준) | 17.64% | 11.76% | +5.89%p [5.29, 6.49], 코호트별 1.4~1.5배 |

**가설 ③ 판정: 지지 (조건부).** 재구매하지 않은 고객까지 넣어 시간을 따라가면, 첫 주문 상위 20% 고객은 재구매에 **1.54배 빠른 속도**로 이릅니다(고객 25% 도달 52일 vs 123일). 다만 **재구매한 고객끼리의 간격은 같아서**(12.5일 vs 12.3일), 차이는 ‘더 자주 사서’가 아니라 ‘다시 오는 고객이 많아서’ 생깁니다. 첫 주문 금액은 순환 논리를 피하려고 썼으며, 관여도가 높은 고객이 원래 크게 사는 것일 수 있어 인과는 아닙니다.

### 다음 구매를 가장 잘 예측하는 건 반복 구매 이력 (시간 분리 검증)

| 1월까지 이력 → 2월 구매율 | 과거 1일 구매 | 2일 구매 | 3일 이상 구매 |
|---|---|---|---|
| 마지막 구매 1~30일 전 | 7.90% | 20.23% | **45.71%** |
| 31~60일 전 | 4.46% | 14.29% | 28.83% |
| 61~90일 전 | 3.93% | 11.59% | 16.48% |
| 91일 이상 | 2.81% | 5.30% | 7.44% |

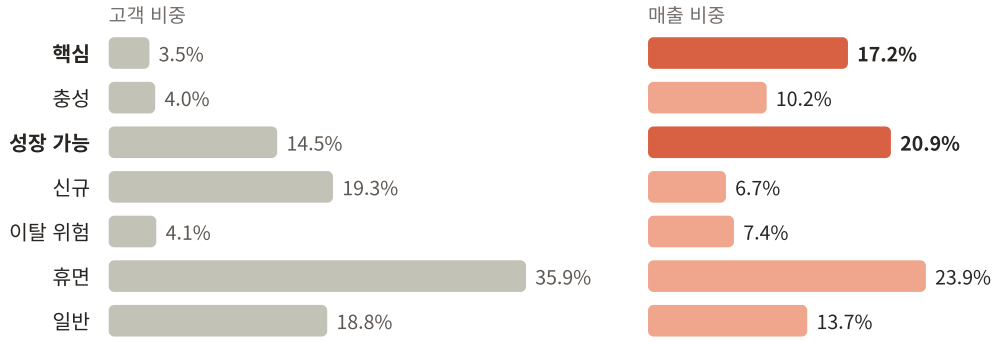
같은 최근성이면 많이 산 고객이, 같은 구매 횟수면 최근에 산 고객이 다음 달에 더 삽니다. 과거 데이터로 미래를 확인한 결과라 같은 기간 비교보다 근거가 강합니다. 30일 안에 다시 산 고객은 첫 구매 전에도 더 많이 둘러보고(조회 중앙값 12 vs 8회, 브랜드 6 vs 4개) 더 오래 고민한 뒤(첫 활동→첫 구매 중앙값 2.7일 vs 0.1일) 샀습니다(구매일 주문 기준 재계산).

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-10. “핵심·성장 가능 고객 18%가 매출의 38%, 고액 휴면 고객이 다음 기회입니다”

**[그림 4-6] RFM 7유형: 고객 비중과 매출 비중**

구매 고객 110,518명 · R·F·M 각 1~5점 · 기준일 3/1




| 유형 | 뜻 | 고객 | 액션 | 움직일 지표 |
|---|---|---|---|---|
| **핵심** | 최근에 자주, 많이 산 고객 | 3,825 | VIP 혜택, 신제품 선공개 | 재구매율 유지 |
| 충성 | 3번 이상 사고 아직 활동 중 | 4,407 | 적립·등급 혜택 | 구매 횟수 |
| **성장 가능** | 최근 2번 샀거나 처음부터 크게 산 고객 | 15,989 | 세 번째 구매 쿠폰 | 30일 재구매율 |
| 신규 | 최근 1번 산 고객 | 21,348 | 2주 안 두 번째 구매 혜택 | 30일 재구매율 |
| 이탈 위험 | 2번 이상 샀지만 오래 안 옴 | 4,577 | 윈백 쿠폰 | 재방문율 |
| 휴면 | 1번 사고 오래 안 옴 | 39,631 | 고액 휴면부터 윈백, 나머지 저비용 자동 메시지 | 재방문율 |
| 일반 | 나머지 | 20,741 | 기본 마케팅 | — |

RFM 규칙(기준일 3/1, 주문 = 구매일): 핵심 = R≥4·4회 이상·M≥4, 성장 가능 = 최근 2회 또는 1회·M≥4. 매출 상위 0.1% 고객을 빼도 핵심 매출 비중은 17.2% → 15.8%로 거의 같습니다.

