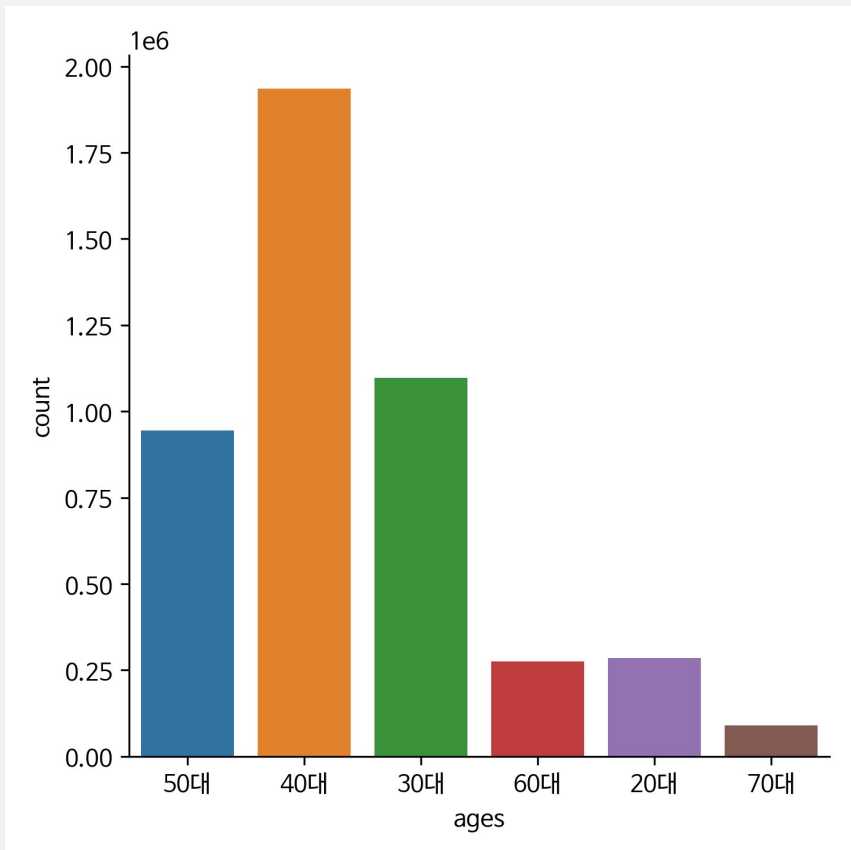


# 롯데 예측모델을 통한 개인화 마케팅(가제)

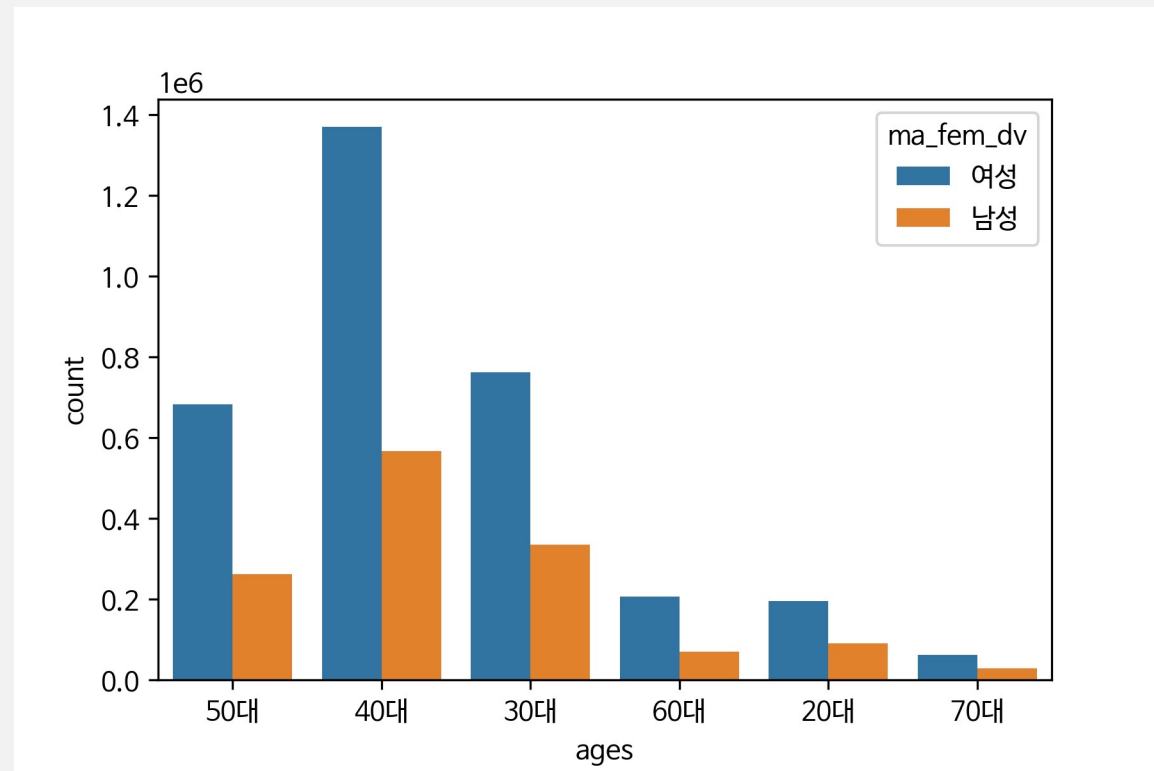
팀이름

- EDA 및 문제 정의
  - EDA 결과를 통해 문제 정의
  - Data cleaning 과정 대략적 설명 (파생변수 설명 포함)
- 모델 개발
  - 모델 설명
  - 성능 검증 과정
- 개인화 마케팅 전략
  - 분석결과 기반 모델 활용 방안(논리적 연계성, 실행 가능성, 독창성 중요)

