

DevCourse Final Project

2024년 4월 1일

—

3 팀 2 조

조민수, 한소현, 정연우

개요

길다면 길고 짧다면 짧았던 데브코스 진행. 코스를 수료하기전, 우리는 배웠던 기술들을 잘 활용할 수 있을까? 최종 프로젝트에선 학습했던 분석 기술들을 복습하는 시간을 갖기로 했다. 어떤 분석을 하기위해, 어떤 스킬을, 왜 쓰는지에 대한 생각을 꾸준히 함으로써, 스킬들에 대해 이해하고 팀원들과 인사이트를 공유하면서 더 나은 방법을 탐구했다.

목표

1. 5개월 동안 학습했던 스킬들을 복습해보고, 데이터에 활용시키자.
2. 분석을 진행하면서 활용했던 스킬의 이유와 원리를 파악하자.

데이터 전처리

데이터 : [[이커머스 고객 세분화 분석 아이디어 경진대회 - DAICON](#)]

데이터는 '데이콘'에서 진행한 대회 '이커머스 고객 세분화 분석 아이디어 경진대회' 데이터를 활용했다. 데이터는 5개의 csv파일로 주어졌고, 모든 데이터 파일을 고객 고유 칼럼인 '고객ID'를 기준으로 합병하고 진행했다.

추가 전처리

그대로 진행하기에는 불필요한 데이터가 많아서 1. 필요한 칼럼은 생성하고 2. 불필요한 칼럼은 버리는 추가 전처리도 진행했다.

1. 추가 칼럼

`['구매금액'] = ['수량'] * ['평균금액']`

`['초중말'] = ['거래날짜'] > strftime(%d)을 활용 | 1~10일['초'] / 11~20일['중'] / 20일~['말']`

`['월'] = ['거래날짜'] > strftime(%b)을 활용 | ['Jan', 'Feb', ... 'Dec']`

`['요일'] = ['거래날짜'] > weekday을 활용 | ['월요일', '화요일', ... '일요일']`

`['가입기간'] = ['가입기간']/12 .. (1년 기준)`

2. 불필요한 칼럼

`['거래ID', '제품ID', '쿠폰코드', 'GST']`: 진행하려는 분석과 관련 없는 칼럼

`['거래날짜', '평균금액']`: 필요한 내용 추출 완료

대회의 목적인 '고객 세분화'를 하기 위해 클러스터링을 진행하기로 했다. 이 때 프로젝트의 목적에 맞게 팀원들이 각각 적절한 클러스터링 기법이 무엇인지 조사하고 비교해보기로 했다.

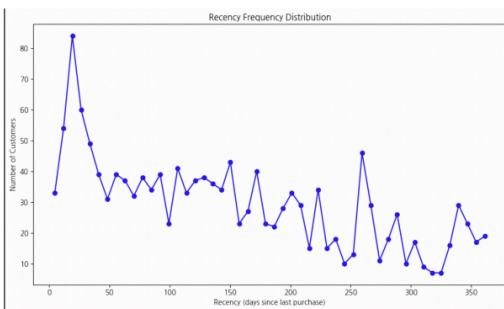
클러스터링

클러스터링을 진행하기 전에 데이터의 칼럼(feature)별로 단위가 다르기 때문에(수량/구매금액) 결과의 왜곡을 막고 추가적으로 모델 성능을 높이기 위해 **Scaling** 작업이 필요했다. 모든 팀원들이 고객 **Sagement**수를 4개로 나누자고 합의를 하고 각자 조사에 들어갔다.

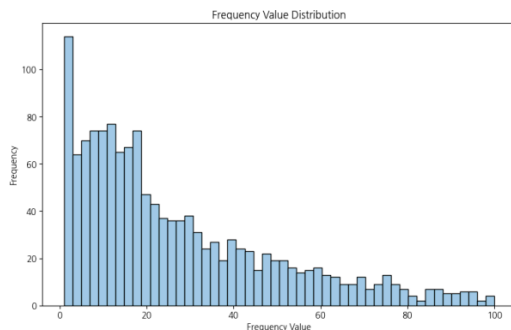
RFM Score

고객 세분화에 주로 사용하는 기법으로 **Recency**(얼마나 최근에), **Frequency**(얼마나 자주), **Monetary**(얼마나 많이)를 비교해서 사용자의 분포와 등급을 분류하는 분석기법이다. 각 컬럼마다 점수를 매겨 **Scaling**을 진행한다.

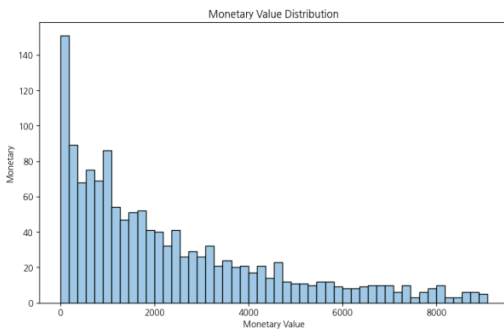
R = 20년1월1일 - 해당 고객이 마지막에 방문한 날짜 (데이터는 19년 데이터만 존재하므로 1년 단위로 분석하기로 했다.)



F = 거래ID의 count값



$M = \text{구매금액} * \text{할인율} * \text{GST의 sum값}$



R은 qcut을 통해 분류를 하기로 했고, F/M의 경우 long-tail 분포를 가졌기 때문에 이상치를 기준으로 나누기로 정했다. (결과적으로 각 컬럼에 1~4점을 매기게 된다)

Log Scaling

전체 데이터가 양수고, 데이터간 크기 차이가 크므로 큰 이상치의 영향력을 줄이기 위해 Log Scaling을 사용해봤다.

Robust Scaling

데이터에 이상치가 많기 때문에 중앙값과 IQR값을 사용해 이상치를 다룰 수 있는 Robust Scaling를 사용하여 이상치의 영향을 줄이려고 했다.

Cox-Box Scaling

마찬가지로 데이터가 모두 양수이므로 정규분포에 가깝게 만들어 추후 클러스터링 모델의 성능을 높이기 위해 Cox-Box Scaling을 사용했다.

각 Scaling 방법을 Loss Function을 통해 실제값과 얼마나 다른지를 확인해보아 가장 Loss가 적은 Scaling 방법은 **RFM** 방법이었다 (**779**)

이후 Elbow Method를 통해 적절한 클러스터 개수를 확인해보니 4개가 나왔다. 다행히 초반에 정해놓은 Sagement갯수와 일치해서 RFM Score을 다시 매길 필요는 없었고, cut의 기준을 조정하여 칼럼값의 가중치를 높여 분류가 정확하게 되도록 했다.

Scaling 방법이 정해진 뒤 이젠 클러스터링을 위해 모델을 선택하기로 했다. 3가지 클러스터링 모델 K-Means / Gaussian Mixture Model / DBSCAN 을 가지고 팀원들끼리 다시 조사를 진행했다. 각각의 특성은 다음과 같다

K-Means

가장 일반적인 클러스터링 방법으로, 간단하고 직관적이어서 결과를 해석하는데 문제가 없다. 대규모, 다양한 데이터셋에 적용시킬 수 있으며 빠른 속도를 자랑한다.

GMM

가우시안 분포(정규분포)를 가진 데이터들을 바탕으로 진행하는 모델로 다양한 형태의 데이터셋에 적용할 수 있는 유연한 특징을 가진다. 이상치 탐지에도 적절하고 소프트 클러스터링을 통해 데이터가 여러 클러스터에 속할 가능성을 고려한다.

DBSCAN

특정 공간 내 데이터 밀도 차이를 기반으로 한 알고리즘으로 복잡하고 기하학적인 분포도를 가진 데이터에 대해서 클러스터링을 잘 수행한다. 미리 클러스터 수를 지정하지 않아도 되는 유연성을 보이고 이상치 탐지에서 좋은 성능을 보여준다

3가지 클러스터링 모델을 모델성능평가지표 (Silhouette Score / BIC)를 사용해서 가장 좋은 성능을 내는 모델을 선택하기로 했다.

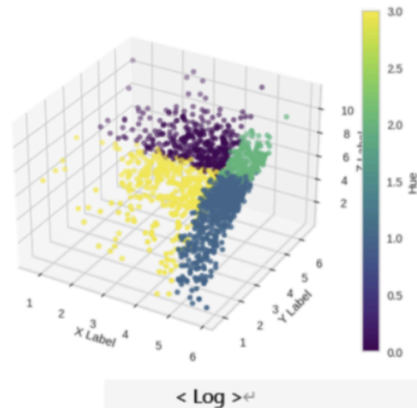
Silhouette Score <> BIC 를 비교하는 과정은 각 지표에서 성능이 가장 좋게 나온 클러스터 수를 보고 선택하기로 했다. GMM 모델에서 가장 최적의 성능을 낼 수 있는 K 수는 10개이고 K-Means에서 Elbow Method를 통해 얻은 최적의 K 수는 4개 였다.

추가적으로 우리는 4개의 Sagement로 나누기로 했으므로 K-Means 클러스터링을 최종적으로 선택하게 되었다. (실제로 Silhouette Score 도 0.5237 점으로 좋은 클러스터링에 속했다.)

데이터 소개

< 클러스터 확인 >				
	고객수	R	F	M
0	195	↑	↑	↑
1	390	↓	↓	↓
2	351	↓	↑	↑
3	532	↑	↓	↓

	해석	등급
0	높은 방문율과 구매율을 가졌지만 방문이 뜸해진 고객	재구매 유도 고객
1	자주 방문하지만 높은 실적을 남기진 않는 금액	일반 고객
2	최근까지도 방문하며 매장에 이익을 가져다 주는 우수 고객	프리미엄 고객
3	방문도 뜸하고 구매율이 적은 고객	이탈 위험 고객

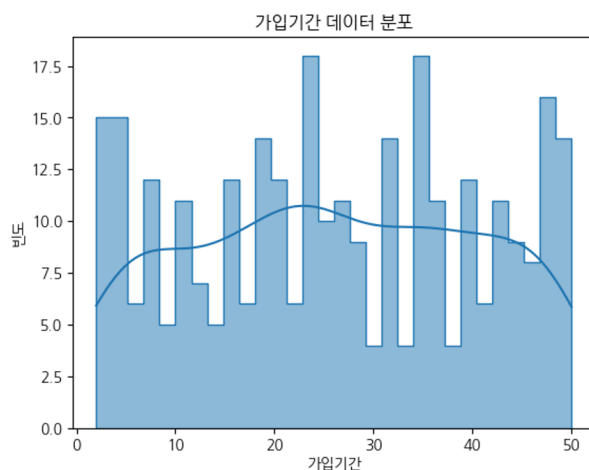


고객은 총 4가지 등급으로 나뉘었다. 일단 가장 분포 수도 많고, 기업입장에서도 고객 유지가 중요하므로 '이탈 위험 고객'이 중요하다고 생각했다. 따라서 '이탈 위험 고객'을 제외한 3가지 등급을 팀원들이 나눠서 분석하고 '이탈 위험 고객'은 팀원들이 공통으로 분석을 하기로 했다.

고객등급 별 분석

I. 일반 고객

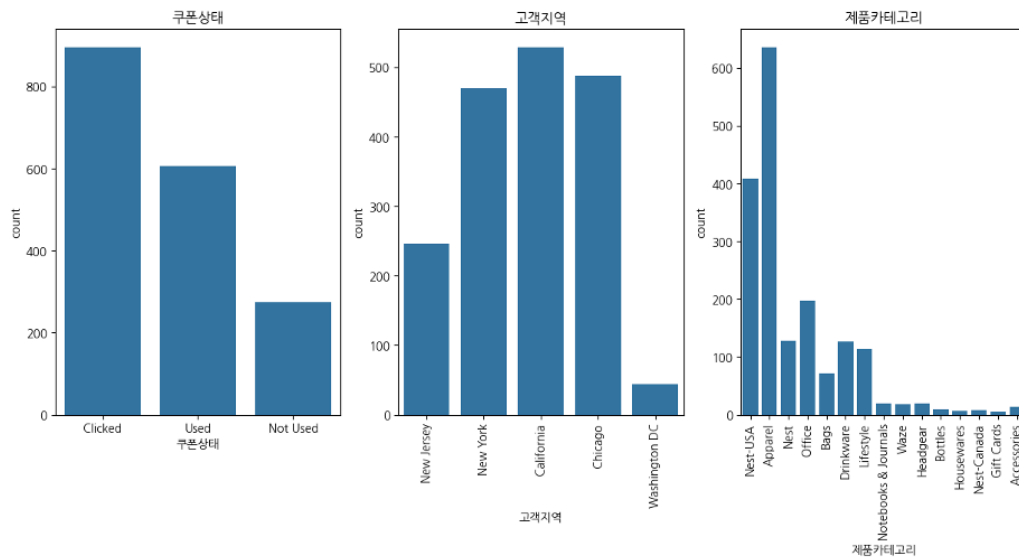
일반 고객은 자주 방문하지만 높은 실적을 남기진 않는 고객층이다. 따라서 제품 구매 빈도와 고가 제품 구매를 유도하는 방안을 세운다.



일반 고객의 가입기간 별 구매빈도 분포를 확인하면 비교적 골고루 분포되어 있으나, 특정기간에 감소하는 경향을 띈다.