### 매출 집중

- 고객 상위 10%가 매출의 43.04%, 상위 20%가 **59.32%** (80/20보다 덜 쏠림)
- 상품 상위 20%가 매출의 80%, 5개월 동안 한 번도 안 팔린 상품 25.3%
- 모든 유형에서 매출의 77~81%가 같은 A등급 상품 → 유형별 대응은 상품 구성보다 **구매 빈도·타이밍**에 맞춤

### 최상위 장기 휴면 (결제액 상위 20% × 88일+)

- 금액 상위 20% 중 **6,189명(28.0%)**이 88일 넘게 구매 없음, 매출의 13.66%
- 그중 63.8%는 1일만 산 고객, 36.2%는 반복 구매 이력 고객 → 추가 구매 유도와 반복 구매 재개를 나눠 접근
- **고액 휴면** (RFM × ABC): 매출 80%를 만드는 A등급 고객 중 이탈 위험·휴면이 **16,633명**, 매출의 22.59%
- K-means로도 ‘오래된 1회·고액’ 그룹(31.4%)이 매출의 34.6% → 휴면 중 고액 고객이 윈백 1순위

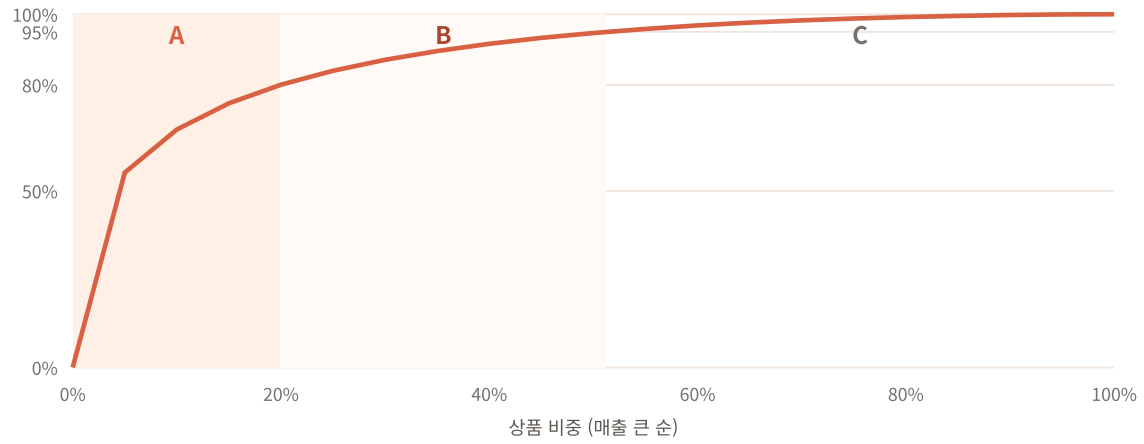
팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-11. “팔린 상품의 20%(A등급 8,140개)가 매출의 80%를 냅니다”

상품별 구매 매출을 큰 순서로 쌓아 누적 80%까지 A, 95%까지 B, 나머지를 C로 나눴습니다. 상위 1% 상품만으로 매출의 31.48%입니다.

**[그림 4-7] 상품 파레토 곡선**

팔린 상품 40,777개 · 매출 6,348,268 (가격 × 1개, 0·음수 가격 제외)




| 등급 | 기준 | 상품 수 | 상품 비중 | 매출 비중 | 판매 건수 비중 | 건당 평균 가격 |
|---|---|---|---|---|---|---|
| **A** | 누적 매출 80%까지 | 8,140 | 19.96% | **80.0%** | 69.13% | 5.71 |
| **B** | 80~95% | 12,743 | 31.25% | **15.0%** | 21.37% | 3.47 |
| **C** | 95~100% | 19,894 | 48.79% | **5.0%** | 9.51% | 2.6 |

**읽는 법** — A등급은 평균 가격도 높아(5.71 vs C 2.6) 매출 집중이 판매량과 가격 모두에서 옵니다. 기간 중 한 번도 팔리지 않은 상품 13,794개(전체의 25.28%)는 등급에서 제외했습니다.
**활용** — A등급은 재고·노출 우선 관리, 장바구니 리마인드 메시지의 대표 상품으로 사용. C등급은 묶음·추천 상품으로 활용.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-12. “RFM 세그먼트 × ABC 등급 교차표: 매출 상위 고객의 상당수가 이미 휴면입니다”

ABC는 고객별 누적 결제액 기준(누적 80% A, 95% B), RFM은 기준일 3/1의 7유형입니다. RFM은 고객 단위라 고객 ABC와 교차했습니다.

| RFM 유형 | A (매출 80%) | B | C | 고객 합계 | 매출 비중 |
|---|---|---|---|---|---|
| **핵심** | 3,825<br>매출 17.2% | 0<br>매출 0.0% | 0<br>매출 0.0% | 3,825 | 17.2% |
| **충성** | 4,136<br>매출 10.1% | 259<br>매출 0.1% | 12<br>매출 0.0% | 4,407 | 10.3% |
| **성장 가능** | 14,079<br>매출 20.1% | 1,760<br>매출 0.8% | 150<br>매출 0.0% | 15,989 | 20.9% |
| **신규** | 801<br>매출 0.5% | 10,666<br>매출 4.5% | 9,881<br>매출 1.7% | 21,348 | 6.7% |
| **이탈 위험** | 3,575<br>매출 6.9% | 925<br>매출 0.4% | 77<br>매출 0.0% | 4,577 | 7.4% |
| **휴면** | 13,058<br>매출 15.7% | 14,317<br>매출 6.0% | 12,256<br>매출 2.2% | 39,631 | 23.9% |
| **일반** | 7,393<br>매출 9.5% | 7,280<br>매출 3.1% | 6,068<br>매출 1.1% | 20,741 | 13.7% |
| **합계** | **46,867**<br>매출 80.0% | **35,207**<br>매출 15.0% | **28,444**<br>매출 5.0% | 110,518 | 100% |

칸 = 고객 수, 아래 작은 글씨 = 전체 매출 대비 비중. 색이 진할수록 고객이 많음.

