



세션 기반 구매 행동 분석 및 전환율 향상 전략 제안

서상우

프로젝트 개요

전자상거래 플랫폼에서는 고객의 구매 여정이 일반적으로 '상품 조회(View)'에서 시작해 '장바구니(Cart)'를 거쳐 '구매(Purchase)'로 이어진다. 그러나 실제 구매까지 이어지는 고객은 전체의 4~7% 수준에 불과하며, 중간 전환 단계에서의 이탈이 매우 빈번하게 발생한다.

따라서 고객이 어느 지점에서 이탈하는지, 어떤 행동이 전환과 연결되는지를 데이터 기반으로 진단하는 것이 필요하다.

(데이터 출처: [eCommerce Behavior Data - Kaggle](#))

➊ 분석 흐름 요약

1. Funnel 전환율 분석

- View → Cart → Purchase 단계별 전환율 계산
- 카테고리 / 시간대 / 가격대 / 요일별 전환율 비교

2. 세션 기반 행동 비교 분석

- 구매 성공 vs 실패 세션 특성 비교
- 세션 길이, 클릭 행동, 평균 가격대 등과 전환율의 관계 분석

3. 전환 예측 모델링

- 로지스틱 회귀모형을 통한 전환 가능성 예측
- 주요 행동 변수의 영향력 정량화

데이터 개요 및 전처리

■ 관측 기간: 2019년 10월 ~ 11월 (2개월)

■ 데이터 크기: 약 2.8억 건 이벤트 로그

컬럼명	설명
event_time	이벤트 발생 시각 (UTC)
event_type	행동 유형 (view, cart, purchase 등)
product_id	상품 고유 ID
category_code	상품 카테고리 (계층형: 예. electronics.smartphone)
brand	브랜드명 (일부 결측 포함)
price	상품 가격
user_id	사용자 고유 ID

■ 전처리 수행 내용 요약:

① 이벤트 정제 및 필터링

- price ≤ 0 또는 NULL 제거
- 이벤트는 ('View', 'Cart', 'Purchase')로 한정
- 결측이 있는 category_code, event_time 레코드 제거

② 세션 정의: 동일 사용자가 30분 이상 이벤트 없이 머물렀을 경우 세션 분리

③ 세션 단위 요약 집계

- 집계 기준: user_id, session_id
- 주요 파생 변수

변수명	설명
session_length	세션 지속 시간 (분 단위)
n_view, n_cart, n_purchase	세션 내 행동 횟수
avg_price	탐색한 상품의 평균 가격

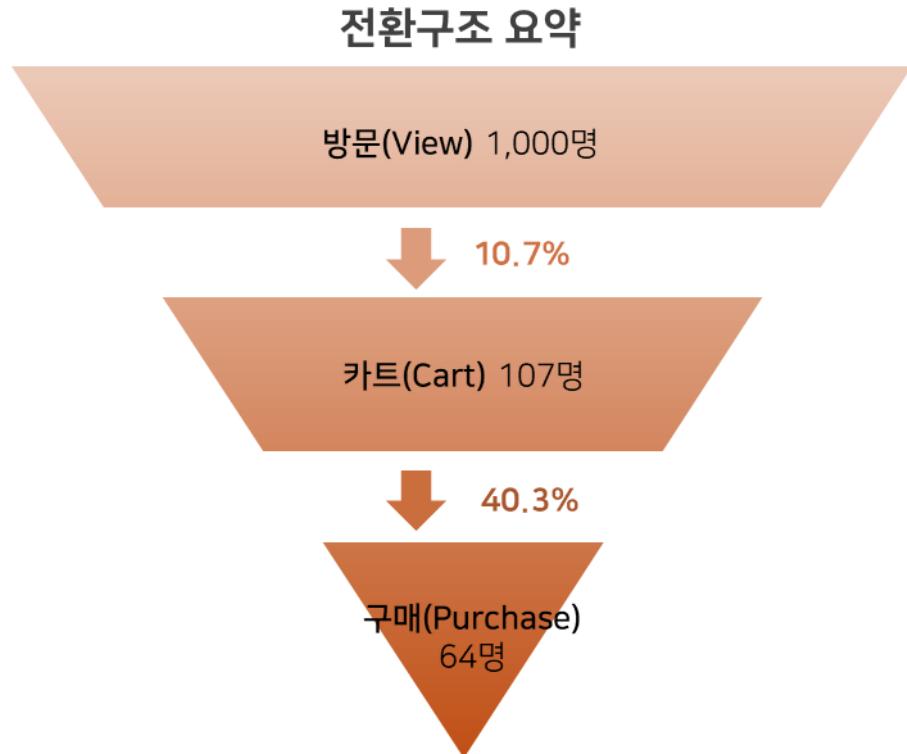
hour_of_day, day_of_week	세션 시작 시각, 요일
category_diversity	카테고리 수
main_category, sub_category	세션의 주요 카테고리 (계층형 파싱)
converted	구매 여부 ($\text{purchase} \geq 1 \rightarrow 1$)