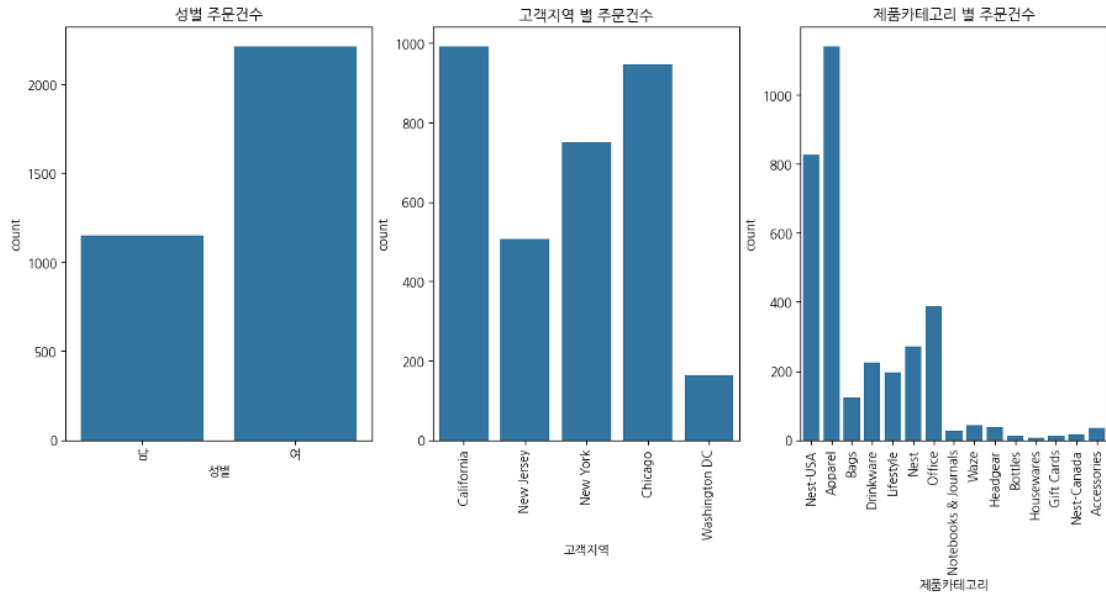
- 특정 기간 감소 고객

10 ~ 14개월, 21~23개월, 29~40개월의 경우 구매 빈도가 떨어지는 경향을 확인할 수 있다. 위의 기간에 해당하는 고객의 경우 쿠폰을 사용한 경우가 적었으며, Apparel, Nest 제품에 관심이 많은 것을 확인할 수 있다.



이들의 관심사에 따라 계절별 신상 제품 또는 고객 스타일의 제품을 “배너”를 통해 고객에게 꾸준히 알릴 수 있도록 한다.

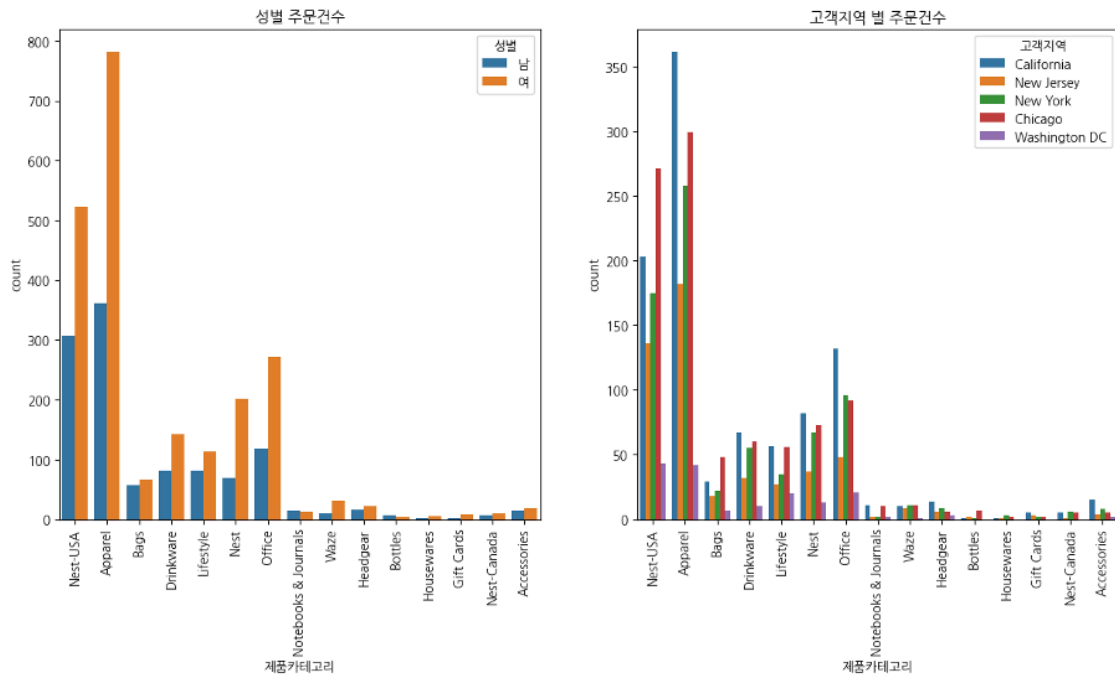
Nest제품은 스마트 홈 제품 라인으로 마트 스피커, 스마트 디스플레이, 스트리밍 기기, 온도조절기, 연기 감지기, 라우터 및 보안 시스템(스마트 도어벨, 카메라, 스마트 락 등)이 있다. 이 제품들은 상대적으로 고가인 제품으로 이 제품에 관심이 많은 고객의 경우 “쿠폰”을 제공해 구매를 유도할 수 있도록 한다. (할인율이 높은 제품 알림으로 고객의 제품 구매 유도)



앞서 [특정 기간에 따른 일반고객]과 [전체 일반 고객] 모두 Apparel, Nest 제품 구매를 많이 하는 것을 확인할 수 있다. 이 제품들의 경우 앞서 [특정 기간에 따른 일반고객]의 제품 구매 유도 방법을 [전체 일반 고객]에게도 적용할 수 있다.

남성보다 여성의 이커머스 사용률이 높은 것을 확인할 수 있다. 남성의 이용 비율이 여성의 이용 비율의 반절 밖에 되지 않는 것으로 보아 남성 고객을 더 많이 유도해야 한다.

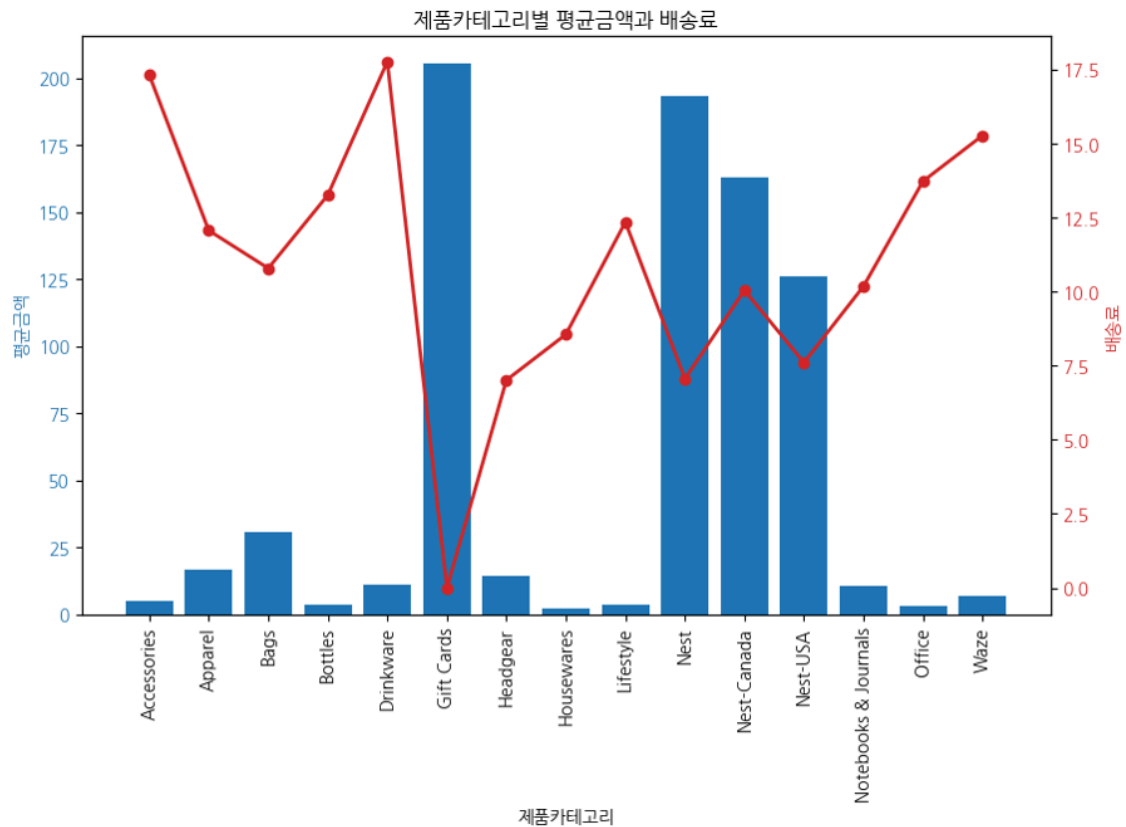
고객지역 별 주문건수는 캘리포니아와 시카고가 비슷하게 높으며 워싱턴 DC의 경우 많이 적은 것을 확인할 수 있다. 지역별 특징이 있는지 확인해봐야 한다. (워싱턴DC의 경우 상권 발달이 잘 되어 있어 이커머스를 활용하지 않는 것인지 등)



여성과 남성의 경우 압도적으로 Apparel, Nest 카테고리 제품에 관심이 많은 것으로 보인다. 이는 전반적으로 어떤 고객이든 두 카테고리의 제품에 관심이 많은 것으로 보인다.

남성의 경우 여성 보다 상대적으로 이용률이 낮으므로 Apparel 제품의 카테고리에 남성 옷의 다양성을 제공해 고객이 다양한 옷을 접할 수 있도록 하여 구매를 유도한다.

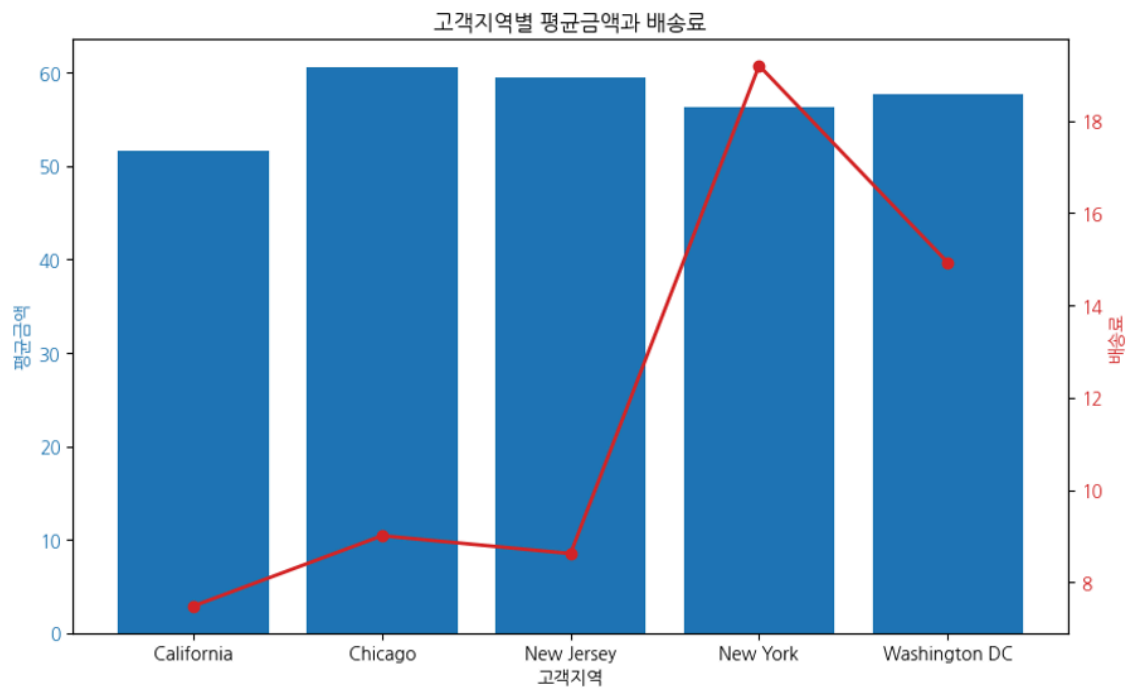
또한 남성의 경우 'Notebooks & Journals'와 'Bottles' 카테고리에서는 여성보다 높은 것으로 관찰된다. 이 카테고리의 제품에 다양성 뿐만 아니라 남성 고객이 선호할 포인트를 찾아 마케팅 전략을 세울 필요가 있어 보인다. 특히 Notebook의 경우 고가 제품으로 마케팅 전략을 잘 세운다면 높은 실적을 세울 수 있을 것으로 보인다.



제품별 평균금액과 배송료를 나타낸 그래프를 확인해보면 Gift Cards, Nest(Canada & USA)를 제외한 전 제품의 배송료가 평균금액보다 훨씬 높은 것을 확인할 수 있다.