**핵심 발견**

- 핵심 고객 3,825명은 전원 A등급 — 유지 대상
- **고액 휴면**(A등급이면서 이탈 위험·휴면) **16,633명이 매출의 22.59%** — 윈백 1순위
- 성장 가능 고객의 88.05%가 이미 A등급 — 세 번째 구매로 핵심 전환 후보
- 신규 고객은 대부분 B·C등급 — 두 번째 구매 혜택으로 등급 상승 유도

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

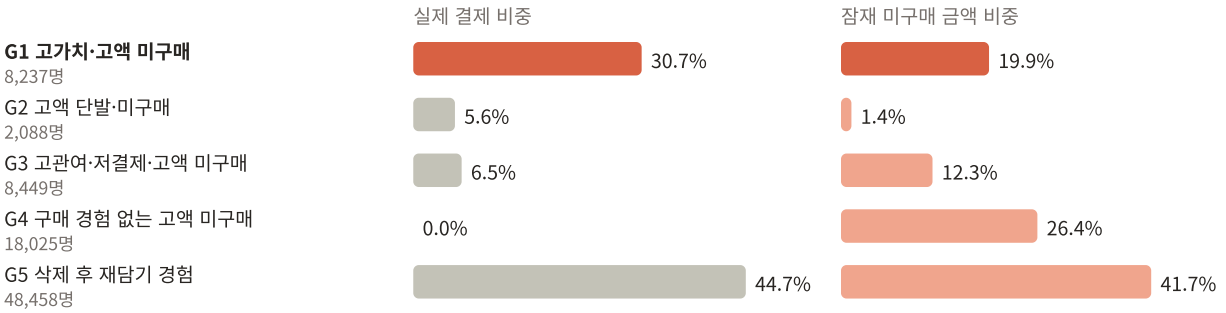
## 4-13. “가장 많이 사는 고객이 가장 많이 남기기도 합니다”

구매 시도 단위(같은 고객·상품의 7일 묶음)와 2×2 가치 세그먼트(방문 4회 이상 × 금액 상위 15%)로 Q1~Q3를 고객 단위로 이었습니다. 잠재 미구매 금액(1,837만)은 산 고객이 남긴 상품까지 포함한 상품 단위 값입니다.

| 질문 | 답 |
|---|---|
| 미구매 금액 상위 10%는 누구인가 | 21.7%가 핵심 고가치형, **50.8%는 구매 경험 없는 고객**. 핵심 고가치형의 57.3%가 미구매 상위 10% (Lift 3.49) |
| 고가치 고객도 장바구니를 버리나 | 핵심 고가치형 시도의 63%가 7일 미구매, 59.8%가 삭제 후 재담기 경험. 평균 구매 소요 19.8시간 (일반형 4.1시간) → 장바구니를 비교·보관 목록으로 씀 |
| 미구매 금액은 왜 큰가 | 서로 다른 상품을 많이 담아서(67.8%), 평균 가격 31.7%, 같은 상품 반복 0.5% |
| 이상 고객을 빼도 같은가 | P99 초과 1.09%가 잠재 금액 28%·실제 결제 25% 차지. 빼도 방향은 같고 규모만 줄어듦 |

### [그림 4-8] 우선 관리 후보군 (서로 겹칠 수 있음)

실제 결제 비중 vs 잠재 미구매 금액 비중




**우선순위 제안.** ① G1 고가치·고액 미구매형 8,237명: 실제 결제 30.7%·잠재 19.9%·30일 후속 구매 8.35%로 근거가 가장 많이 겹침 → 담아 둔 상품 요약 알림, 7~14일 재알림 ② G3 고관여·저결제·고액 미구매형: 시도 전환율 15.6%로 가장 낮음 → 묶음·무료배송 기준 안내 ③ G5 재담기 경험형: 재담기 직후 실시간 안내 ④ G4 구매 경험 없는 고액 미구매형은 규모는 크지만 근거가 약해 탐색 대상.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-14. “방치형은 ‘처음 와서 담고, 다시 오지 않은’ 신규 방문자입니다”

장바구니에 담은 고객을 7일 뒤 장바구니 상태(4유형)로 나누고, 고객 특성(첫 방문 여부 · 담기 전 방문 횟수 · 7일 안 재방문 · 이전 구매)으로 다시 쪼갰습니다.

| 유형 | 고객 | 7일 미구매 | 7일 구매율 | 첫 방문에 담음 | 7일 안 재방문 | 미구매자 재방문 | 담기 전 구매 경험 |
|---|---|---|---|---|---|---|---|
| **방치형** | 241,722 | 216,603 | 10.39% | 80.13% | 23.33% | 21.17% | 0.08% |
| **일부 삭제형** | 94,641 | 53,580 | 43.39% | 72.62% | 62.2% | 59.3% | 0.14% |
| **재담기형** | 29,411 | 12,793 | 56.5% | 68.83% | 74.59% | 70.14% | 0.12% |
| **전부 삭제형** | 18,606 | 15,509 | 16.65% | 68.91% | 44.0% | 43.53% | 0.06% |

**읽는 법**

· 방치형의 **80.1%는 첫 방문에 담았고**, 미구매 방치형의 **78.8%는 7일 안에 다시 오지 않았습니다**(재방문 21.2%). 사이트에 다시 올 이유를 만들어 주는 외부 알림(푸시·이메일)이 필요한 이유입니다.

· 재담기형·일부 삭제형은 70% 안팎이 7일 안에 다시 와서 장바구니를 고칩니다. 이미 사이트 안에 있으니 **재방문 순간의 사이트 안 안내**(쿠폰 배너·재고 알림)가 맞습니다.