1. Funnel 전환율 분석

- 전자 상거래 플랫폼의 고객 행동은 일반적으로 다음 세 가지 주요 단계로 구성됨:
 - 상품 조회(View) → 장바구니 담기(Cart) → 구매(Purchase)
- 이 분석에서는 이 세 가지 이벤트 간 전환 흐름을 세션 단위로 측정하고, 전환율이 낮은 구간과 전환 차이를 야기하는 요인을 규명하는데 중점을 둠

1) 전체 전환 구조

단계	세션당 전환율
View → Cart	10.71%
Cart → Purchase	40.29%
View → Purchase	6.39%



- 전체 View 세션 중 약 6.4%만이 실제 구매로 이어짐
- 가장 큰 이탈은 View → Cart 단계에서 발생하며, 약 90%의 세션이 장바구니에 도달하지 못하고 이탈

2) 카테고리별 전환율

■ 메인 카테고리 전환율

메인 카테고리	전체 세션 수	구매 세션 수	구매 전환율
appliances	1,794,224	123,007	6.9%
electronics	2,004,919	131,701	6.6%
computers	841,180	41,988	5.0%

- Appliances 및 Electronics (생활/디지털 기기 영역) 은 비교적 높은 구매

전환율을 보이는 반면,

- Computers 카테고리는 가격대가 다소 높고 탐색 시간이 길어 전환율이 상대적으로 낮은것으로 추정됨

■ 서브 카테고리 전환율

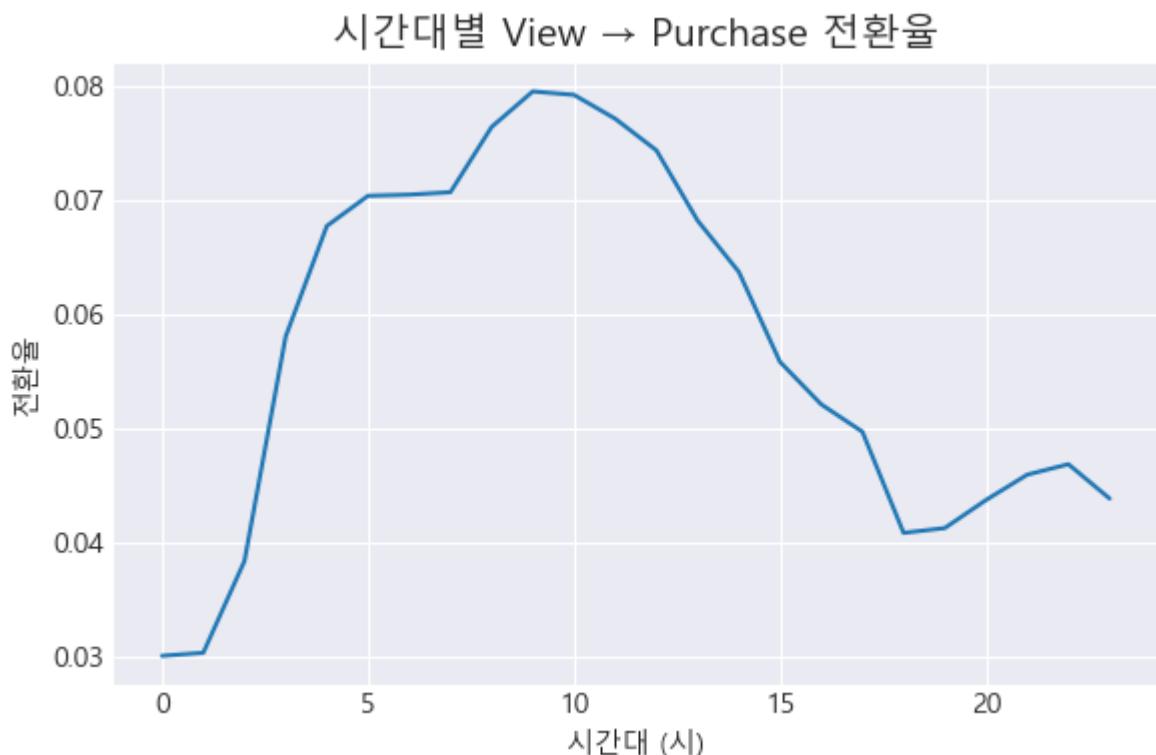
서브 카테고리	전체 세션 수	구매 세션 수	구매 전환율
iron	63,049	5,723	9.1%
ironing_board	12,049	940	7.8%
personal	54,952	4,257	7.7%
environment	433,206	32,059	7.4%
video	475,095	35,045	7.4%
telephone	75,573	5,538	7.3%
audio	747,005	54,307	7.3%
kitchen	1,169,980	76,532	6.5%
tablet	126,377	7,720	6.1%
notebook	398,692	23,140	5.8%
sewing_machine	60,988	3,496	5.7%
peripherals	144,520	7,818	5.4%
clocks	551,362	28,316	5.1%
ebooks	13,307	549	4.1%
components	139,391	5,440	3.9%
desktop	145,270	5,041	3.5%
camera	29,507	775	2.6%