## EDA 및 문제 정의 : 성별, 연령별 구매 현황

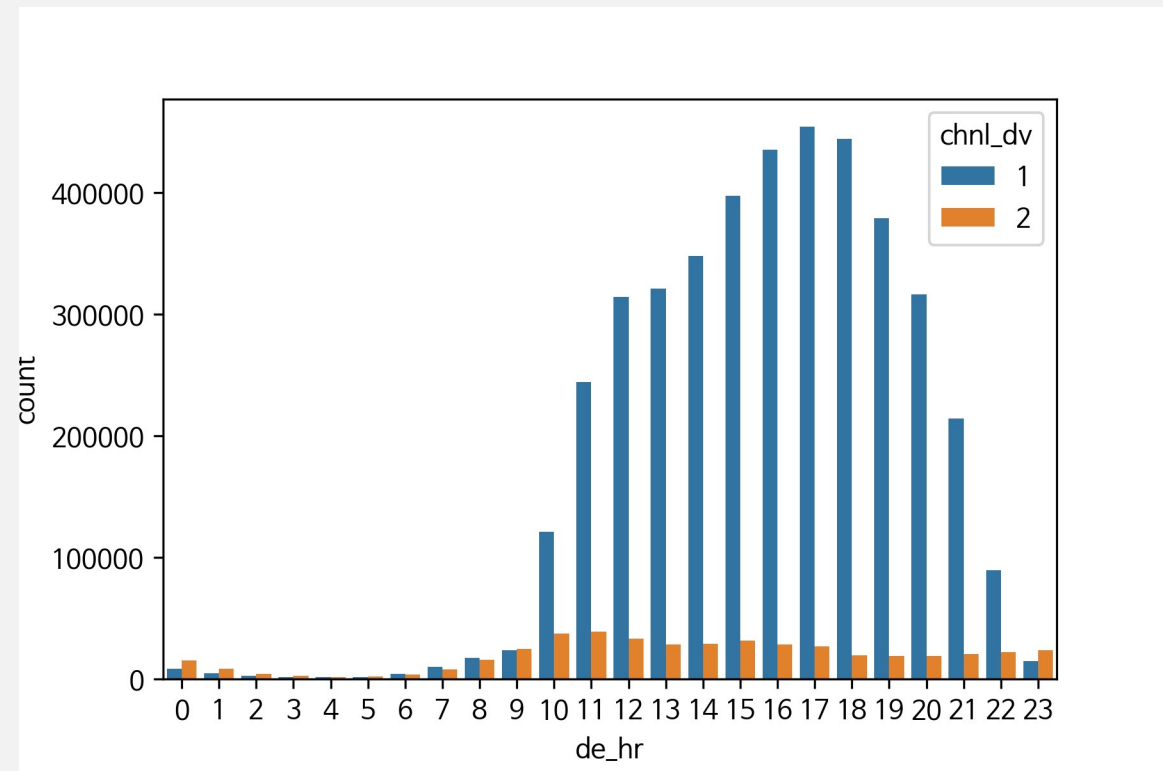
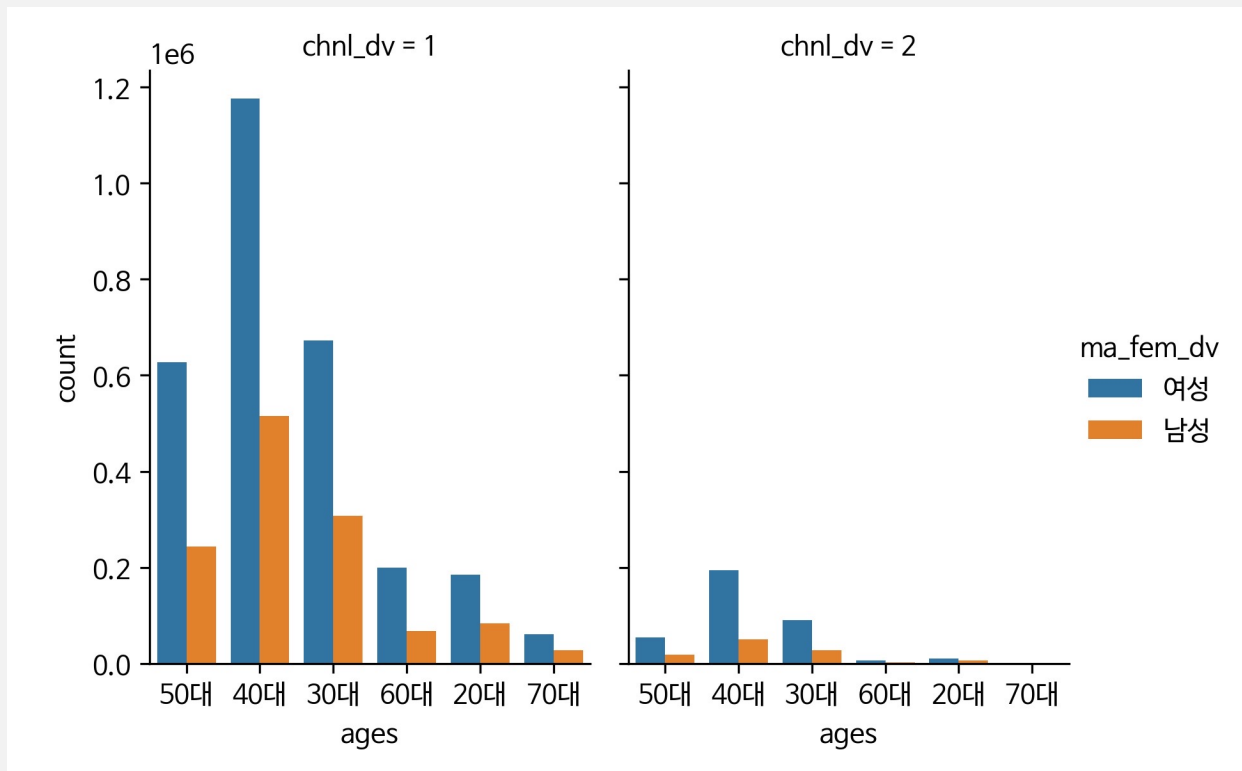