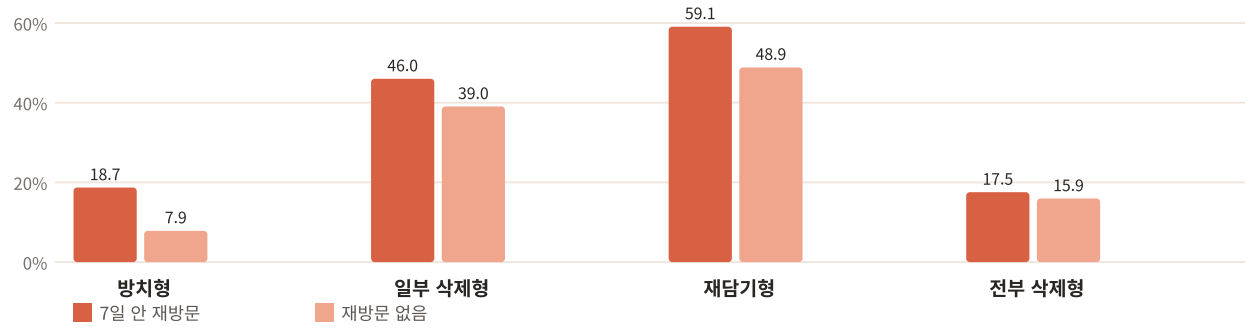
· 담기 전 구매 경험이 있는 고객은 0.1% 안팎뿐입니다. 장바구니를 처음 쓰는 시점이 곧 첫 구매 직전이라, **4유형은 사실상 신규 고객의 첫 구매 전환 문제**입니다.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-15. “방치형도 다시 오거나, 전에 몇 번 와 본 고객은 2~3배 더 삽니다”

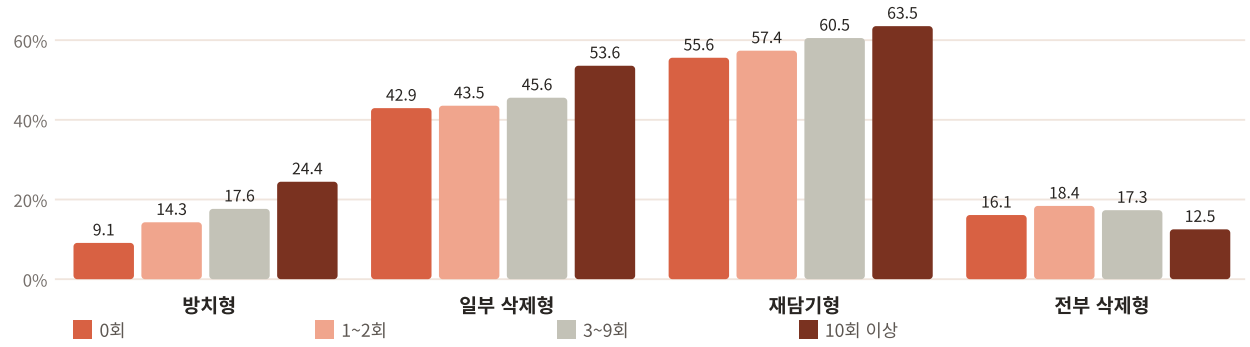
### [그림 4-9] 7일 안 재방문 여부별 7일 구매율

첫 담기 세션 이후 7일 안에 다른 세션이 있었는지




### [그림 4-10] 첫 담기 전 방문(세션) 횟수별 7일 구매율

0회 = 첫 방문에 바로 담음




| 방치형 세부 | 고객 | 미구매 | 7일 구매율 |
|---|---|---|---|
| 첫 방문에 담음 | 193,680 | 175,844 | 9.21% |
| 재방문에 담음 | 48,042 | 40,759 | 15.16% |
| 7일 안 재방문 없음 | 185,324 | **170,754** | **7.86%** |

**시사점** — 리마인드 1순위는 **첫 방문에 담고 다시 오지 않은 방치형 약 17만 명**(7일 구매율 7.9%)입니다. 다시 온 방치형은 18.7%로 2배 이상 사므로, 리마인드의 목표는 ‘구매’보다 먼저 **‘다시 오게 하기’**로 잡는 것이 맞습니다. 반대로 재담기형은 재방문 여부와 상관없이 절반 이상 사므로 할인보다 결제 마찰 제거가 우선입니다.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-16. “핵심 고객은 ‘많이 둘러보고, 담고 빼며 고르는’ 고객입니다”

구매 고객을 RFM 7유형으로 나누고 방문·탐색·장바구니 행동을 비교했습니다. (중앙값 기준, 주문당 금액은 평균)

| 유형 | 고객 | 평균 주문 | 주문당 금액 | 방문(세션) | 조회수 | 주문당 조회 | 삭제 경험 | 첫 방문에 첫 구매 | 첫 방문→첫 구매 | 마지막 구매 후 |
|---|---|---|---|---|---|---|---|---|---|---|
| **핵심** | 3,825 | 5.78 | 49.32 | 30 | 104 | 19.3 | 99.48% | 26.51% | 3.8일 | 16일 |
| **충성** | 4,407 | 3.23 | 45.7 | 18 | 60 | 18.7 | 97.89% | 25.16% | 3.9일 | 33일 |
| **성장 가능** | 15,989 | 1.44 | 57.93 | 7 | 21 | 15.5 | 87.9% | 27.48% | 1.9일 | 25일 |
| **신규** | 21,348 | 1.0 | 19.95 | 3 | 8 | 8.0 | 71.38% | 36.73% | 0.1일 | 26일 |
| **이탈 위험** | 4,577 | 2.27 | 45.14 | 11 | 36 | 16.5 | 94.32% | 31.72% | 0.9일 | 102일 |
| **휴면** | 39,631 | 1.0 | 38.24 | 3 | 10 | 10.0 | 76.48% | 39.61% | 0.1일 | 115일 |
| **일반** | 20,741 | 1.14 | 36.61 | 4 | 11 | 10.0 | 75.94% | 35.44% | 0.2일 | 70일 |

**읽는 법**

· **핵심·충성**: 방문 18~30회, 조회 60~104회, 거의 모두(98~99%) 장바구니에서 빼 본 경험. 주문 한 번에 19회 안팎을 둘러봅니다 → 비교·탐색이 쉬운 화면(찜·비교·재입고 알림)이 유지에 중요합니다.