- Appliances 및 Electronics (생활/디지털 기기 영역) 은 비교적 높은 구매 전환율을 보이는 반면,
- Computers 카테고리는 가격대가 다소 높고 탐색 시간이 길어 전환율이 상대적으로 낮은것으로 추정됨

전략적 제안

- 생활가전 및 개인 디지털 기기 제품 등 저관여 제품은 전환율이 높아 ROI 높은 마케팅 후보군으로 활용 가능
- 컴퓨터, 카메라, 구성 부품 등 고관여 제품은 탐색 대비 구매 적음
→ 상세페이지 개선, 비교 콘텐츠, 후기/리뷰 강조 필요
- 세부 카테고리별 Funnel 성과 모니터링을 통한 운영 전략 정교화 필요

3) 시간대별 전환율



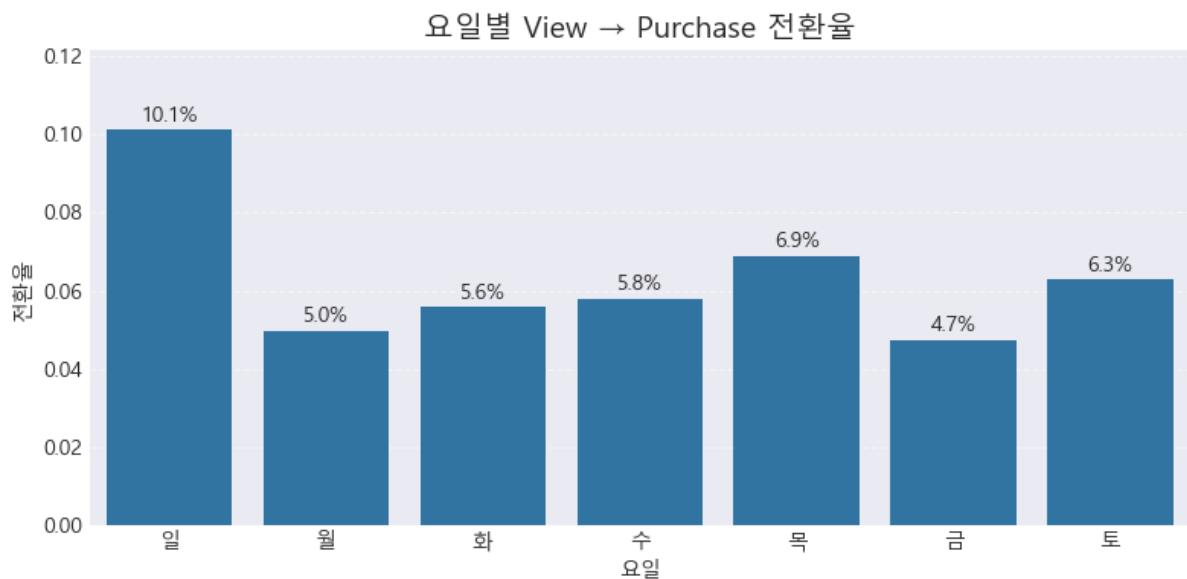
- 고객의 구매 전환율은 시간대에 따라 뚜렷한 차이를 보임
- 주요 결과 요약
 - 전환율 최고 구간: 오전 9시 ~ 11시
 - 전환율 하락 구간: 오전 11시 이후 저녁 시간대까지 점진적 하락
 - 이후 야간 시간대에는 전환율이 저조한 상태를 유지하며, 오전 시간대부터 전환율 상승