배송료에 대한 불만이 고객에게 나올 수 있음을 인지할 필요가 있다. 고객의 제품 주문량이 많은 달(블랙프라이 또는 12월 등)에는 배송료 '무료' 또는 '할인' 등의 이벤트를 진행하면 좋을 것으로 보인다.

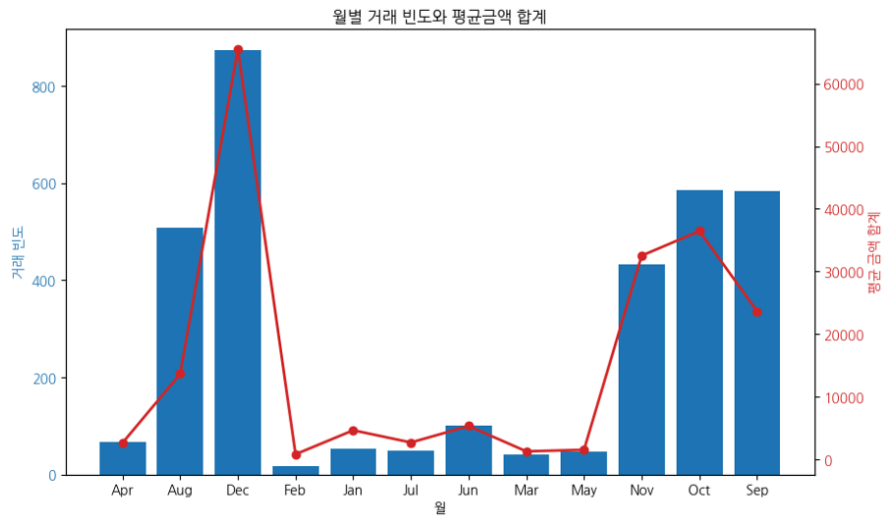
또한 Apparel 제품의 경우 주문건수가 굉장히 많은 제품이지만 배송료가 비싼 점이 고객의 제품 구매에 있어 걸림돌이 될 수도 있다는 점 또한 인지해야 할 필요가 있다. (상대적으로 저가의 제품이 배송료가 높다는 점 인지해야 함)



제품 뿐만 아니라 지역에서도 배송료 차이가 많이 나는 것으로 보인다. 평균 금액은 비슷하나, 뉴욕의 경우 압도적으로 배송료가 비싼 것을 확인할 수 있다. 이는 배송지가 뉴욕에서 많이 멀어 배송료가 높은 것인지 등 다른 이유도 찾아 봐야 할 것으로 보인다.

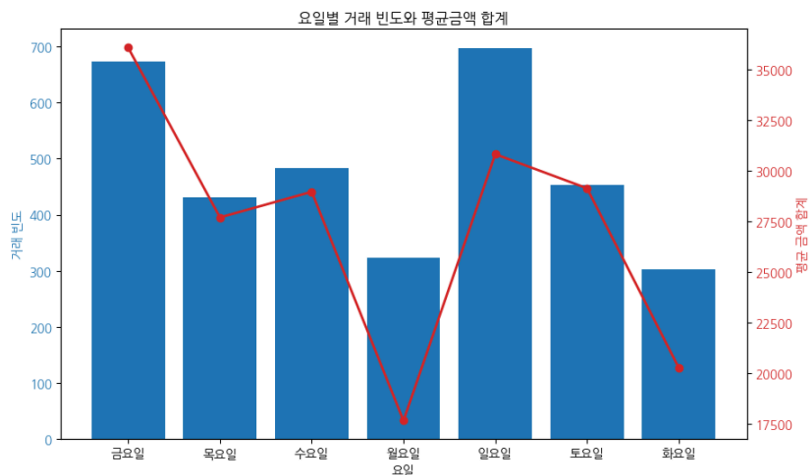
	제품카테고리	월	판매량
0	Accessories	Dec	13
1	Apparel	Dec	278
2	Bags	Sep	33
3	Bottles	Sep	5
4	Drinkware	Sep	63
5	Gift Cards	Oct	8
6	Headgear	Sep	12
7	Housewares	Aug	2
8	Lifestyle	Sep	72
9	Nest	Dec	119
10	Nest-Canada	Dec	6
11	Nest-USA	Dec	280
12	Notebooks & Journals	Aug	16
13	Office	Dec	96
14	Waze	Dec	16

12월에 판매량이 가장 높은 제품들이 많이 있다, 12월이 전체적으로 거래량이 많은 것을 암시한다.



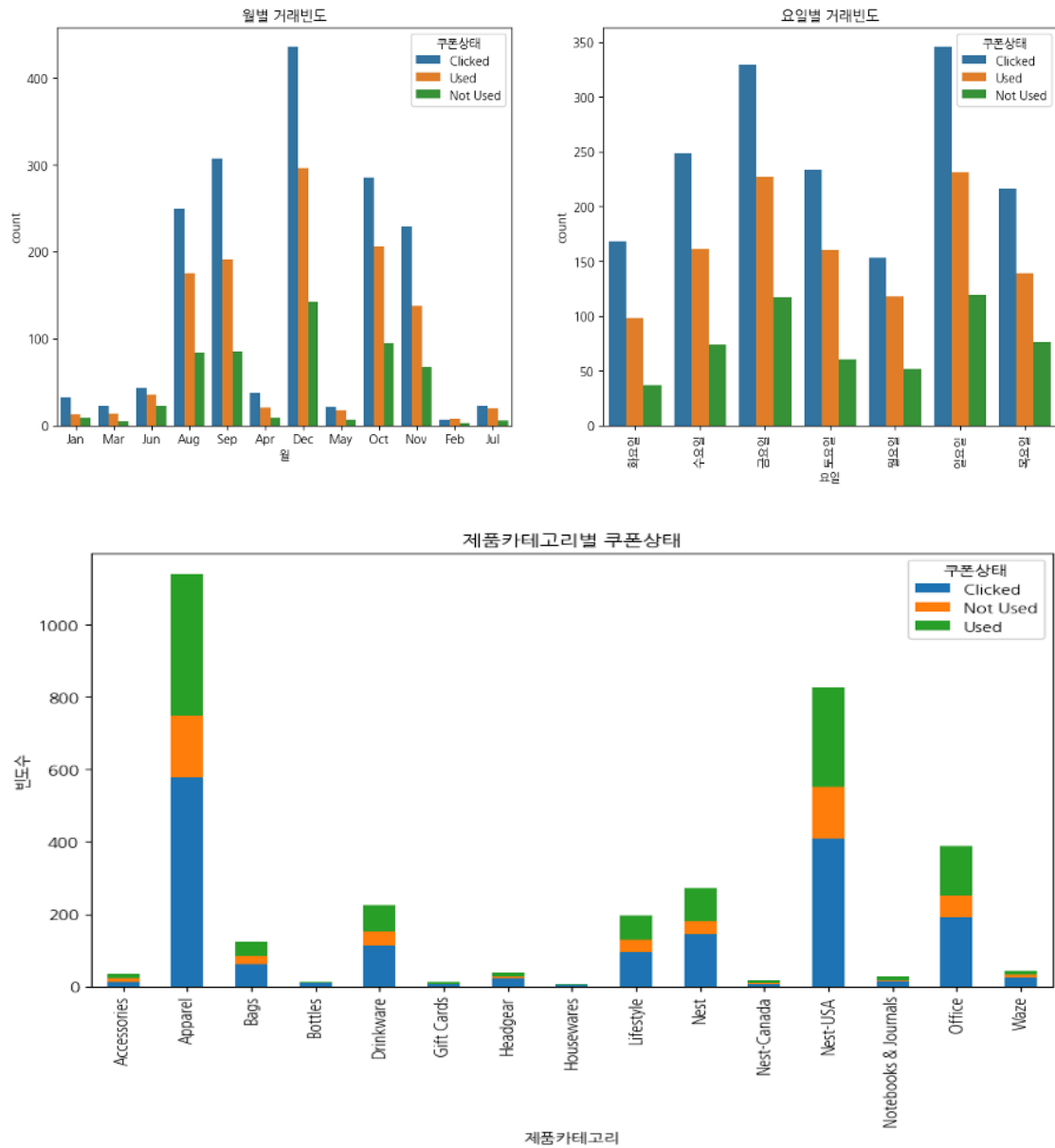
8~12월 거래 빈도가 높다. 12월과 11월의 경우 크리스마스와 블랙프라이데이가 있어 높은 것으로 파악된다. 8~12월의 경우 거래빈도가 높으므로 해당 달의 경우 평균 금액을 높이기 위해 고가의 제품을 많이 전시하면 좋을 것 같다.

1~7월의 경우 빈도수와 평균 금액 합계도 적다. 이 달의 경우 고객의 관심사를 분석할 필요가 있어 보인다.



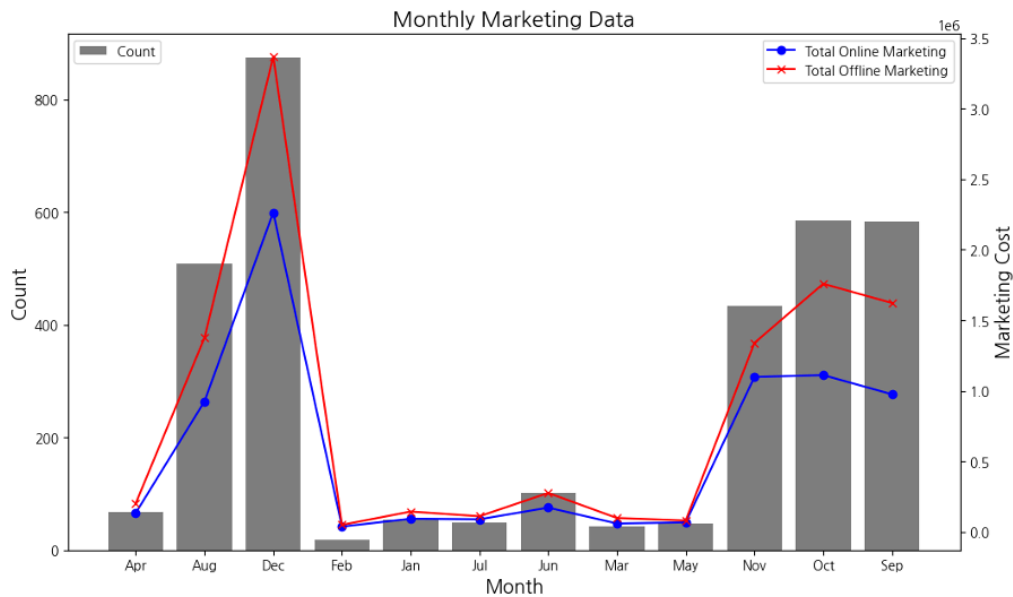
금요일과 일요일 거래량이 많은 것을 확인할 수 있다. 금요일과 일요일 고객이 선호하는 **Apparel, Nest** 카테고리의 제품을 더욱더 활발하게 홍보할 수 있도록 한다. 특히 금요일은 평균 금액 합계가 거래빈도수보다 높은 것을 확인할 수 있다. 거래빈도수와 금액 사이의 관계가 비례해 보인다. 이때는 고가의 제품도 많이 노출시키면 좋을 것 같다.

상대적으로 거래량이 적은 월요일과 화요일은 '직장인'을 타겟으로 마케팅을 하면 좋을 것 같다. 예를들어 '월요병'을 이길 수 있는 제품 등의 온라인 홍보를 하면 좋을 것 같다.



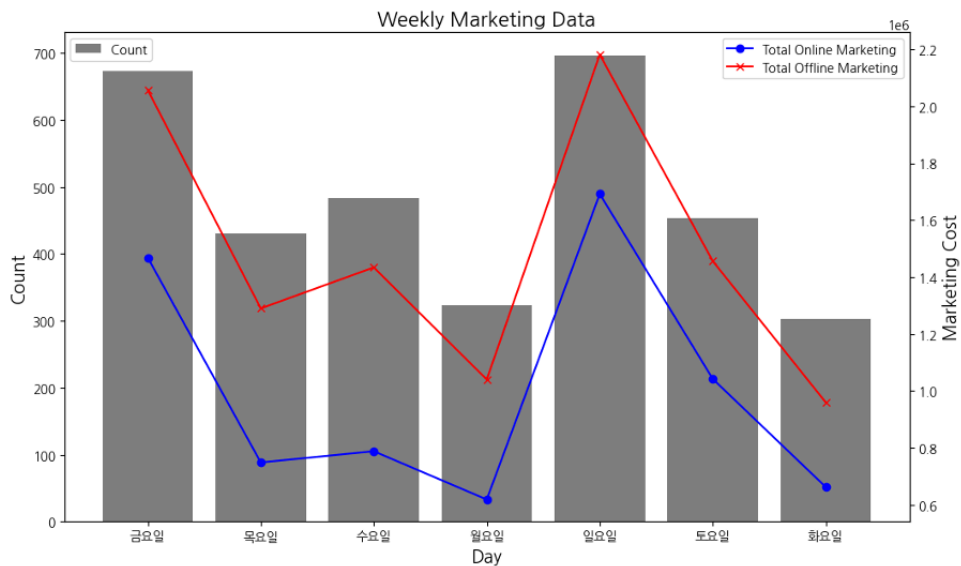
대부분 쿠폰 사용량이 적다. 거래가 많은 8,9,10,11,12월은 쿠폰 사용률을 높여 더 많은 제품을 구매할 수 있도록 유도하며, 위의 달은 특히 금요일과 일요일날 행사를 많이 하면 좋을 것으로 보인다. 또한 11월과 12월은 달의 특성상 고가의 제품에 할인 쿠폰을 많이 제공하면 좋을 것 같다.