· **성장 가능**: 주문당 금액이 57.93로 가장 높고, 첫 방문 후 첫 구매까지 1.9일 → 큰 첫 주문 고객에게 세 번째 구매 혜택.

· **신규·휴면**: 방문 3회, 조회 8~10회, 첫 방문 당일 바로 삼(0.1일), 첫 방문에 첫 구매한 비율 37~40% → **‘들어와서 바로 사고 떠나는’ 목적형 고객**. 첫 구매 직후 두 번째 구매 이유를 만들어야 합니다.

· **이탈 위험**: 마지막 구매 후 102일, 주문당 금액 45.14 → 과거 가치가 있던 고객이라 윈백 우선.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 4-17. “7일 미구매 고객의 93%는 구매 경험이 없고, 고가치 고객의 방치는 3,289명입니다”

7일 안에 사지 않은 고객 298,485명을 RFM 유형(기간 종료 3/1 기준)으로 다시 나눴습니다. ‘구매 없음’은 기간 내내 한 번도 사지 않아 RFM 유형이 없는 고객입니다.

| RFM 유형 | 방치형 | 일부 삭제형 | 재담기형 | 전부 삭제형 | 합계 |
|---|---|---|---|---|---|
| **핵심** | 458 | 502 | 170 | 35 | 1,165 |
| **충성** | 643 | 512 | 191 | 55 | 1,401 |
| **성장 가능** | 2,188 | 1,591 | 474 | 168 | 4,421 |
| **신규** | 2,100 | 1,202 | 354 | 175 | 3,831 |
| **이탈 위험** | 361 | 328 | 120 | 21 | 830 |
| **휴면** | 2,480 | 1,888 | 609 | 197 | 5,174 |
| **일반** | 2,006 | 1,360 | 355 | 168 | 3,889 |
| **구매 없음** | 206,367 | 46,197 | 10,520 | 14,690 | 277,774 |
| **합계** | **216,603** | **53,580** | **12,793** | **15,509** | **298,485** |

**타깃 정리**

- **첫 구매 전환 (구매 없음 × 방치형) 206,367명** — 장바구니 리마인드 + 첫 구매 혜택
- **고가치 방치 (핵심·충성·성장 가능 × 방치형) 3,289명** — 담아 둔 상품 요약 알림, 할인 없이도 반응 가능
- **장바구니를 방치한 고액 휴면 (고액 휴면 × 방치형 7일 미구매) 1,269명** — 리마인드와 윈백을 한 메시지로

주의: RFM 유형은 기간 종료 시점 기준이라 장바구니에 담은 뒤의 구매가 포함됩니다. 유형별 7일 구매율은 부풀려지므로 사용하지 않고, 미구매 고객 수만 봅니다. 담기 전에 구매 경험이 있던 고객은 4유형 전체의 0.1%(382명)뿐이라 이전 구매 여부로는 나누지 않았습니다.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

# Chapter.5 해석과 마케팅 제언

## 5-1. "가설 ①은 부분 지지, ②는 지지, ③은 조건부 지지"

### ① View→Cart와 Cart→Purchase 전환율에 차이가 있으며, 이탈은 View→Cart에서 가장 크고 가격이 높을수록 커질 것이다 부분 지지

| 세부 주장 | 근거 | 판정 |
|---|---|---|
| 두 단계 전환율이 다르다 | 같은 방문·같은 상품 11.10% vs 15.31% | 맞음 |
| 이탈은 View→Cart가 가장 크다 | 같은 상품 기준 이탈률 88.9% vs 84.69%, 빠진 조합 740만 vs 78만 → 비율·수 모두 조회→담기. 단 순서 무관이면 두 단계가 비슷 (14.99% vs 14.84%) | 맞음 (순서 기준) |
| 가격이 높을수록 이탈이 크다 | 조회→담기 13.50% → 8.41% / 담기→구매 15.32% vs 16.62%로 고가가 오히려 높음. 카테고리 85.28%에서 같은 방향 | 담기 단계만 |

**수정 가설:** 이탈은 상품을 보고 담는 단계에서 가장 크고 가격은 이 결정에만 작용한다. 담은 뒤의 이탈은 가격보다 다시 오는지(B2 16.92%, Q2)가 좌우한다.

### ② 장바구니 도달 이후의 행동 유형에 따라 최종 구매전환율과 구매까지 소요시간에 차이가 있을 것이다 지지

| 근거 | 결과 |
|---|---|
| 7일 유형 4가지 (재담기 분리) | 재담기형 56.5% · 2.9시간 / 일부 삭제형 43.39% · 57분 / 전부 삭제형 16.65% · 7분 / 방치형 10.39% · 11분 ($\chi^2$=64,173) |
| 첫 1시간 행동 | 재조회·삭제·재담기 29.22% vs 무행동 6.79% |
| 구매 시도 | 삭제 후 재담기 39.26% vs 삭제만 6.66%, 구매의 57%가 1시간 안 |

네 사람의 방식이 달라도 결론이 같습니다. 단, 많이 담고 고치는 고객은 원래 관여도가 높은 고객이라 행동이 구매를 '일으킨다'고는 볼 수 없습니다.