- ✓ 전환율의 흐름은 “오전에 집중, 오후에 감소, 야간 침체” 패턴을 보임

💡 전략적 제안

- 고객의 실제 구매 전환은 이른 시간대에 집중 됨
→ 마케팅 집중 타이밍 / 리마인드 캠페인 적기
- 야간 시간대 트래픽 대비 전환율 저조
→ 리타겟팅보다 탐색 유도 중심 콘텐츠 전환 필요

4) 요일별 전환율



- 일요일이 전환율 최고 요일로, 전체 평균 대비 약 1.5~2배 가량 높은 전환율 기록
- 토요일과 목요일도 상대적으로 높은 전환율을 보여 주말과 주중 후반의 구매 집중 경향
- 반면, 금요일은 전환율이 가장 낮은 요일로, 탐색은 많지만 실제 전환은 덜 발생하는 흐름

- 평일 초반(월~수)은 전환율이 낮은 편이나 점차 회복되는 경향

전략적 제안

- 일요일은 메인 프로모션 집중, 알림 송출 최적
- 목, 토요일은 주말 직전/초기 마케팅 타이밍으로 활용
- 금요일은 탐색을 유도하는 콘텐츠 운영, 전환 보다는 리마케팅 기회 확보

5) 가격대별 전환율

가격대	전체 세션 수	구매 세션 수	구매 전환율
~\$50	591,020	43,829	7.4%
\$50~\$100	595,720	40,587	6.8%
\$100~\$200	1,063,957	79,432	7.5%
\$200~\$500	1,590,291	101,005	6.4%
\$500~	799,335	31,843	4.0%

- \$100~\$200 구간이 전환율 최고치(7.47%)로, 구매력이 있으면서도 비교적 부담이 낮은 상품군으로 해석됨
- \$500 이상 고가 제품은 전환율이 4% 미만으로 크게 하락
→ 장바구니 이탈/보류 고객 다수 존재 예상
- 저가 제품군(~\$50)도 높은 전환율(7.4%) 기록 → 소액 반복구매 가능성이 큼

전략적 제안

- ~\$100 이하 제품군은 충동구매 및 즉시 구매 전환율이 높은 특징을 가져 마감 임박 또는 한정 수량을 강조하면 효과적일 것으로 예상
- \$100~\$200 구간은 전환율 최고 구간으로 가성비 중심 기획 상품, 프로모션 집중 후보로 적합

- \$500 이상 고가 제품은 바로 구매하기 보다는 반복 탐색 후 구매하는 경향이 높으므로 리뷰/비교 콘텐츠 제공 및 신뢰도 강화 필요

2. 세션 기반 행동 분석

1) 분석 목표 및 분석 방법

- **분석 목표:** 세션 단위로 고객의 행동 패턴을 분석함으로써, 구매 전환 성공 여부에 영향을 미치는 행동 특성을 도출하는 것을 목표로 함
- **분석 방법**
 - 비교 기준: 구매 전환된 세션 vs 구매 미전환 세션
 - 분석 대상 변수
 - session_length: 세션 길이(분)
 - n_view: 탐색 물품 수
 - n_cart: 장바구니에 담은 물품 수
 - unique_brands: 탐색한 브랜드 수
 - avg_price: 탐색 물품의 평균 가격