- 구매자 연령 비율
  - 40대, 30대, 50대 순으로 비율이 큼



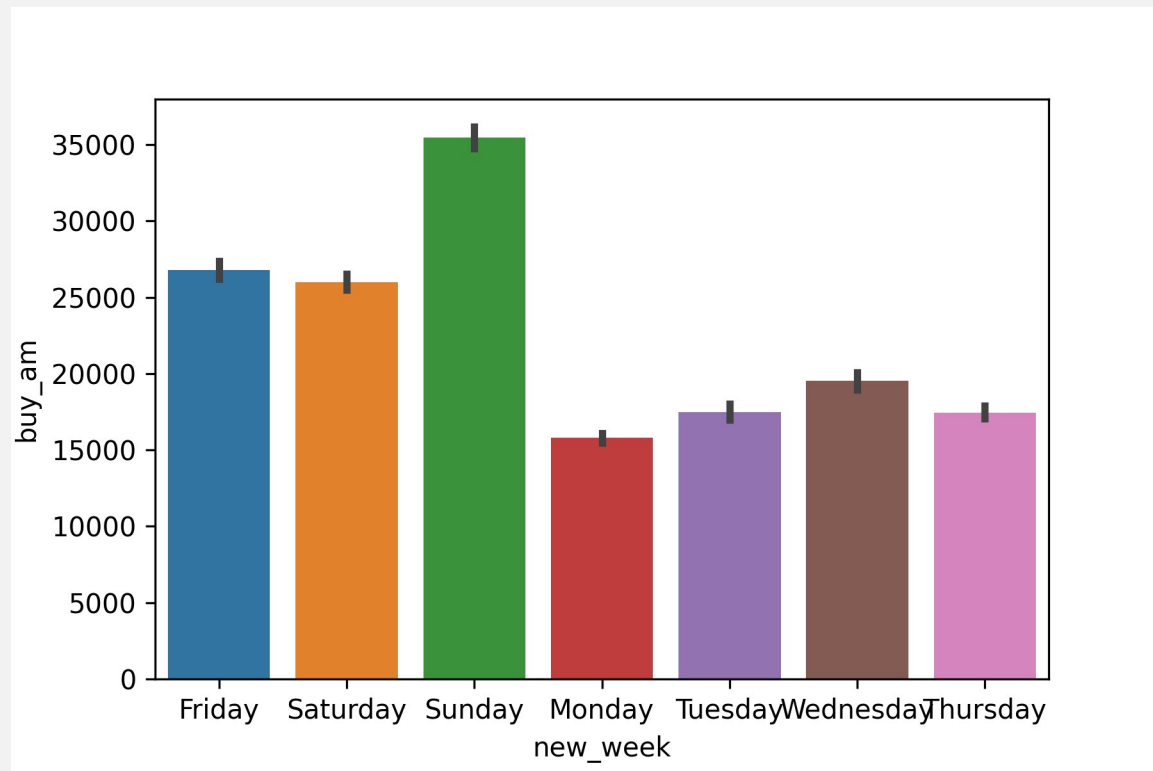
- 연령대별 성별 비율
  - 모든 성별에서 여성고객 비율이 남성고객의 2배 이상