### ③ 상위 N% 고가치 고객은 첫 구매 후 재구매까지 걸리는 기간이 일반 고객보다 짧을 것이다 지지 (조건부)

생존 분석(재구매 안 한 고객 포함): 첫 주문 상위 20%의 재구매 위험비 1.54배, 고객 25%가 재구매하기까지 52일 vs 123일, 7·30·90일 모든 시점에서 약 1.5배. 상위 10%(1.55배)와 11월 이후 코호트(1.51배)에서도 같음. 단 재구매한 고객끼리의 간격은 12.48일 vs 12.28일로 같음. **수정 가설:** 첫 주문이 큰 고객은 다시 돌아오는 비율이 높아 재구매에 더 빨리 이르지만, 돌아온 고객끼리의 구매 주기는 같다. 다음 구매는 반복 구매 이력과 최근성이 가장 잘 예측한다.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 5-2. “세 가지 분석 질문에 대한 답”

### Q1. 구매 전환과 이탈 행동은 고객의 행동 패턴·시간대·가격대에 따라 어떻게 다른가?

- **병목은 담기 단계**: 같은 방문에서 본 상품의 11.10%만 담기고, 담은 상품의 15.31%가 그 방문에서 팔림(순차 전환 1.70%)
- **행동 패턴이 가장 큰 차이**: 같은 상품 조회 5회+ 담기율 36.26% vs 1회 9.55%, 삭제 후 재담기 32.17% vs 재담기 없음 7.16%
- **가격대는 담기 단계에서만**: 조회→담기 13.50% → 8.41%, 담기→구매는 가격이 높다고 낮아지지 않음(14~17%, 고가 16.62%). 카테고리 85.28%에서 같은 방향. 첫 방문·재방문 차이는 작음
- **시간**: 담은 뒤 구매는 중앙값 13.87분, 86.17%가 1시간 안. 시간대는 UTC 기준이라 해석에 주의

### Q2. 구매 의도가 높지만 구매하지 않은 고객은 누구이며, 규모는 얼마인가?

- **규모**: 담은 고객 38.4만 명 중 7일 안에 안 산 고객 29.8만 명(77.65%), 활동 고객의 18.2%. 놓친 금액 952만(남은 상품 707만, 상위 1% 제외 795만)
- **누구**: ① 담고 그대로 둔 방치형 72.6%(21.7만) — 산 고객의 절반이 11분 안에 삼 ② 담은 뒤 다시 오지 않은 고객 68.5%(20.4만) ③ 상품 1개만 담은 고객 48.2% ④ 다시 와서도 안 산 9.4만 명
- **금액**: 고가 상품을 처음 담은 고객이 놓친 금액의 절반. 큰 금액은 많은 상품을 담아 둔 소수 고객(상위 1%가 16.5%)과 고가치 고객에게 몰림

### Q3. 지속적인 관리 가치가 높은 고객의 행동 특성은 무엇이며, 유형별로 어떤 전략이 효과적일까?

- **특성**: 반복 구매 이력과 최근성이 다음 구매를 가장 잘 예측(45.7% vs 7.9%). 첫 구매 전에 더 많이 둘러보고 더 크게 삼. 장바구니를 비교·보관 목록처럼 많이 쓰고, 7일 뒤에도 돌아와 삼(8.4%)
- **규모**: 30일 재구매율 12.9%, 79.4%는 한 번만 삼. 상위 20% 고객이 매출의 59.3%
- **전략**: 신규 — 두 번째 구매 지원 · 성장 가능 — 다음 구매 지원 · 핵심·충성 — 편의 중심 유지 · 최상위 장기 휴면 — 복귀 시험

---

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 5-3. “전략 1 · 돌아오게 하고, 구매를 돕는다”

> **근거** — 장바구니에 담은 고객의 **77.65%**가 7일 안에 사지 않았고, 미구매 고객의 **68.5%**는 7일 안에 다시 오지 않았습니다. 같은 7일 기준으로 장바구니를 다시 고친 재담기형은 **56.5%**가 샀지만, 손대지 않은 방치형은 **10.39%**만 샀습니다. 돌아올 계기를 만들고, 돌아온 순간 선택과 구매를 도와야 합니다.

### “① 복귀 알림”

| | |
|---|---|
| **대상** | 담은 뒤 구매하지 않았고 아직 돌아오지 않은 고객 (방치형 미구매 21.7만 명, 그중 재방문 없는 17.1만 명이 중심) |
| **실행** | 미발송 / 1시간 / 24시간 알림을 비교합니다. 발송 직전 구매·재방문 여부를 다시 확인하고, 관심 상품으로 바로 연결되는 알림을 1회 보냅니다. 첫 시험은 할인 없이 시작합니다. |
| **예시** | “관심 상품을 이어서 확인하세요.” → 홈 화면이 아니라 해당 상품·장바구니로 이동 |

### “② 돌아온 뒤 화면 지원”

| 고객이 하는 행동 | 화면에서 제공할 도움 |
|---|---|
| 한 상품을 다시 확인 | 상품 특징·리뷰·배송 조건을 바로 확인 (고가 상품은 담기 전 정보 보강: 리뷰·용량당 가격) |
| 여러 상품을 비교·조정 | 비교할 상품을 골라 용량·가격·사용 정보를 나란히 확인 |
| 구매할 상품을 결정 | 선택 상품의 총액과 배송 조건을 확인하고 구매 화면으로 이동 |

### “③ 종료·중복 처리”

구매하면 미구매 안내를 중단합니다. 장바구니를 방치한 고액 휴면처럼 윈백 대상과 겹치면 메시지를 하나로 합칩니다.

| | |
|---|---|
| **기대 효과 (가정)** | 방치형 미구매 고객 216,603명의 5%(10,830명)가 알림으로 돌아와 담아 둔 상품을 산다면 |
| **결과** | **K1 6.74% → 7.37% (+0.63%p) · 매출 +3.5% (3%면 K1 7.12% · 매출 +2.1%, 10%면 K1 8.00% · 매출 +7.0%)** |

성과 판단: 알림은 첫 담기 후 7일 구매율, 화면 지원은 배정 후 7일 구매율을 비교하고 재방문율은 중간 지표로 봅니다. 재담기와 구매의 관계는 관측 결과이며, 알림·비교 기능의 효과와 1시간·24시간의 적합성은 시험으로 확인합니다.