1~7월달의 경우 쿠폰 사용으로 고객의 구매를 유도하는 것보다 고객의 관심사를 통해 이커머스 사용률을 늘리는 것이 더 좋을 것 같다.



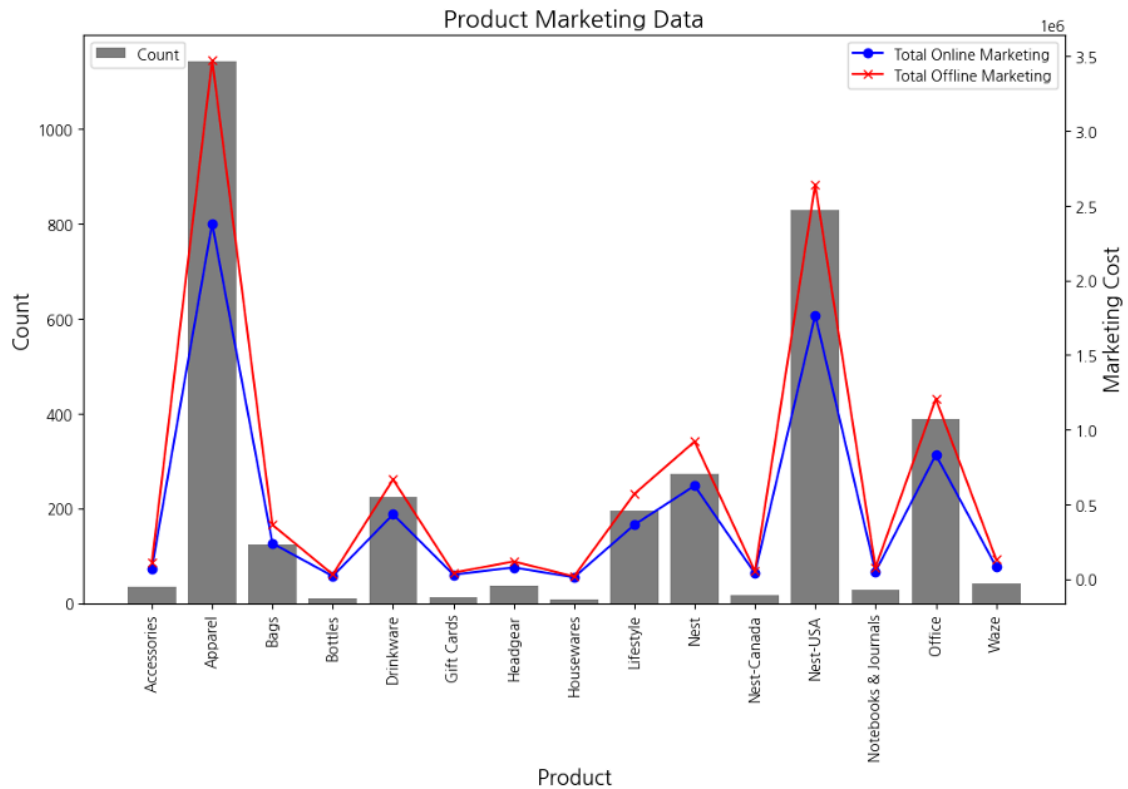
8~12월 달은 거래량 뿐만 아니라 매출 합계, 마케팅 비용 모두 높은 것을 확인할 수 있다. 이는 마케팅이 어느정도 거래량과 매출에 영향을 끼치는 것으로 파악할 수 있다.

따라서 특히 11,12월의 경우 마케팅을 더욱 공격적으로 하면 좋을 것으로 보인다.



요일별

마케팅 비용도 앞서 본 거래 빈도수와 매출 합계 그래프와 비슷한 형태를 보인다. 평일의 경우 온라인 마케팅 비용을, 주말의 경우 오프라인 마케팅 비용을 상대적으로 더 높이는 것이 좋을 것 같다. 특히 '직장인'들이 출퇴근 하며 핸드폰 하는 시간대를 노리면 좋을 것 같다.



제품 또한 거래 빈도수에 비례하게 마케팅 비용을 사용하고 있다. 온라인, 오프라인 그래프 모두 같은 동향을 보이고 있다. 이미 거래빈도가 많은 제품에 특히 고가에 해당하는 물건에 마케팅 비용을 사용하면 좋을 것 같다.

[결론]

일반고객의 경우 자주 방문하지만 높은 실적을 남기진 않는 고객이다. 비록 높은 실적을 남기진 않지만 꾸준히 방문하며 이커머스 거래를 많이 하는 고객이기 때문에 관리가 필요하다.

이 고객의 가입기간 별 거래 빈도를 봤을 때 10 ~ 14개월, 21~23개월, 29~40개월 3개구간에서 거래 빈도가 낮아지는 것을 확인할 수 있다. 가입기간이 이 구간에 해당하는 고객의 경우, 고객이 평소 구매했던 제품의 카테고리에서 신제품 또는 할인중인 제품을 “메시지” 또는 “알림” 등으로 고객에게 알려 거래 빈도수를 높이게 해야 한다.

또한 가입기간과 상관없이 일반고객 대부분 Apparel, Nest 카테고리의 제품을 많이 구매하는 것을 확인할 수 있다. 두 카테고리의 제품은 온라인, 오프라인 마케팅 비용을 많이 사용한 카테고리이기도 하다. 공격적인 마케팅을 한 만큼 고객에게 많이 노출되어 거래량이 많은 것으로 유추할 수 있다. 다만, 상대적으로 많이 마케팅을 한 것에 비해 평균금액의 총 액이 드라마틱한 높은 금액을 보이지 않는다. 이 두 카테고리의 경우

마케팅 비용을 줄이고, 다른 전략을 세우는 것이 좋을 것 같다. 예를 들어 고가의 제품을 더욱 공격적으로 노출 시키거나 신제품 홍보 등에만 마케팅 비용을 사용하는 등 마케팅 비용을 효율적으로 사용할 필요가 있어 보인다.

더불어, Apparel의 경우 거래의 대부분을 차지하는 만큼 다양성을 보유할 필요가 있다고 본다. 특히 “남성”의 구매 제품 카테고리에서 1위를 한 만큼 “남성” 고객의 마음을 사로잡을 수 있는 방안이 필요해 보인다. 다양한 종류의 남성 의류, 남성이 좋아하는 브랜드 유치 등 “남성”고객을 잡기 위한 마케팅 방안이 필요해 보인다.

따라서 대부분의 거래와 매출을 담당하는 Apparel과 Nest의 경우 전략적인 마케팅을 할 필요가 있어 보인다. 그러나 다른 카테고리의 제품에 비해 두개의 제품에 너무나 많은 마케팅 비용을 투자하고 있다. 이들 대신 고가의 제품에 대한 마케팅 비용을 늘려 고객에게 노출시킬 필요가 있다. 일반 고객의 경우 실적이 낮기 때문에 더더욱 고가의 제품을 구매할 수 있도록 유도해야 한다. 고가의 제품이지만 일반 고객들이 필요할 만하고, 사기에 부담 없는 금액대의 제품을 고객에게 노출시킬 필요가 있다.

또한 금요일과 일요일, 8~12월에 거래량이 뿐만 아니라 마케팅 비용 또한 높다. 11월, 12월의 경우 블랙프라이데이, 크리스마스 등 특별한 날이 있기 때문에 고객들이 제품을 많이 구매한다. 이 경우 많은 마케팅 비용을 투자하지 않아도 거래량이 계속 많을 것으로 생각한다. 따라서 이 달은 마케팅 비용을 줄이고 다른 달에 투자하는 것이 효율적일 것 같다. 날이 좋은 달은 오프라인 마케팅을 통해 고객이 단순히 시각적으로 즐기는 것 뿐만 아니라 체험을 통해 즐길 수 있도록 하여 거래 빈도를 높일 수 있도록 한다. 따라서 계절별 오프라인, 온라인 마케팅 전략을 짤 필요가 있다.

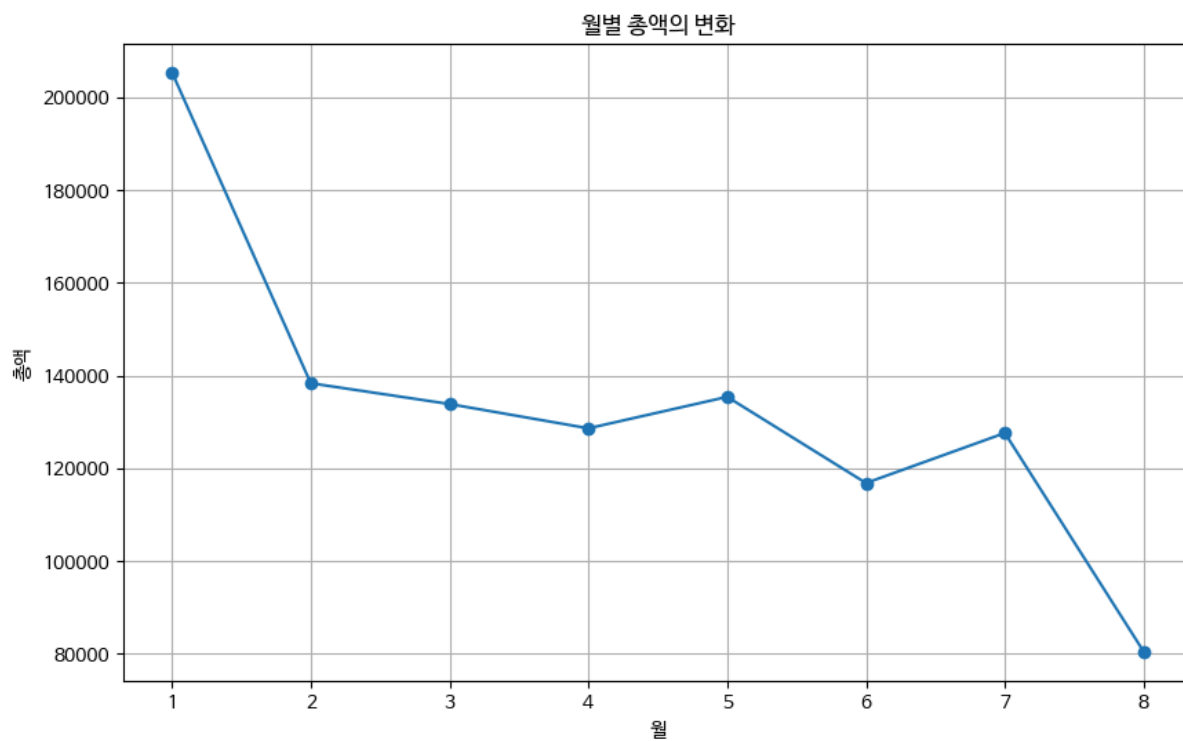
요일도 마찬가지로 평일의 경우는 온라인 마케팅 비용을 늘리고 오프라인 마케팅 비용은 줄이는 것이 좋아 보인다. 특히, “직장인”을 타겟으로 평일 출퇴근 시간대에 핸드폰을 많이 하는 상황을 노려 그 시간대에 온라인 마케팅을 노출 시키면 좋을 것 같다.

고가의 쿠폰 사용량은 적은 편이다. 이는 고객이 현명한 구매를 하지 못하는 것을 나타낸다. 뿐만 아니라 배송료 또한 제품 구매 금액에 비해 높은 경우가 많다. 이러한 경우 고객이 배송료와 제품 둘 다 비싸게 느껴 구매를 하지 않는 경우가 생길 수도 있다. 고객이 쿠폰을 사용할 수 있도록 유도해야 하며, 배송료의 경우 특정 금액 이상 절감 또는 무료 행사를 진행해 구매할 수 있도록 유도해야 한다. 또한 11,12월의 경우 거래 빈도수가 많으므로 이 달은 배송료 무료 혜택 등을 통해 고객이 더욱더 많은 제품을 자주 살 수 있도록 한다. 특히 뉴욕의 경우 다른 지역에 비해 배송료가 많이 비싸므로 뉴욕에 거주 중인 고객에겐 배송료에 대한 혜택을 많이 줄 필요가 있다.