## EDA 및 문제 정의 : 구매채널, 시간대별 구매 현황

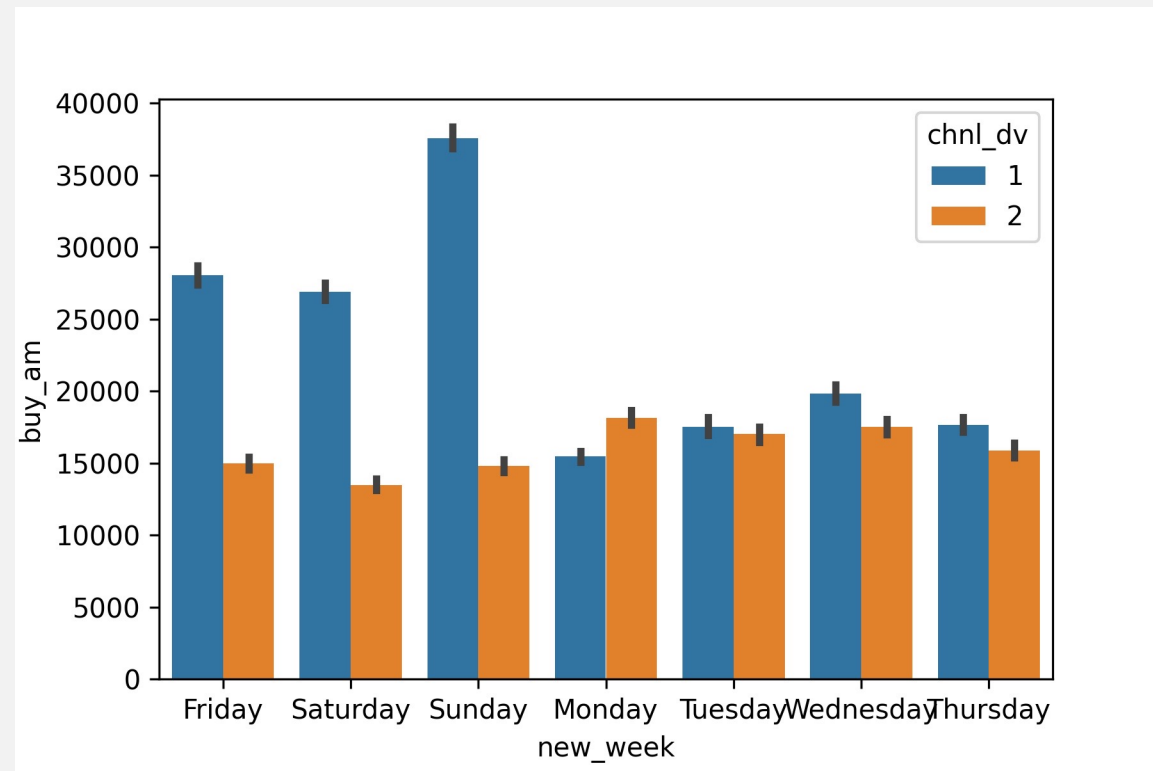


- 연령, 성별에 따른 구매채널
  - 모든 연령, 성별이 대부분 오프라인 매장에서 구매

- 시간대별 온라인/오프라인 구매 현황

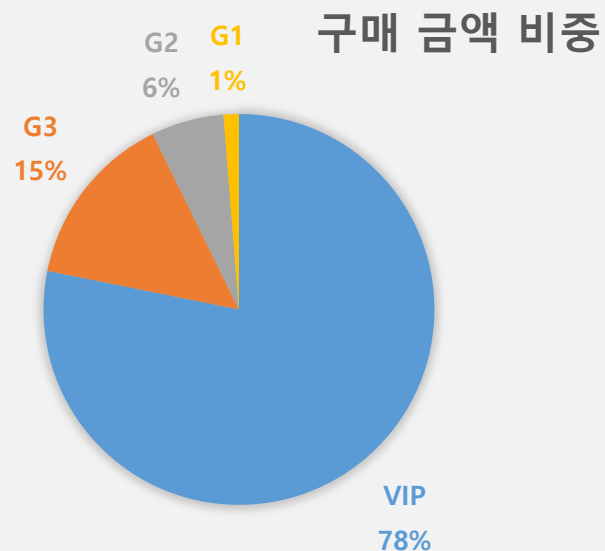


요일별 구매금액



- 요일별 온라인/오프라인 구매 비율
  - 오프라인 : 금,토,일 주말에 구매가 집중
  - 온라인 : 요일별 편차가 크지 않음

# EDA 및 문제 정의 : 온라인 채널 구매 특징



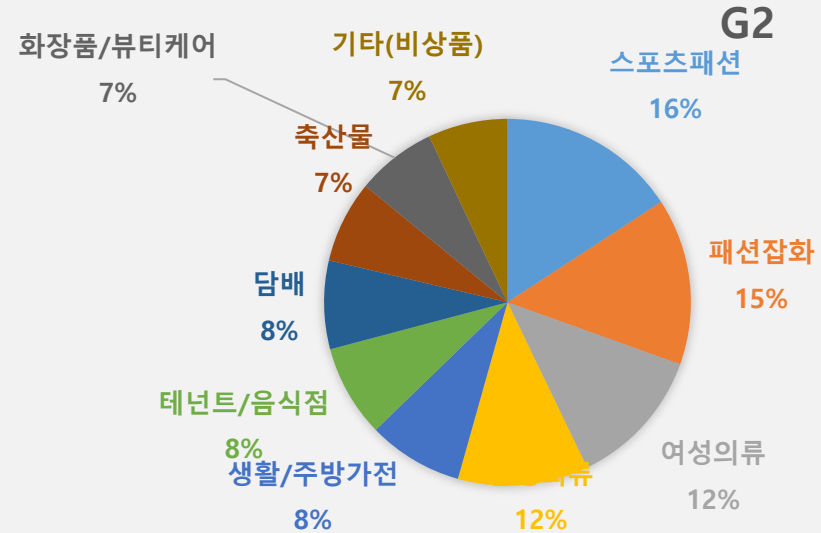
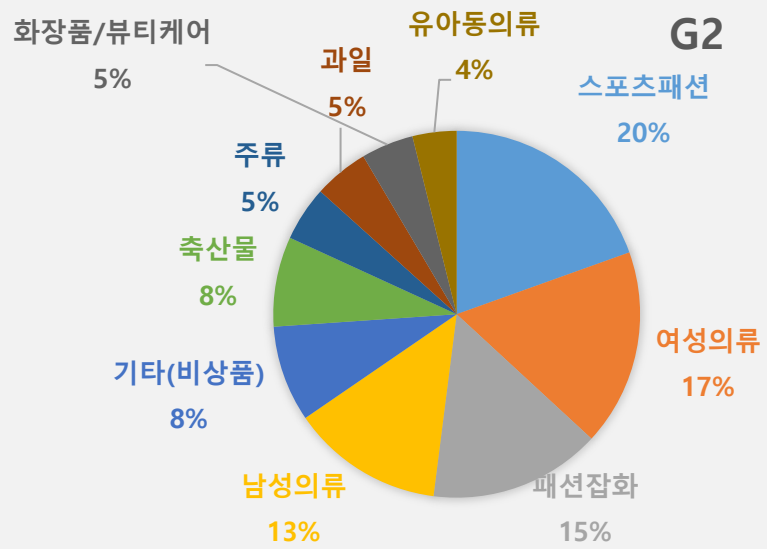
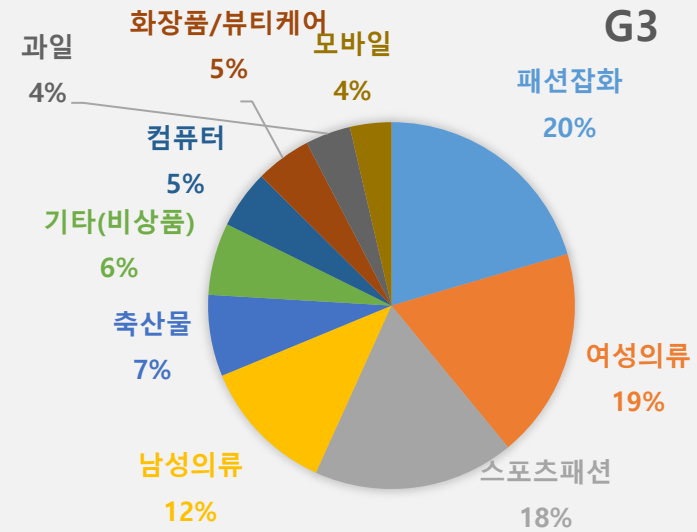
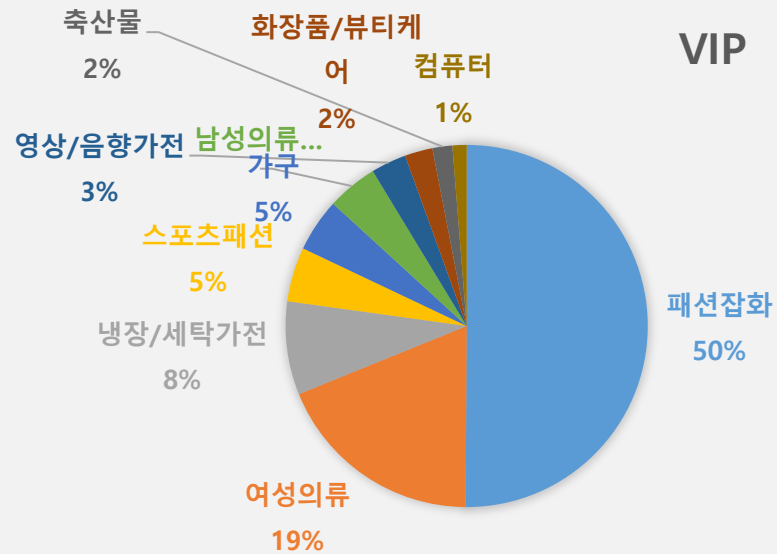
구매 등급	온라인 연간 구매 총액	총액 비율
VIP	79,499,526,508	78%
G3	14,811,185,876	15%
G2	6,104,430,211	6%
G1	1,290,925,602	1%
총합계	101,706,068,197	100%