---

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 5-4. “전략 2 · 네 그룹별 다음 행동을 만든다”

> **근거** — 첫 구매 고객의 **12.93%**만 30일 안에 다시 샀습니다. 구매일 수별 다음 달 구매율은 1번 **7.90%** · 2번 **20.23%** · 3번 이상 **45.71%**로, 처음 산 고객과 반복 구매한 고객은 다르게 관리해야 합니다.

### 신규 · 21,348명 / 6.7% (고객 / 구매금액 비중)

| | |
|---|---|
| 특성 | 최근에 처음이자 한 번만, 소액(주문당 19.95)으로 산 고객. 방문 3회·조회 8회 안팎으로 필요한 것만 당일에 사고 떠나는 목적형 |
| 실행 | 구매 직후 관련 정보를 주고, 7일째에도 재구매가 없으면 두 번째 구매 혜택 안내 |
| 왜 유의미한가 | 두 번째 구매가 단골로 가는 문입니다. 과거 1번 산 고객의 다음 달 구매율은 7.90%, 2번은 20.23%로 2.5배. 재구매까지 걸린 기간 중앙값 12.3일 안에 이유를 만들어야 휴면으로 넘어가지 않습니다 |
| 기대 효과 (가정) | 신규의 5%(1,067명)가 주문당 19.95로 한 번 더 산다면 → **매출 +0.34% · 30일 재구매 고객 +1,067명** |

### 최상위 장기 휴면 · 6,189명 / 13.66% (고객 / 구매금액 비중)

| | |
|---|---|
| 특성 | 결제액 상위 20% 중 88일 넘게 사지 않은 고객. 전체 구매 고객의 5.6%지만 과거 구매금액의 13.66%. 63.8%는 하루만 산 고객, 36.2%는 반복 구매 이력 |
| 실행 | 1회 구매자는 관련 상품으로 새 구매 이유 제시, 반복 구매자는 이전 상품 재구매 경로 제공. 복귀 혜택을 함께 시험 |
| 왜 유의미한가 | 이미 크게 산 경험이 있어 새 고객보다 복귀 가능성과 기여가 큽니다. 1회·반복 구매자의 필요가 달라 메시지를 나눕니다 |
| 기대 효과 (가정) | 5%(309명)가 주문당 41.13(전체 평균)으로 돌아와 한 번 산다면 → **매출 +0.20%** |

### 성장 가능 · 15,989명 / 20.9% (고객 / 구매금액 비중)

| | |
|---|---|
| 특성 | 최근에 1~2번 산 고객 중 주문 금액이 가장 큰 고객(주문당 57.93). 88%가 이미 매출 상위 A등급 |
| 실행 | 1구매일 고객은 두 번째, 2구매일 고객은 세 번째 구매 지원. 마이페이지 혜택과 실제 조건에 맞는 등급 안내 |
| 왜 유의미한가 | 구매 횟수가 늘수록 다음 달 구매율이 오릅니다(2번 20.23% → 3번 이상 45.71%). 다음 구매 한 번이 핵심 고객으로 가는 단계입니다 |
| 기대 효과 (가정) | 5%(799명)가 주문당 57.93으로 다음 구매를 한다면 → **매출 +0.73%** |

### 핵심·충성 · 8,232명 / 약 27.4% (고객 / 구매금액 비중)

| | |
|---|---|
| 특성 | 최근에 여러 번 사는 단골. 방문 18~30회·조회 60~104회, 98~99%가 장바구니에서 빼 보며 비교하는 고객 |
| 실행 | 최근 구매 목록·찜·비교·관심 상품 알림. 샘플 선택·신제품 선공개 등 VIP 혜택 |
| 왜 유의미한가 | 할인 없이도 사는 고객이라 할인은 비용 낭비입니다. 비교하며 고르는 행동에 맞춘 편의가 이탈을 막고 구매 주기를 지킵니다 |
| 기대 효과 (가정) | 5%(412명)가 주문당 47.38로 한 번 더 산다면 → **매출 +0.31%** |

신규·성장 가능·핵심/충성은 RFM 7유형, 최상위 장기 휴면은 결제액 상위 20% × 88일 이상의 별도 기준이라 일부 겹칠 수 있습니다. 구매금액 비중은 과거 기여이며 회수 가능 매출이 아닙니다. 인원은 분석 시점 규모이고, 7일째 혜택 등 시점과 혜택 수준은 실행 제안입니다.