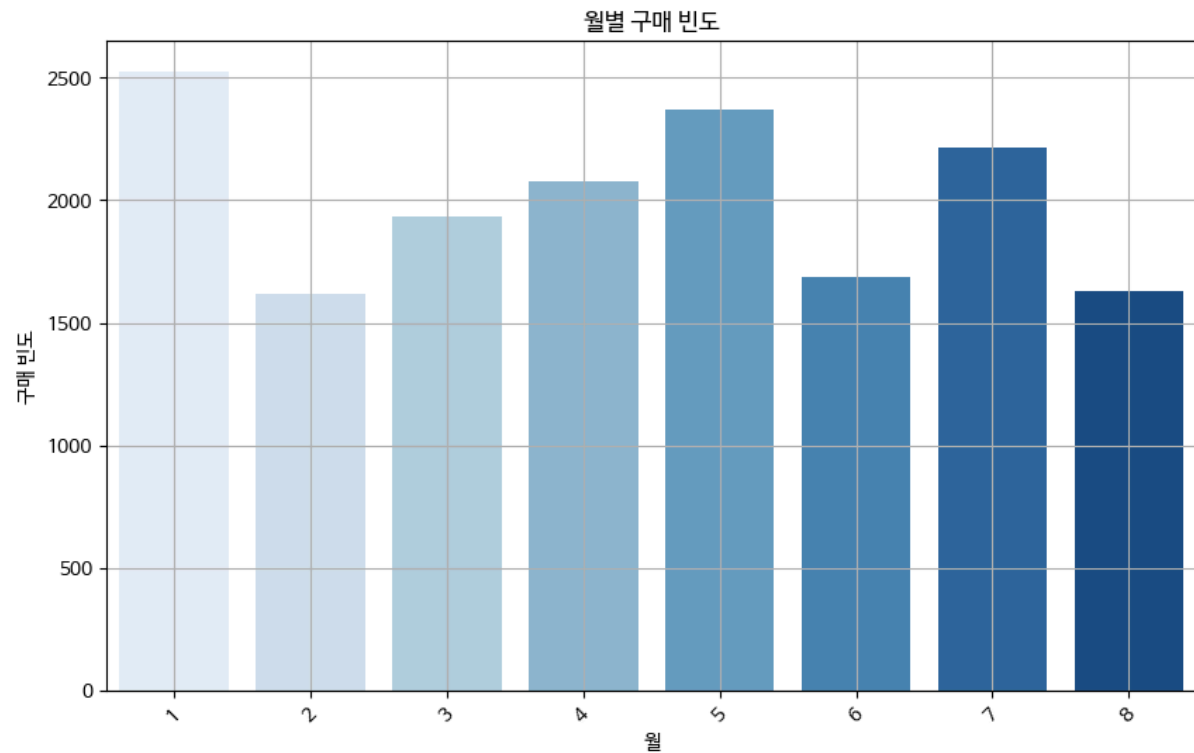
II. 재구매 유도 고객

재구매 유도 고객은 높은 판매량, 높은 매출을 기록했었지만 최근 방문량이 낮아진 고객층이다.

1. 월별 구매액 및 구매 빈도 분석

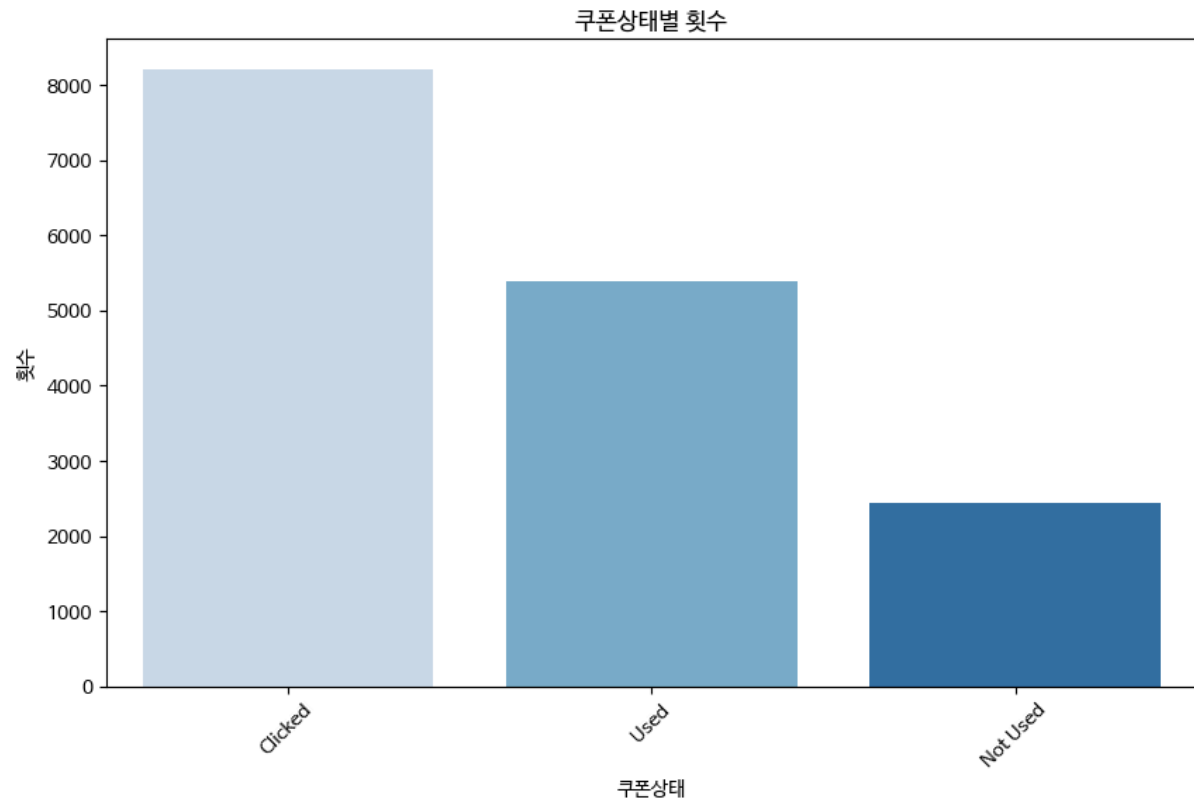


2월, 3월, 4월, 6월, 8월에 사용자의 구매 금액이 감소하는 것을 알 수 있다. 이중 2월 6월 8월인 경우 사용자들이 구매한 금액이 크게 감소한 것을 알 수 있다.

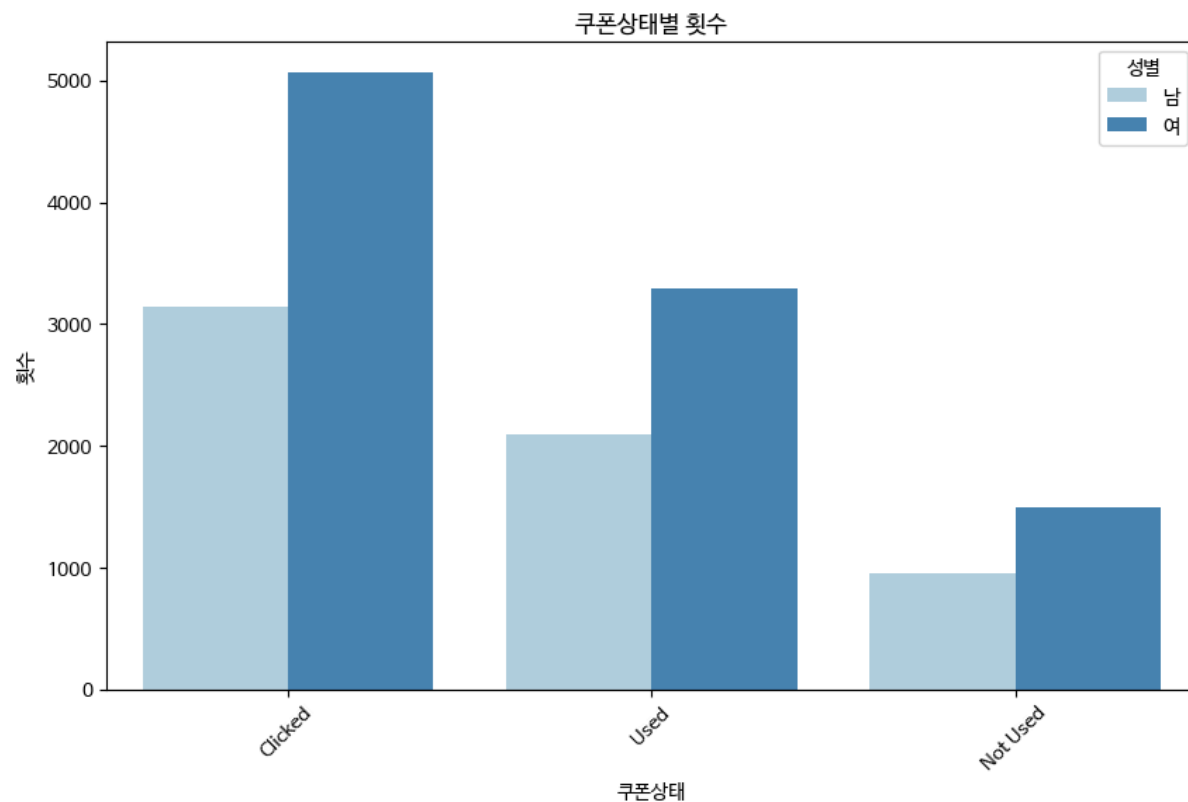


구매 빈도를 보았을 때 2월, 6월, 8월의 구매 빈도가 다른 월에 비해 낮은 것으로 보아 판매하는 제품의 종류가 크게 다르지 않을 것으로 판단된다. (만약 금액에서 차이가 나지만 구매 빈도에서 차이가 나지 않는다면 제품의 종류가 달라졌을 것이다.)

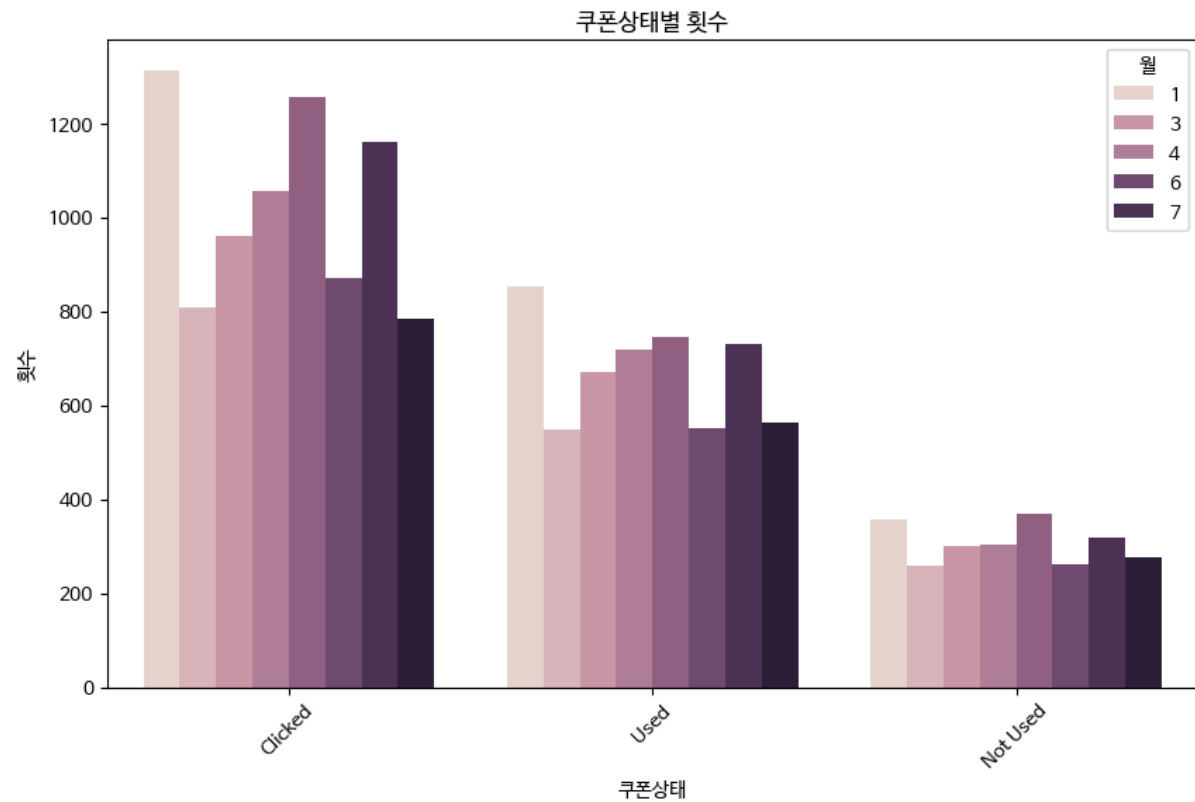
2. 쿠폰 사용 여부 분석



쿠폰을 사용한 사람들보다 쿠폰을 사용하지 않은 사람들이 더 많다는 것을 알 수 있다. (Clicked > Used > Not Used)



성별을 나눠 비교했을 때 성별에 상관없이 쿠폰 사용 여부는 비슷했지만 남성보다 여성의 분포가 많다는 것을 알 수 있다.