- 구매 금액 순으로 4등급으로 분류
- 상위 25%에 속하는 고객이 78% 매출을 차지

매출 금액 순	VIP	G3	G2	G1
1	패션잡화	패션잡화	스포츠패션	스포츠패션
2	여성의류	여성의류	여성의류	패션잡화
3	냉장/세탁가전	스포츠패션	패션잡화	여성의류
4	스포츠패션	남성의류	남성의류	남성의류
5	가구	축산물	기타(비상품)	생활/주방가전
6	남성의류	기타(비상품)	축산물	테넨트/음식점
7	영상/음향가전	컴퓨터	주류	담배
8	화장품/뷰티케어	화장품/뷰티케어	과일	축산물
9	축산물	과일	화장품/뷰티케어	화장품/뷰티케어
10	컴퓨터	모바일	유아동의류	기타(비상품)
11	생활/주방가전	주류	테넨트/음식점	주류
12	계절가전	유아동의류	생활/주방가전	과일
13	기타(비상품)	완구	완구	모바일
14	구기/필드스포츠	상품권	모바일	과자
15	건강용품	테넨트/음식점	담배	완구

- 등급별 높은 매출을 보이는 카테고리 항목이 다름
- 등급별로 적절한 마케팅 전략을 세우는 것이 중요

## EDA 및 문제 정의 : 온라인 채널 구매 특징



- 등급별 구매 금액 상위 10개 카테고리 비중

## 예측모델 : 행렬 분해를 활용한 추천 시스템

Cust_2	가구	건강식품	건강용품	건강산물	계절가전	공구/안전용품	과일	과자	구기/필드스포츠	...	침구/수예
M343264688	18541000.0	3978000.0	238000.0	8980.0	0.0	54700.0	794380.0	80350.0	1871000.0	...	0.0
M384121563	0.0	0.0	59960.0	0.0	0.0	0.0	0.0	109300.0	605000.0	...	990000.0
M268374518	0.0	281000.0	12390.0	0.0	0.0	0.0	104400.0	13130.0	0.0	...	0.0
M055769504	0.0	17400.0	3000.0	0.0	0.0	0.0	0.0	10500.0	0.0	...	0.0
M288629527	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	8052000.0
...	...	...	...	...	...	...	...	...	...	...	...
M549598435	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
M698192014	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0

- 추천 상품을 띄울 수 있는 온라인 판매 특성을 이용 : 추천 시스템으로 개인화 마케팅 전략 수립 가능
- 카테고리별로 고객이 연간 구매한 금액을 모두 합친 데이터를 활용
  - Sparse matrix : 모든 카테고리를 구매하지 않아서, 구매 금액이 없는 경우가 많음
  - 총 구매 고객이 26917명, 대분류 카테고리를 사용하여 58개 변수 사용
  - 고객 데이터가 추가로 있는 경우, 더 세분화된 카테고리 및 상품 분류 적용 가능
- 고차원 행렬을 저차원 행렬로 분해해 구매가 없는 값을 추정