---

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

## 5-5. “기대 효과와 실행 순서, 검증 방법”

| 전략 | 가정 | 결과 (가정 시나리오) |
|---|---|---|
| **전략 1 · 복귀 알림 + 화면 지원** | 방치형 미구매 고객의 5%가 산다면 | **K1 6.74% → 7.37% · 매출 +3.5%** |
| **전략 2 · 신규** | 5%가 두 번째 구매 | **매출 +0.34%** |
| **전략 2 · 최상위 장기 휴면** | 5%가 돌아와 한 번 구매 | **매출 +0.20%** |
| **전략 2 · 성장 가능** | 5%가 다음 구매 | **매출 +0.73%** |
| **전략 2 · 핵심·충성** | 5%가 한 번 더 구매 | **매출 +0.31%** |
| **전략 2 합계** | 네 그룹 각 5% | **매출 약 +1.6%** |

매출 기준 5개월 6,348,268. 그룹별 주문당 금액(신규 19.95 · 성장 가능 57.93 · 핵심·충성 47.38, 장기 휴면은 전체 평균 41.13)으로 1회 추가 구매를 가정했습니다. 혜택 비용은 빼지 않았고, 인과가 확인된 값이 아닙니다.

### “실행 순서”

| 단계 | 실행 |
|---|---|
| **첫 시험** | 복귀 알림(전략 1-①), 신규의 두 번째 구매 혜택 |
| **후속** | 돌아온 뒤 화면 지원(전략 1-②), 최상위 장기 휴면 복귀 |
| **확대** | 성장 가능·핵심·충성 관리 — 효과와 비용을 확인하며 확대 |

실행 순서는 효과가 입증된 순위가 아니라 규모와 실행 난이도를 고려한 순서입니다.

### “비교 실험”

| 실험 | 대상 · 배정 | 주 지표 | 함께 볼 것 |
|---|---|---|---|
| 복귀 알림 시점 | 방치형 미구매 무작위 3그룹: 미발송 / 1시간 / 24시간 | 첫 담기 후 7일 구매율 | 7일 재방문율 · 수신 거부율 |
| 돌아온 뒤 화면 지원 | 재방문 고객 무작위 배정 | 배정 후 7일 구매율 | 비교 기능 사용률 |
| 신규 두 번째 구매 혜택 | 신규 50:50, 7일째 안내 | 첫 구매 후 30일 재구매율 | 혜택 비용 · 고객당 구매금액 |
| 성장·휴면·단골 관리 | 그룹별 무작위 배정 | 배정 후 30일 구매율 | 고객당 구매금액 · 혜택 비용 |

### 통계적 확인 (핵심 비교 · 표본이 커서 차이의 크기로 판단)

| 비교 | 결과 | 판정 |
|---|---|---|
| 조회→담기 저가 vs 고가 (같은 상품) | +5.09%p · 고가는 저가의 0.62배 | 유의 |
| 담기→구매 저가 vs 고가 (같은 상품) | -1.3%p · 고가가 약간 높음 | 가격 효과 없음 |
| 7일 구매율 재담기형 · 일부 삭제형 vs 방치형 | +46.1%p [45.5, 46.7] · +33.0%p [32.7, 33.3] | 유의 |
| 30일 재구매율 첫 주문 상위 20% vs 나머지 | +5.89%p [5.29, 6.49] · 1.5배 | 유의 |
| 재구매까지 시간 상위 20% vs 나머지 (생존 분석) | 위험비 1.54배, 로그순위 $\chi^2$=830 | 빠름 |
| 재구매자끼리 간격 상위 20% vs 나머지 | 12.48일 vs 12.28일, p=0.643 | 차이 확인 안 됨 |
| 30일 재구매율 급등일 vs 그 외 첫 구매 | -1.77%p [-2.37, -1.18] | 유의 |

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)

# Chapter.6 한계점 및 결론

### 한계점

- **인과 아님** — 행동·가격과 구매의 관계는 관찰된 상관. 효과는 A/B 테스트로 확인 필요
- **단위 차이** — Q1은 같은 방문 조합, Q2·Q3는 고객. 같은 방문 미구매의 16.92%는 7일 안 다른 방문에서 구매
- **시간대** — UTC 기준, 판매 지역 현지 시간 미확인
- **관찰 기간** — 10월 이전 기록 없음, 2월 말 고객은 7일·30일 관찰 불가로 제외
- **주문·수량·통화 없음** — 금액은 상한선·범위로 보고
- **이상 고객** — 상위 1%가 놓친 금액의 16.5%. 봇 여부 판단 불가

### 결론

- **담기 전** — 이탈이 가장 큰 곳은 상품을 보고 담는 단계(11.10%). 가격은 이 결정에만 작용 → 고가 상품의 정보 보강·담기 장벽 완화
- **담은 후** — 담은 고객의 77.65%가 7일 안에 미구매, 방치형이 72.6%. 사는 고객은 1시간 안에 삼. 방치형 대부분은 첫 방문에 담고 다시 오지 않음 → 돌아올 계기를 만들고 돌아온 뒤 구매를 돕는다
- **산 후** — 30일 재구매율 12.93%, 첫 달이 고비 → 신규·성장 가능의 다음 구매, 핵심·충성 유지, 최상위 장기 휴면 복귀

## “두 전략과 기대 효과 (가정)”

| 전략 | 대상 | 기대 효과 (가정) |
|---|---|---|
| ① **돌아오게 하고, 구매를 돕는다** | 담고 사지 않은 고객 · 복귀 알림 + 돌아온 뒤 화면 지원 | 방치형 5% 구매 시 K1 6.74% → 7.37%, 매출 +3.5% |
| ② **네 그룹별 다음 행동을 만든다** | 신규 · 최상위 장기 휴면 · 성장 가능 · 핵심·충성 | 각 그룹 5% 추가 구매 시 매출 약 +1.6% |

두 전략은 비교 실험으로 효과를 확인한 뒤 확대하고, 이후 대시보드를 만들어 K1과 단계별 지표를 매달 확인할 계획입니다.

팀 정제 데이터 cosmetics_2019Oct_2020Feb.csv (19,583,742행, 2019-10-01 ~ 2020-02-29, UTC)