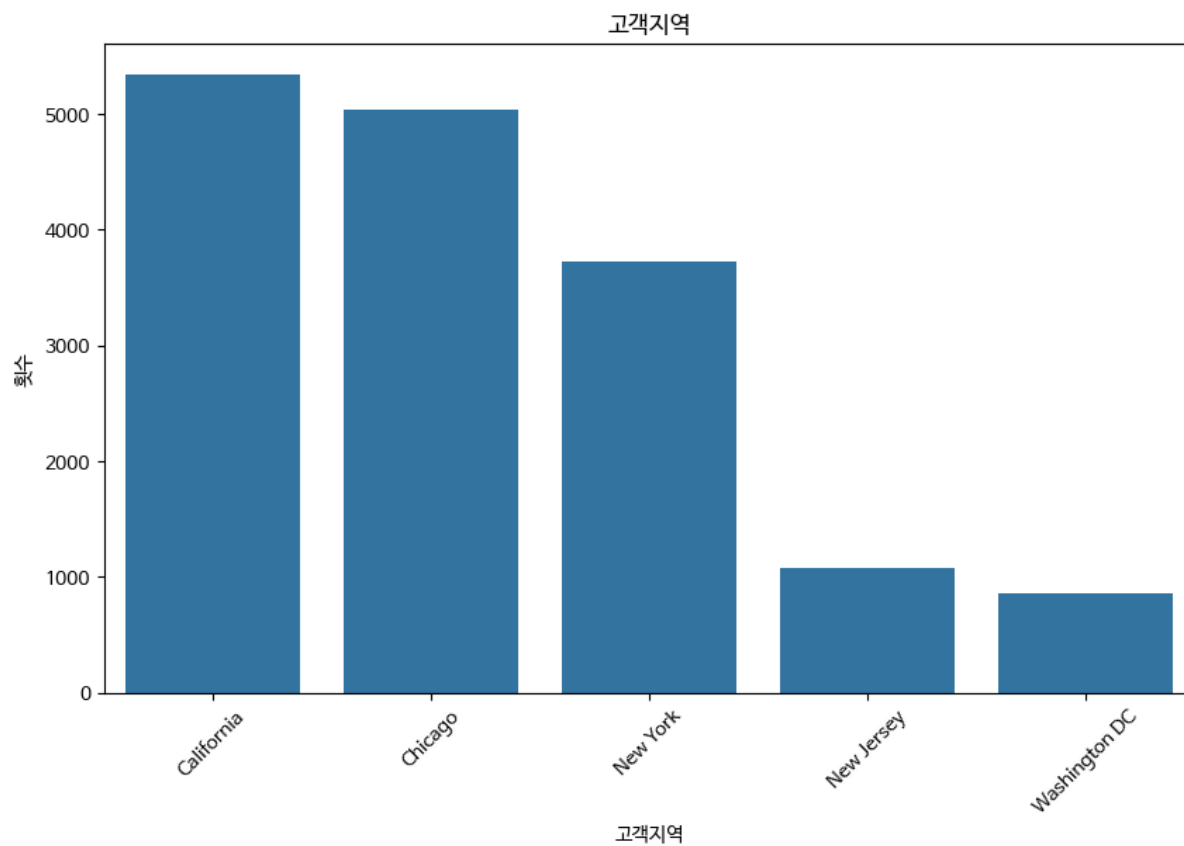


월별로 시각화했을 때 구매 빈도가 낮았던 2월, 6월, 8월에 다른 달에 비해 전체적으로 낮다는 것을 알 수 있다.

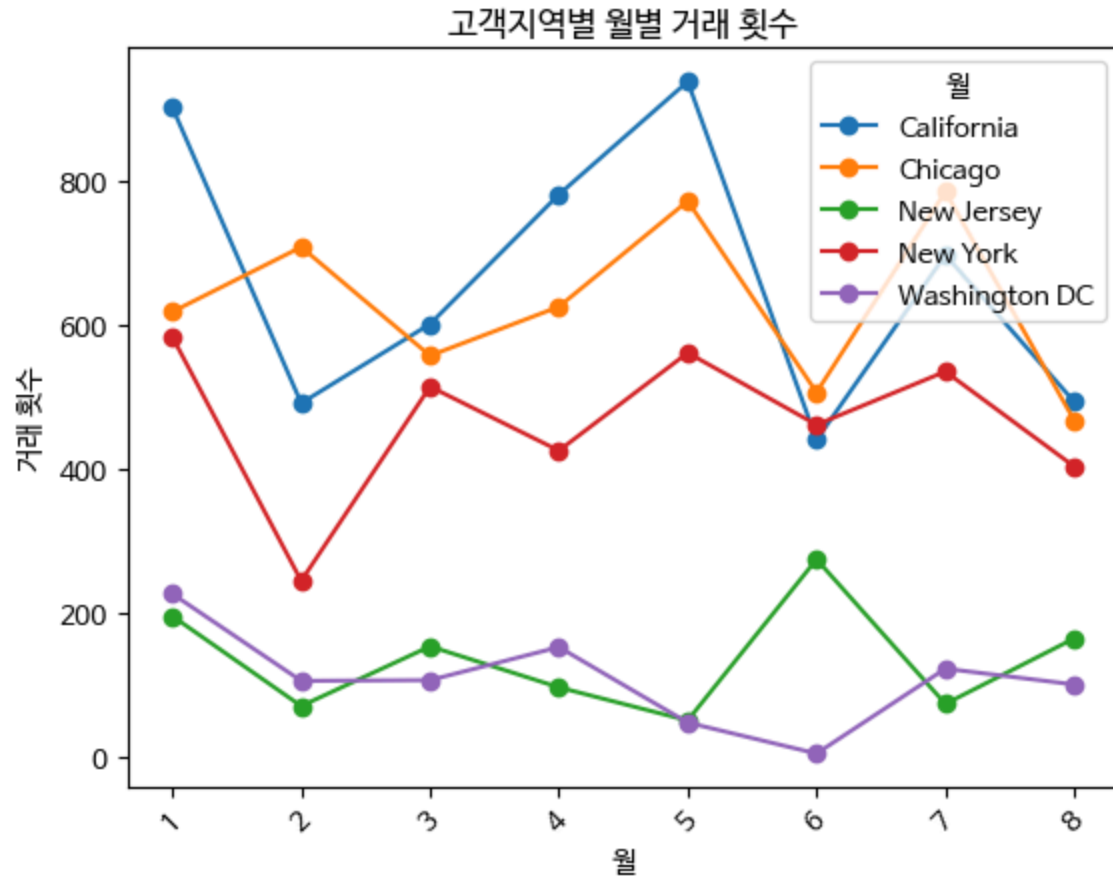
월	
1	10.0
2	20.0
3	30.0
4	10.0
5	20.0
6	30.0
7	10.0
8	20.0

월별 할인율을 보았을 때 2월, 6월, 8월의 할인율이 낮지 않기 때문에 해당 달의 할인율 때문에 구매빈도가 낮다고 판단하기는 어렵다.

3. 지역별 분석

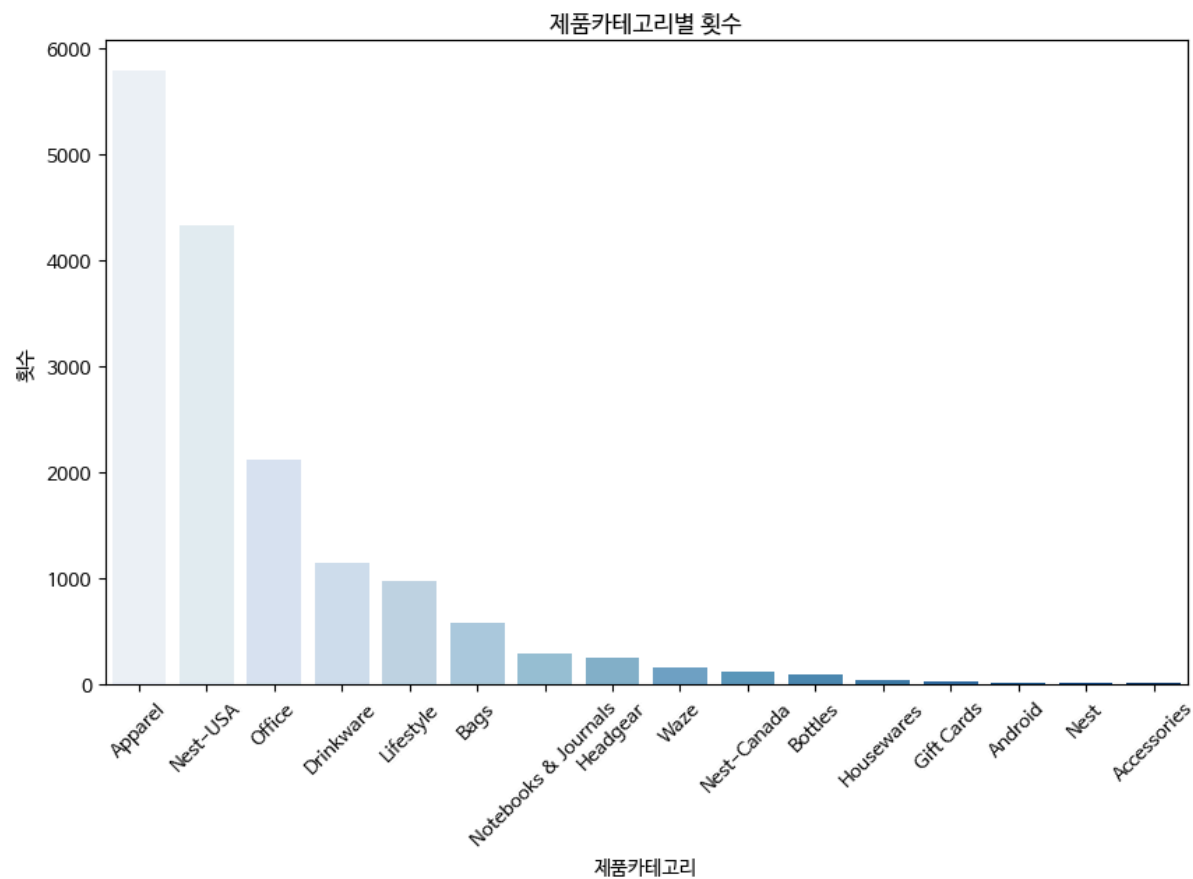


5개의 지역이 있고 그 중 California, Chicago, New York에서 고객 사용 빈도가 다른 두 지역보다 많다는 것을 알 수 있다.

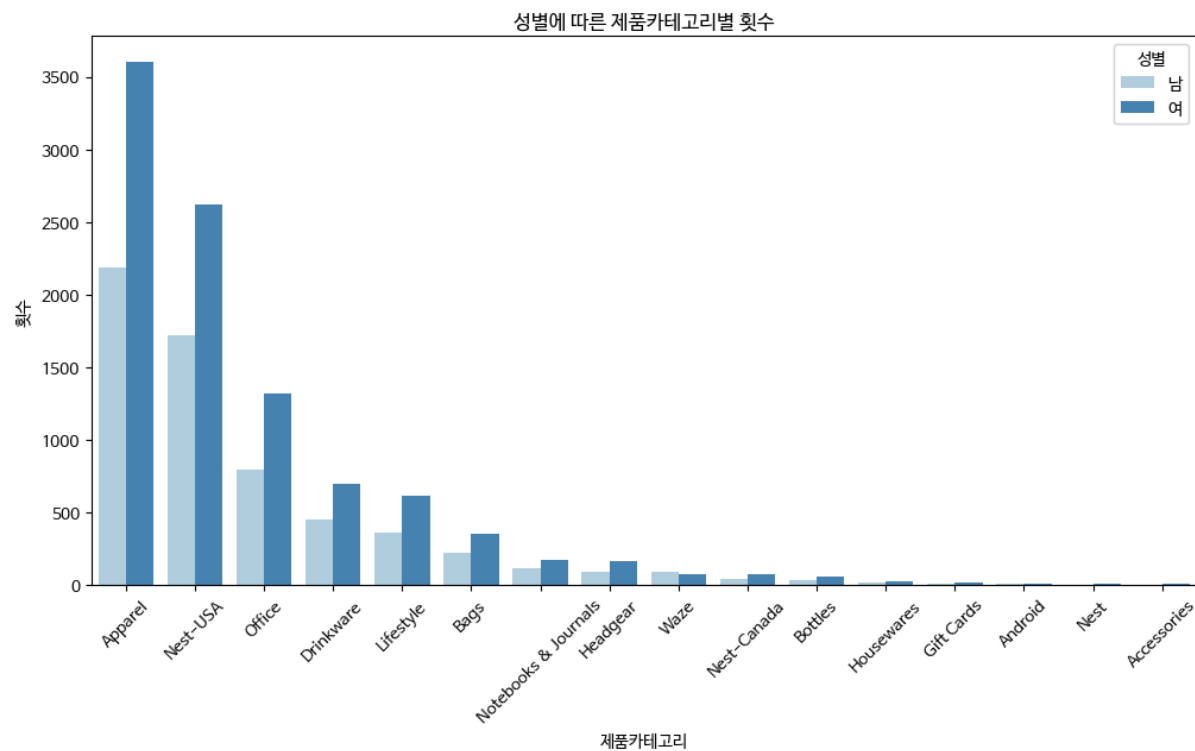


월별 지역 거래 빈도를 보았을 때 거래량이 많은 3개의 지역의 거래횟수가 감소할 때 거래액이 감소했다는 것을 알 수 있다. 예외적으로 2월의 **Chicago**는 거래횟수가 증가했지만 나머지 두 지역의 거래량이 크게 감소했기 때문에 2월 전체의 구매빈도 및 거래액이 감소했다는 것을 알 수 있다.