- $R = PQ^t$  행렬분해 활용

$$\begin{aligned} R &= PQ^t \\ \hat{R} &\approx PQ^t \end{aligned}$$

- 오차제곱을 최소로 하는  $\hat{R} = PQ^t$  추정해 잠재적으로 구매할 가능성이 높은 카테고리 선별

$$P, Q^t = \operatorname{argmin} e^2$$

- 과대적합 문제를 줄여 예측력을 높이기 위해 L2 norm 정규화 활용 (Convex Optimization)
  - 경사하강법(Gradient Descent) 알고리즘으로 최적화 솔루션 추정

$$e^2 = (R - \hat{R})^2 + \frac{\beta}{2} (\|P\|_2^2 + \|Q\|_2^2)$$

- 상품별 가격 단위 표준화 위해 MinMax 정규화 적용

```
from sklearn.preprocessing import MinMaxScaler
```

- 대분류 카테고리를 식품류, 패션류, 가전류, 그 외 네 가지로 구분했고, 저차원 행렬 K=4로 활용

$$\widehat{R_{n \times p}} \approx P_{n \times 4} Q_{4 \times p}^t$$

- 경사하강법 최적화 알고리즘으로 추천 행렬  $\hat{R}$  추정 후, 본래 가격 단위로 되돌림

```
class MatrixFactorization():  
    def __init__(self, R, k, learning_rate, reg_param, epochs, verbose=False):  
        ....  
factorizer = MatrixFactorization(R, k=4, learning_rate=0.1, reg_param=0.1, epochs=10, verbose=True)
```

## 예측모델 : 행렬 분해를 활용한 추천 시스템

- 등급별로 나누어 추천시스템 모델 적용 & 개인별로 구매 이력 없는 행렬 원소 추정

등급	Cust	가구	건강용품	...	계절가전	패션잡화	퍼스널케어	헬스/피트니스	화장품/뷰티케어	추천 1순위 카테고리	추천 2순위 카테고리
VIP	M343264688	17,163,898	3,145,548	...	6,527,170	38,503,130	2,902,103	1,165,818	24,574,730	패션잡화	화장품/뷰티케어
	M384121563	16,537,444	3,027,318	...	6,342,832	37,098,090	2,795,617	1,123,926	23,677,841	패션잡화	화장품/뷰티케어
	M268374518	14,503,636	2,653,079	...	5,662,146	32,530,323	2,459,742	1,000,007	20,727,233	패션잡화	화장품/뷰티케어
	M055769504	12,866,029	2,351,696	...	5,113,509	28,853,624	2,189,508	900,855	18,345,727	패션잡화	화장품/뷰티케어
	M288629527	19,626,681	3,603,276	...	7,306,435	44,033,611	3,314,197	1,323,617	28,114,120	패션잡화	화장품/뷰티케어
G3	M860584542	183,000	56,800	...	262,000	234,000	57,000	172,000	191,000	냉장/세탁가전	스포츠패션
	M193883297	336,000	135,000	...	454,000	380,000	134,000	222,000	309,000	냉장/세탁가전	컴퓨터
	M290219002	210,000	70,700	...	310,000	280,000	76,900	171,000	203,000	냉장/세탁가전	컴퓨터
	M600419132	269,000	104,000	...	408,000	321,000	106,000	177,000	227,000	냉장/세탁가전	컴퓨터
	M075635632	181,000	55,700	...	272,000	267,000	66,800	163,000	187,000	냉장/세탁가전	컴퓨터
G2	M342276582	157,747	25,757	...	125,828	160,235	52,429	117,290	109,923	냉장/세탁가전	생활/주방가전
	M139925797	121,054	27,880	...	176,448	152,245	37,942	112,523	109,344	냉장/세탁가전	생활/주방가전
	M368642492	148,661	29,725	...	158,921	154,814	51,420	101,255	97,191	냉장/세탁가전	여성의류
	M142912521	111,439	41,476	...	174,220	176,018	53,929	123,249	125,388	컴퓨터	생활/주방가전
	M255009646	93,857	22,950	...	146,544	140,032	32,108	101,783	91,848	냉장/세탁가전	생활/주방가전
G1	M041756011	95,503	34,227	...	86,324	61,793	40,563	85,899	58,611	냉장/세탁가전	가구
	M077412930	80,867	24,550	...	71,421	44,480	29,495	70,875	42,806	냉장/세탁가전	가구
	M889321303	88,560	29,855	...	79,367	54,143	35,669	78,325	51,411	냉장/세탁가전	가구
	M116331199	83,070	27,092	...	74,219	49,477	33,042	72,765	47,416	냉장/세탁가전	가구
	M648527975	100,367	37,557	...	91,683	66,813	44,120	90,065	64,302	냉장/세탁가전	가구

- 최적화 가장 큰 금액 순으로 개인별 추천 카테고리 2개 선정
  - 고객 데이터가 더 많은 경우, 세부 카테고리 및 상품 분류를 사용해 더 세밀한 추천이 가능

## 개인화 마케팅 전략 : 온라인 채널 등급별 구매 비중

카테고리	VIP	G3	G2	G1
패션잡화	94.4%	4.1%	1.2%	0.3%
여성의류	86.8%	9.2%	3.4%	0.6%
냉장/세탁가전	96.9%	2.6%	0.4%	0.0%
스포츠패션	62.7%	24.4%	10.6%	2.3%
남성의류	69.6%	19.7%	8.7%	2.0%
가구	95.1%	4.1%	0.6%	0.2%
화장품/뷰티케어	75.6%	16.0%	5.9%	2.4%
영상/음향가전	96.9%	1.9%	0.9%	0.2%
축산물	60.0%	26.0%	11.3%	2.7%
기타(비상품)	47.8%	31.8%	16.7%	3.7%
컴퓨터	64.3%	27.8%	6.5%	1.4%
생활/주방가전	69.7%	15.2%	9.4%	5.7%
과일	46.6%	33.8%	16.2%	3.4%
계절가전	91.9%	5.6%	0.9%	1.6%
모바일	51.0%	33.6%	12.0%	3.5%
유아동의류	55.5%	27.8%	14.2%	2.5%
주류	44.7%	30.2%	18.5%	6.7%
구기/필드스포츠	87.4%	8.6%	2.9%	1.1%