2) 분석 결과

변수	구매 세션 평균	미구매 세션 평균	주요 해석
session_length	14.6 분	6.0 분	전환 세션은 약 2.4 배 길

n_view	8.6 회	5.4 회	탐색량이 많을수록 구매로 이어짐
n_cart	1.4 회	0.1 회	구매전환에는 장바구니 행동이 핵심
unique_brands	2.2 개	2.0 개	큰 차이 없음
avg_price	\$ 248.7	\$ 316.5	고가 제품 탐색일수록 전환 저하 가능성

- 세션이 길고 탐색량이 많은 세션일수록 전환율이 높음
 - 단기/단순 탐색 세션은 대부분 구매로 전환되지 않음
- 장바구니 행동 여부가 구매 전환의 핵심 분기점
- 고가 제품 탐색 시 전환율이 낮음

전략적 제안

- 상품 리스트 내 '장바구니' 버튼의 시각적 강조 및 접근성 개선
 - 고객이 탐색 중 바로 장바구니에 담을 수 있도록 UI/UX 개선 + 유도 문구
- 최근 본 제품에 대한 장바구니 담기 리마인드 알림 활용
 - 세션 종료 시점 또는 이탈 직전 타이밍에 알림 노출
- 고가 제품 탐색 후 이탈률이 높은 제품군에 대해
 - 구매 망설임 해소를 위한 리뷰/비교 콘텐츠, 보증 및 혜택 정보 적극 제공

3. 전환 예측 모델링

1) 분석 목표

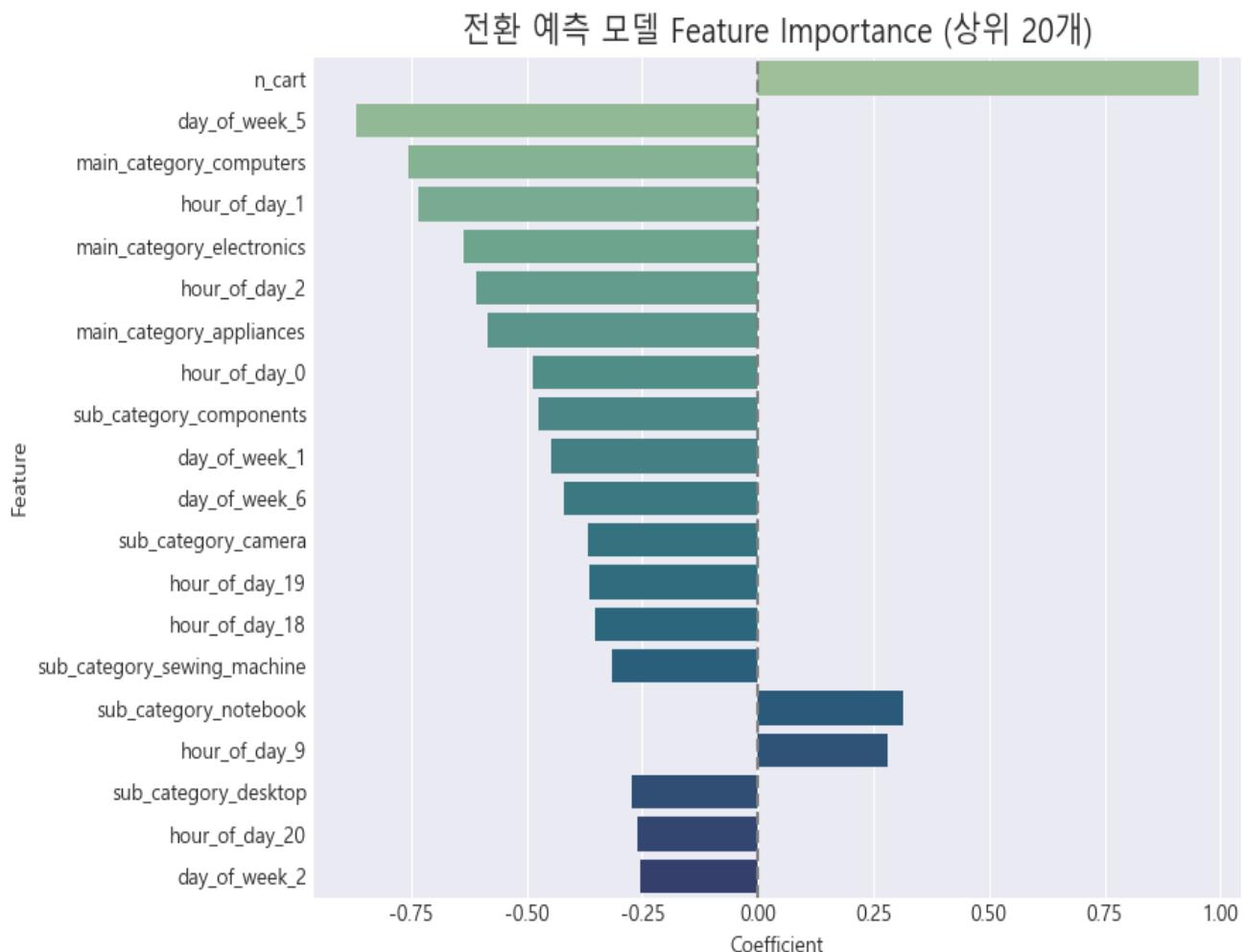
- 세션 단위 데이터를 활용하여, 해당 세션이 구매로 전환될 확률을 예측하는 분류 모델 구축