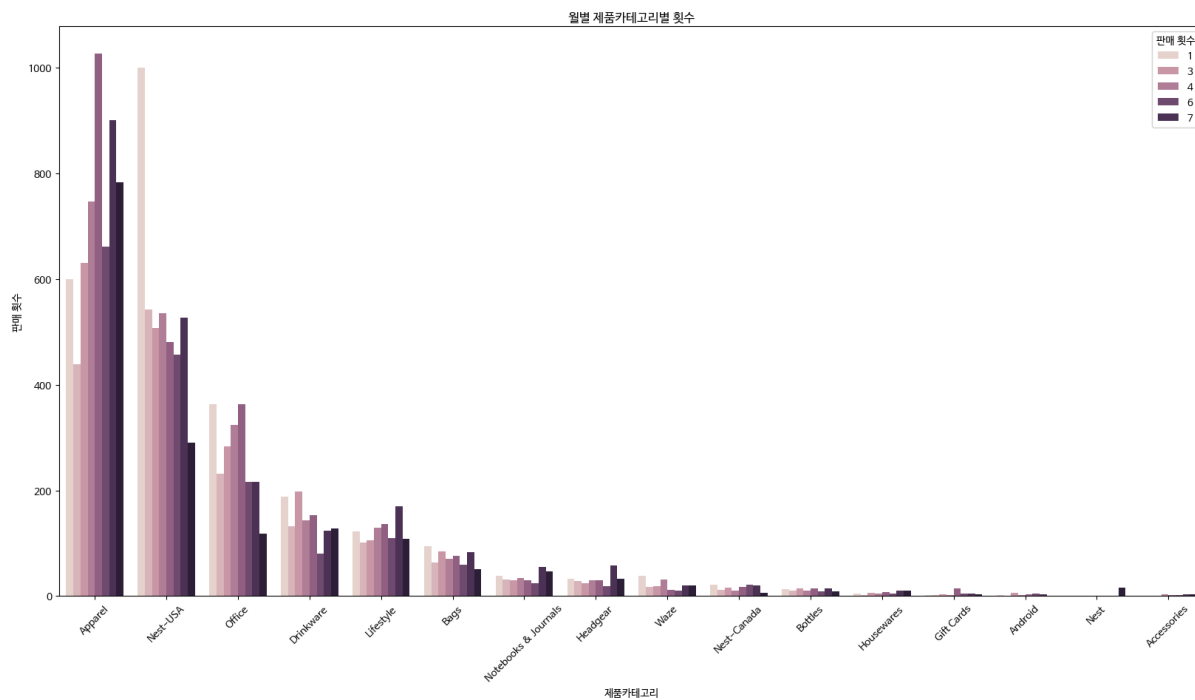
4. 제품카테고리 분석



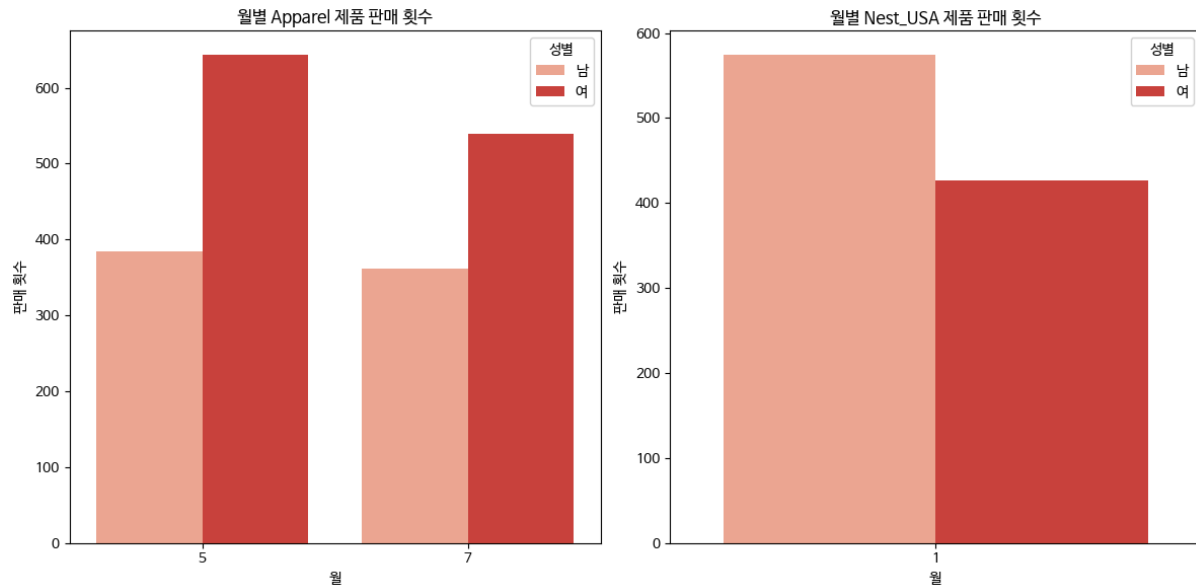
거래빈도가 가장 많은 카테고리 3개는 **Apearel**, **Next-USA**이다. 이 2개의 카테고리가 전체 거래량의 **63.14%**를 차지한다.



성별로 제품카테고리 거래량을 비교했을 때 제품별로 거래 비율을 다르지만 제품별 거래 순위는 다르지 않다는 것을 알 수 있다.

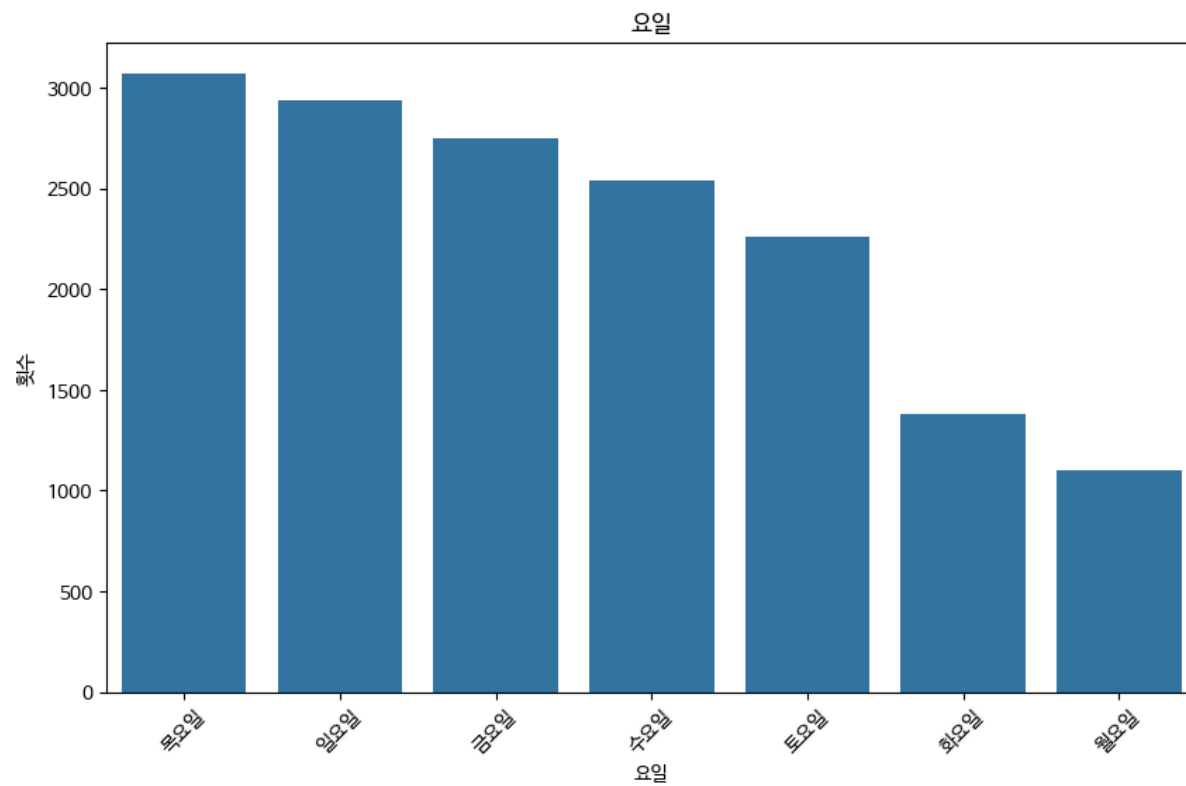


월별로 제품카테고리 판매 빈도를 보았을 때 1월의 **Nest-USA**의 거래량이 가장 많았다. 또한 거래량이 가장 많은 **Apparel**의 경우 2월에 가장 적게 거래되고 5월, 7월에 고객들이 많이 구매한다.

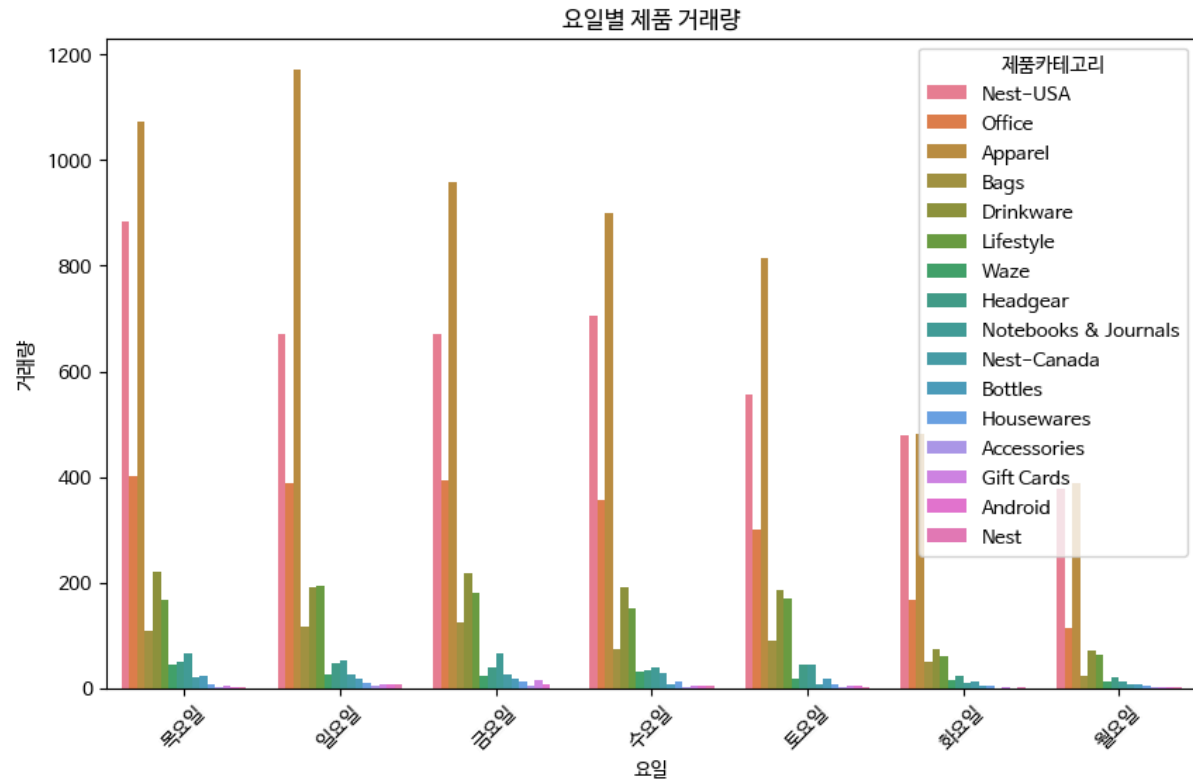


Apparel의 가장 많이 팔린 달(5, 7월) 2개와 **Nest-USA**가 가장 많이 팔린 달(1월)을 비교했다. 전반적인 카테고리 거래량을 보았을 때는 여성의 비중이 컸지만 **Nest-USA**의 1월을 보면 남성이 여성보다 거래를 많이 했다는 것을 알 수 있지만 여성의 거래량 또한 적지는 않다.

5. 요일별 분석



목요일의 거래량이 가장 많고 월요일이 가장 적다.



카테고리별로 나뉘었을 때 목요일에 **Apparel**의 거래량이 일요일에 비해 조금 적었지만 **Nest-USA**의 거래량이 많았기 때문에 총 거래량이 가장 많았고, 화요일과 월요일은 **Apparel**, **Nest-USA**의 거래량의 유사했다.

[결론]

재구매 유도 고객 중 **Apparel**, **Next-USA** 카테고리를 많이 구매하기 때문에 이 카테고리들에 대해서 혜택을 주면 판매량이 증가할 것으로 예상된다. 그중 **Apparel**의 경우 5월과 7월에 가장 판매가 많이되기 때문에 해당 달에 집중적으로 혜택을 주는 것이 좋다고 판단된다.

Nest-USA 같은 경우 1월의 거래량이 가장 많기 때문에 1월에 혜택을 주면 판매량이 증가할 것으로 예상된다. 또한 1월의 구매량은 여성보다 남성이 많아 남성 마케팅에 집중하되 여성 구매량도 적지 않기 때문에 여성 사용자 또한 고려해야 한다.

요일별 분석에서 목요일에 **Nest-USA**의 판매량이 가장 많았고 일요일에 **Apparel**의 판매량이 가장 많았으므로 해당 요일에 마케팅을 하면 효율적일 것이라고 판단된다.

III. 프리미엄 고객

프리미엄 고객은 모든 측면에서 가게에 이익을 가져다 주는 고객층이다. 따라서 해당 고객층을 유지하는 것에 주의를 기울여야 한다. 그러기 위해선 할인, 경품, 이벤트같은 서비스 차원에서 힘을 쏟는 전략을 세웠다.

1. 고객층 분석

우선 고객층에 대해 분석을 진행했다. 총 7개의 칼럼을 (성별, 고객지역, 시간대(초중말/월/요일), 제품카테고리, 쿠폰상태) 기준으로 분석해봤다.