VIP : 패션잡화, 여성의류, 가전에서 높은 매출

→ 특정 카테고리 전용관을 만들고,  
신상품 출시 및 이벤트 소식을 알려  
구매를 높일 수 있는 마케팅 전략 필요

G 등급 : 생필품 카테고리에서 매출 비중 증가

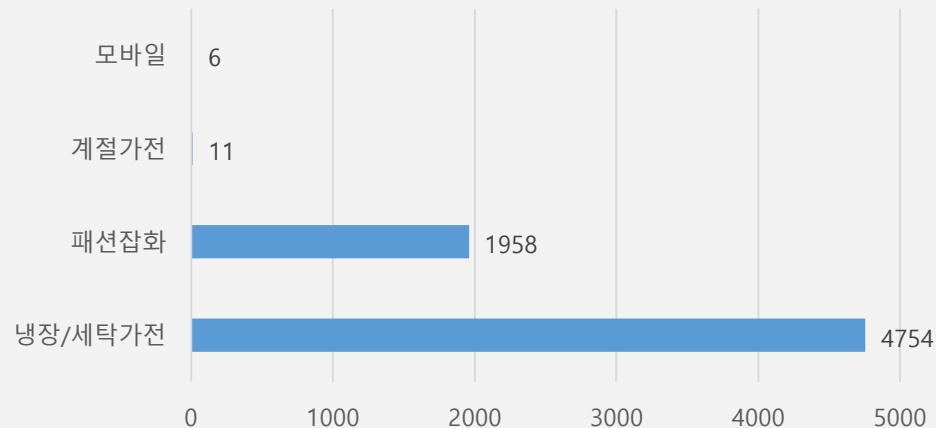
→ 자주 구매하는 생필품 특성상,  
재구매율을 높일 수 있도록 주기적으로 할인쿠폰  
및 이벤트 등을 활용한 마케팅 전략 필요

- 상위 25% 구매 고객이 카테고리 별 매출의 상당 부분을 차지
  - 패션잡화 매출의 경우, VIP 등급이 95%를 차지
  - 생필품 카테고리에서 하위 등급 비중 증가