- 이를 통해, 주요 행동 변수들의 전환 영향력을 정량화하고, 전략 설계 기준으로 활용

2) 모델 설계 계획

항목	내용
모델 종류	Logistic Regression → 해석 중심, 기본 기준선 모델로 적합
타겟 변수	converted (1 = 구매 세션, 0 = 미구매 세션)
입력 변수	<p>session_length 세션 길이</p> <p>n_view 탐색 물품 수</p> <p>n_cart 장바구니에 담은 물품 수</p> <p>avg_price 탐색 물품의 평균 가격</p> <p>hour_of_day 세션 시작 시간대</p> <p>day_of_week 요일</p> <p>main_category 물품의 주요 카테고리</p> <p>sub_category 물품의 서브 카테고리</p>
평가 지표	Accuracy, ROC AUC
해석 포인트	각 행동 변수의 계수 해석 → 전환 기여도 정량 분석

2) 모델링 결과 요약



변수	Coefficient
<code>n_cart</code>	0.95
<code>day_of_week_5</code>	-0.87
<code>main_category_computers</code>	-0.76
<code>hour_of_day_1</code>	-0.73
<code>main_category_electronics</code>	-0.64
<code>hour_of_day_2</code>	-0.61
<code>main_category_appliances</code>	-0.58
<code>hour_of_day_0</code>	-0.49
<code>sub_category_components</code>	-0.47

day_of_week_1	-0.45
day_of_week_6	-0.42
sub_category_camera	-0.37
hour_of_day_19	-0.36
hour_of_day_18	-0.35
sub_category_sewing_machine	-0.32
sub_category_notebook	0.32
hour_of_day_9	0.28
sub_category_desktop	-0.27
hour_of_day_20	-0.26
day_of_week_2	-0.25
hour_of_day_10	0.25
hour_of_day_8	0.23
hour_of_day_11	0.22
day_of_week_3	-0.2
hour_of_day_12	0.19
sub_category_clocks	-0.18
sub_category_ebooks	-0.18
day_of_week_0	0.17
hour_of_day_17	-0.17
hour_of_day_16	-0.17
hour_of_day_21	-0.16
sub_category_kitchen	-0.15
sub_category_peripherals	-0.14
sub_category_environment	-0.13
sub_category_video	0.12
hour_of_day_7	0.12
hour_of_day_6	0.11
sub_category_telephone	-0.11

hour_of_day_13	0.11
hour_of_day_15	-0.1
sub_category_iron	0.1
hour_of_day_5	0.1
hour_of_day_23	-0.1
hour_of_day_22	-0.1
sub_category_tablet	-0.09
hour_of_day_3	-0.07
hour_of_day_4	0.06
sub_category_ironing_board	-0.05
sub_category_personal	-0.05
day_of_week_4	0.04
hour_of_day_14	0.03
n_view	-0.03
session_length	0.03
sub_category_audio	-0.01
avg_price	0