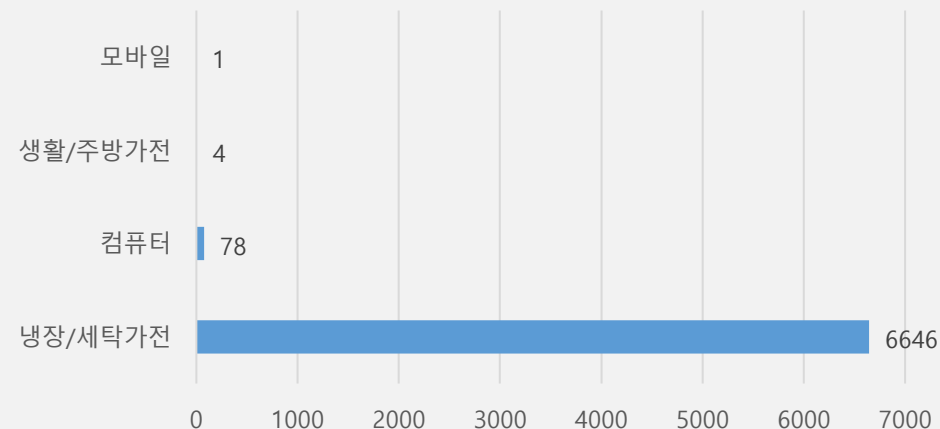
# 개인화 마케팅 전략 : 행렬 분해를 활용한 추천 시스템

- 등급별 추천 카테고리 빈도

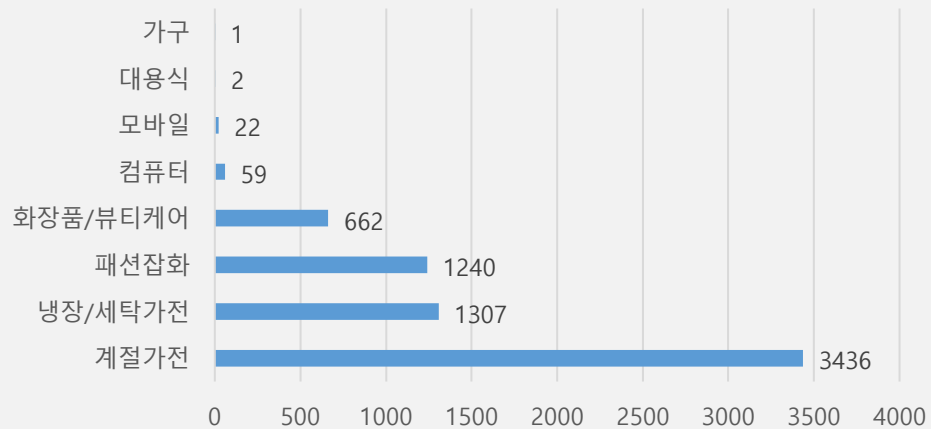
## VIP - 1번째 추천 카테고리



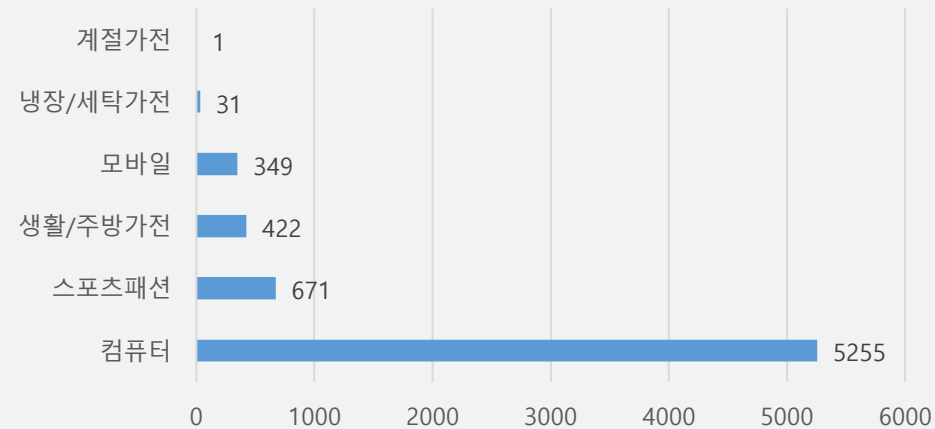
## G3 - 1번째 추천 카테고리



## VIP - 2번째 추천 카테고리



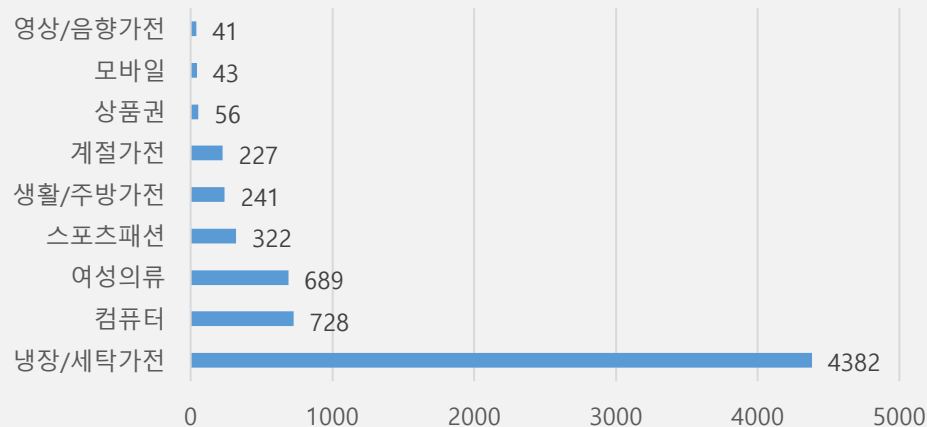
## G3 - 2번째 추천 카테고리



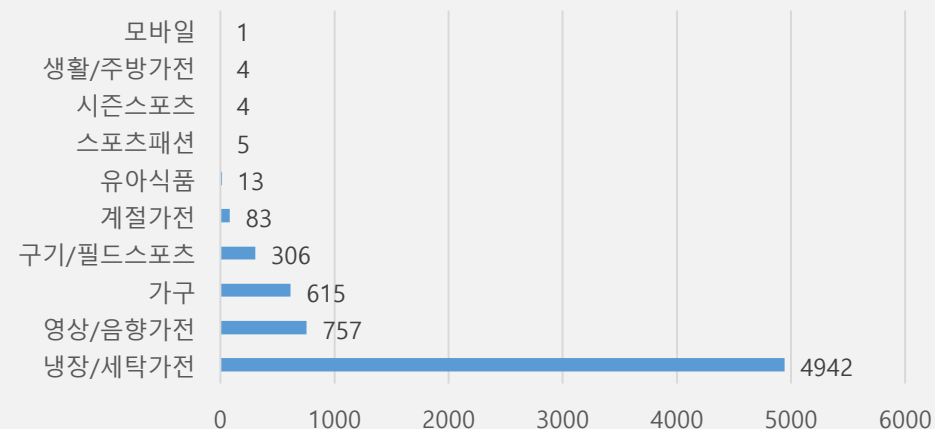
# 개인화 마케팅 전략 : 행렬 분해를 활용한 추천 시스템

- 등급별 추천 카테고리 빈도

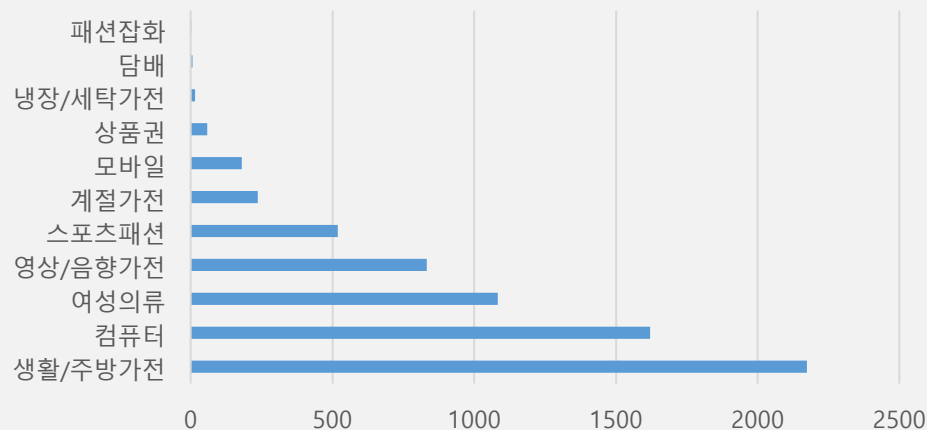
## G2 - 1번째 추천 카테고리



## G1 - 1번째 추천 카테고리



## G2 - 2번째 추천 카테고리



## G1 - 2번째 추천 카테고리