■ 전환 확률을 높이는 요인

- '장바구니 담기'는 전환 확률을 가장 강하게 끌어올리는 변수
→ 장바구니 행동이 있는 세션은 구매 가능성이 매우 높음
- 오전 시간대 (9~11시)의 세션은 전환 확률이 유의미하게 높음
→ 고객의 집중력과 결제 의향이 높은 시간대
- 일요일 세션은 다른曜일에 비해 전환율이 높음
- 노트북, 영상기기, iron(다리미) 관련 제품은 전환 확률을 끌어올리는 경향

■ 전환 확률을 낮추는 요인

- 금요일은 전환율이 가장 낮은曜일
→ 탐색은 활발하지만 구매는 유보되는 패턴

- 심야 시간대는 전반적으로 전환율이 낮음
→ 피로도, 결제 망설임 등 심리적 장벽 존재
- 전자제품, 컴퓨터, 카메라 제품은 전환율이 낮음
→ 가격 부담과 정보 탐색 소요 시간이 많아 전환 유보

■ 전환에 유의한 영향을 끼치지 않은 변수들

- **제품 가격**: 고가 제품 탐색이 전환에 영향을 미칠 것으로 예상했으나, 실제 데이터에서는 직접적인 영향력 없음
- **세션 길이**: 일부 전환 분석에서는 의미 있었으나, 다변량 회귀모델에서는 유의하지 않음
→ 세션이 길다고 반드시 구매로 이어지지 않음
- **탐색 제품 수**: 전환율에 뚜렷한 영향 없음

■ (참고) 모델 성능

클래스 (converted)	Precision	Recall	F1-score	Support
0 (미구매)	0.94	0.99	0.97	868,695
1 (구매)	0.54	0.13	0.2	59,370
Accuracy		0.94		
ROC AUC		0.84		

- 전체 정확도(Accuracy)는 94%로 높지만, **클래스 불균형**으로 인해 Recall은 13%로 낮음
- AUC는 0.84로 양호한 분류 성능을 보였으며, 후속 개선 여지가 있음
- Precision(0.54)은 괜찮지만, Recall을 개선하려면 언더/오버샘플링, 가중치 부여, 모델 변경 등의 방법 고려 가능

마무리: 종합 결론 및 제언

결론 요약

- 본 분석은 실제 eCommerce 고객 세션 로그를 바탕으로, Funnel 구조 분석 → 행동 패턴 비교 → 전환 예측 모델링까지 일관된 흐름으로 수행되었음
- 전체 전환율은 6.39%로 낮은 수준이며, 전환은 특정 시간·요일·제품군·행동 조합에 집중되는 경향을 보였음
- 특히 장바구니 담기 행동이 구매 전환의 핵심 분기점으로 확인되었고, 시간대/카테고리/가격대에 따른 패턴 차이도 유의미했음
- 로지스틱 회귀 분석을 통해 유의한 변수와 무의미한 변수 간의 경계가 명확히 정리되었으며, 전략적 해석 가능성이 확보되었음

전략적 제언

- Funnel 전환 흐름 상 가장 큰 이탈 구간(View → Cart)에 집중 대응 전략 필요
→ 상품 리스트 내 CTA 강화, 장바구니 유도형 UX/UI 개선
- 전환율이 높은 시간대(오전)와 요일(일요일)에 마케팅 예산 집중
→ 프로모션 타이밍, 광고 노출 최적화
- \$100~\$200 증가 제품군을 중심으로 신규 고객 유입 전략 설계
→ 대표상품 구성, 세트 번들링, 리뷰 강조 마케팅 활용
- 고가 제품군은 상세정보·비교 콘텐츠·장바구니 기반 리마케팅 구조 적용
→ 탐색은 활발하나 전환율 낮은 특성 반영
- 단순 탐색량(n_view)이나 세션 길이(session_length)는 전환율에 큰 영향 없음
→ 의미 있는 행동(장바구니, 특정 시간, 브랜드 집중 탐색)을 유도하는 전략이 효과적