< 성별 >

	총 판매수량	총 거래횟수	총 매출	평균 매출	평균 배송료	평균 오프라인비용	평균 온라인비용	고유 고객수	총 판매수량_비용	총 매출_비용	총 거래횟수_비용
성별											
남	50824	11404	1097818.69	96.266108	9.963713	2903.367240	1949.829004	175	0.373461	0.381274	0.37549
여	85265	18967	1781522.01	93.927453	10.215217	2936.784942	1935.693429	261	0.626539	0.618726	0.62451

< 고객지역 >

	총 판매수량	총 거래횟수	총 매출	평균 매출	평균 배송료	평균 오프라인비용	평균 온라인비용	고유 고객수	총 판매수량_비용	총 매출_비용	총 거래횟수_비용
고객지역											
California	40642	8838	877981.04	99.341598	10.676041	3019.891378	1918.275575	126	0.298643	0.304924	0.291001
Chicago	55059	11465	1080225.34	94.219393	10.246071	2947.893589	1941.955157	153	0.404581	0.375164	0.377498
New Jersey	11583	2636	250909.92	95.185857	9.363930	2736.191199	1873.616855	39	0.085113	0.087141	0.086793
New York	23074	5872	526673.54	89.692360	9.497745	2836.665531	2008.520131	92	0.169551	0.182915	0.193342
Washington DC	5731	1560	143550.86	92.019782	9.678250	2855.833333	1922.453583	26	0.042112	0.049855	0.051365

< 시간대 - 초중말 >

	총 판매수량	총 거래횟수	총 매출	평균 매출	평균 배송료	평균 오프라인비용	평균 온라인비용	고유 고객수	총 판매수량_비용	총 매출_비용	총 거래횟수_비용
초중말											
초	41975	9031	866206.58	95.914802	9.764517	2920.595726	2014.286079	140	0.308438	0.300835	0.297356
중	48649	10404	982719.97	94.455976	10.404428	3195.934256	2002.963690	133	0.357479	0.341300	0.342564
말	45465	10936	1030414.15	94.222216	10.145134	2668.763716	1821.533963	163	0.334083	0.357865	0.360080

< 시간대 - 월 >

	총 판매수량	총 거래횟수	총 매출	평균 매출	평균 배송료	평균 오프라인비용	평균 온라인비용	고유 고객수	총 판매수량_비용	총 매출_비용	총 거래횟수_비용
월											
Jan	5533	1132	110784.73	97.866369	15.491237	3223.586572	2161.224647	50	0.040657	0.038476	0.037272
Feb	6100	1517	137081.28	90.363401	17.143164	3066.183256	2103.202426	1	0.044824	0.047609	0.049949
Mar	8603	1909	158505.09	83.030430	14.376029	2536.930330	1560.253735	25	0.063216	0.055049	0.062856
Apr	14561	1535	202430.06	131.876261	10.645414	3341.693811	2113.867270	2	0.106996	0.070304	0.050542
May	7378	1847	126075.84	68.259794	8.906145	2051.164050	1639.671971	2	0.054215	0.043786	0.060815
Jun	10699	1902	150411.58	79.080747	8.866945	2659.305994	1727.527292	33	0.078618	0.052238	0.062626
Jul	11904	2568	185607.85	72.277200	9.422574	2176.207165	1731.322418	1	0.087472	0.064462	0.084554
Aug	16641	3552	249466.09	70.232570	9.957185	2755.489865	1820.730935	88	0.122280	0.086640	0.116954
Sep	16051	3690	316531.83	85.780984	9.763870	2770.189702	1726.029363	57	0.117945	0.109932	0.121497
Oct	14966	3567	352156.82	98.726330	9.085091	3064.900477	1873.102871	49	0.109972	0.122305	0.117448
Nov	12664	3524	455510.78	129.259586	7.825948	3067.111237	2222.278734	53	0.093057	0.158200	0.116032
Dec	10989	3628	434778.75	119.839788	8.588170	3956.312018	2475.345196	75	0.080749	0.150999	0.119456

<시간대 - 요일>

	총 판매수량	총 거래횟수	총 매출	평균 매출	평균 배송료	평균 오프라인비용	평균 온라인비용	고유 고객수	총 판매수량_비율	총 매출_비율	총 거래횟수_비율
요일											
월요일	6604	2733	232936.39	85.231025	7.992012	2933.260154	2021.115229	39	0.048527	0.080899	0.089987
화요일	5344	2589	234614.41	90.619703	8.493167	2946.427192	1897.292696	27	0.039268	0.081482	0.085246
수요일	24104	5409	543356.98	100.454239	10.037040	2895.230172	1695.129131	85	0.177119	0.188709	0.178098
목요일	23987	4913	486761.76	99.076279	9.743403	2910.136373	1507.110767	83	0.176260	0.169053	0.161766
금요일	31290	5240	547161.71	104.420174	10.884523	2931.870229	1871.259101	78	0.229923	0.190030	0.172533
토요일	24361	5001	428096.65	85.602210	11.222905	2917.076585	2287.693677	64	0.179008	0.148679	0.164664
일요일	20399	4486	406412.80	90.595809	10.750531	2955.416852	2384.040588	60	0.149895	0.141148	0.147707

<제품카테고리>

	총 판매수량	총 거래횟수	총 매출	평균 매출	평균 배송료	평균 오프라인비용	평균 온라인비용	총 판매수량_비율	총 매출_비율	총 거래횟수_비율
제품카테고리										
Accessories	934	183	4724.14	25.814973	11.051148	3282.513661	2077.699781	0.006863	0.001641	0.006025
Android	19	19	289.35	15.228947	7.205263	2578.947368	1740.852632	0.000140	0.000100	0.000626
Apparel	19246	10072	337361.15	33.494951	9.344010	2850.744639	1889.989626	0.141422	0.117166	0.331632
Bags	8210	1070	85254.07	79.676701	13.497056	2828.596131	1839.851486	0.060328	0.029609	0.035231
Bottles	1171	145	3872.74	26.708552	12.779034	2793.103448	1922.621379	0.008605	0.001345	0.004774
Drinkware	18876	1919	125181.05	65.232439	14.697822	2896.508598	1911.745857	0.138703	0.043476	0.063185
Gift Cards	108	103	13651.54	132.539223	0.000000	3115.533981	1953.446311	0.000794	0.004741	0.003391
Headgear	2200	431	33068.16	76.724269	9.541601	2826.682135	1898.391183	0.016166	0.011485	0.014191
Housewares	1051	56	2118.50	37.830357	19.573750	2678.571429	1947.421429	0.007723	0.000736	0.001844
Lifestyle	13682	1744	40851.59	23.424077	13.977930	2799.082569	1904.045419	0.100537	0.014188	0.057423
Nest	2462	1904	450173.87	236.435856	7.365572	3319.852941	2153.074785	0.018091	0.156346	0.062691
Nest-Canada	230	165	35493.92	215.114667	9.232485	3033.939394	2047.436303	0.001690	0.012327	0.005433
Nest-USA	12756	8256	1511819.46	183.117667	7.212678	3007.279554	2004.130159	0.093733	0.525058	0.271838
Notebooks & Journals	6643	376	75286.86	200.231011	20.778059	2647.340426	1826.703936	0.048814	0.026147	0.012380
Office	47744	3608	156065.11	43.255297	14.231749	2856.125277	1899.465327	0.350829	0.054202	0.118798
Waze	757	320	4129.19	12.903719	8.637781	2933.750000	1913.128250	0.005563	0.001434	0.010536

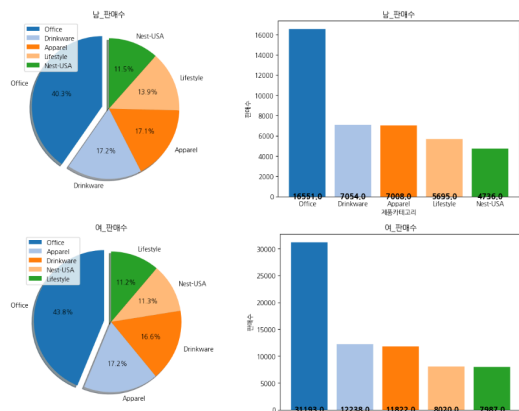
<쿠폰상태>

	총 판매수량	총 거래횟수	총 매출	평균 매출	평균 배송료	평균 오프라인비용	평균 온라인비용	고유 고객수	총 판매수량_비율	총 매출_비율	총 거래횟수_비율
쿠폰상태											
Clicked	68585	15408	1450302.76	94.126607	10.265616	2923.435877	1941.106698	221	0.503972	0.503693	0.507326
Not Used	19631	4629	463003.35	100.022327	9.744264	2927.003672	1948.663556	72	0.144251	0.160802	0.152415
Used	47873	10334	966034.59	93.481187	10.073485	2924.191988	1937.411614	143	0.351777	0.335505	0.340259

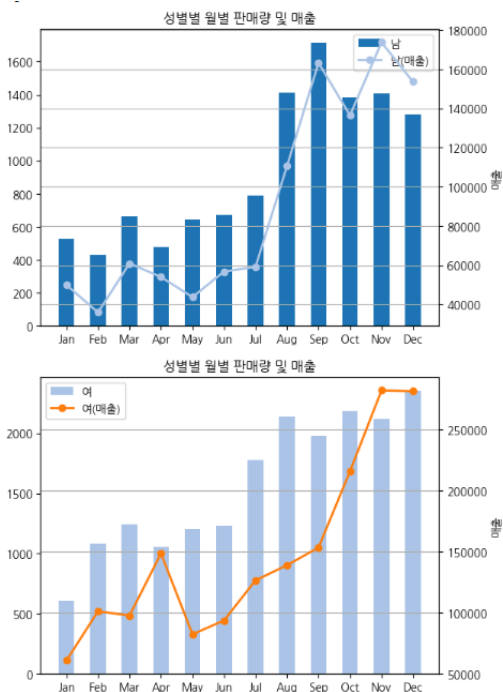
종합해보자면 프리미엄 고객층은 여성이 남성에게 비해 총 판매수량(+167%), 총 거래횟수(+166%), 총 매출(+162%)을 기록했고, 전체적으로 60%이상의 비중을 차지하고 있다. 지역에서는 **Chicago**(미국 중서부)와 **California**(미국 서부)에서 판매수량, 매출, 거래횟수의 70%를 담당하고 있다. 시간대별로 살펴보자면, 초(1일~10일) 중(11일~20일) 말(21일~) 시기별로는 큰 차이가 없고, 그나마 중순에 소비패턴이 늘어난 것을 볼 수 있었다. 월단위로 보자면 연초에는 소비가 적다가 4월에 한번 소비가 늘어나고 8월부터 연말까지도 높은 소비율을 보였다. 4월에는 연말 신고 기간, 따뜻해진 날씨, 학기의 시작 8월 이후는 명절, 학기의 시작, 연말 휴가 등 여러 요인으로 인한 것으로 보인다. 따라서 이 시기를 적절하게 이용하면 될 것 같다. 요일단위에서는 예상한 그대로 금요일에 판매수량이 가장 많았지만 의외로 수요일에는

거래량이 적음에도 매출이 금요일과 비슷하게 나왔다. 제품별로는 **Office**(사무용품)이 제일 많은 판매량을 가졌지만 그에 비례해 매출은 나오지 않았다. 그에 이어서 **Apparel**, **Drinkware**, **Lifestyle**, **Nest-USA**가 순위에 있지만 Apparel은 판매수량 대비 매출이 준수하게 나오고, Nest-USA가 판매되는 수량도, 금액도 매우 높기 때문에 해당 제품에 대한 관리가 필요하다. 참고로 다른 **Nest+@** 제품들도 높은 수익성을 보장하고 있다. 쿠폰같은 경우 **Clicked**(클릭만하는 고객)가 전체의 50%를 차지하고 **Used**(쿠폰을 사용한 고객)는 35%밖에 안됐다. 따라서 클릭에서 사용으로 전환할 수 있도록 쿠폰이 넓은 범용성을 가지고 가야한다.

2. 판매량, 매출이 높은 제품에 대한 장려 (상위 5)



성별별로 판매량이 높은 제품들을 확인해 봤다. 비율을 봤을 때 남성/여성간의 제품 선호도는 큰 차이가 보이지는 않았다. (각 성별 다 비슷한 제품을 구매를 한다) 그나마 여성은 Drinkware보다 Apparel을 선호하는 것으로 보인다. 하지만 판매량은 여성이 압도적으로 많으므로 여성에게 집중적으로 마케팅을 해야한다,

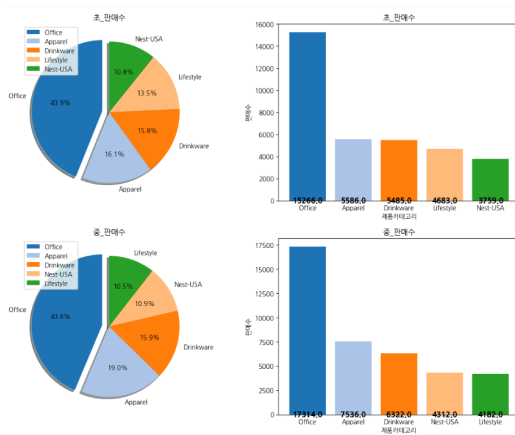


성별별 월별 판매량, 매출이다.

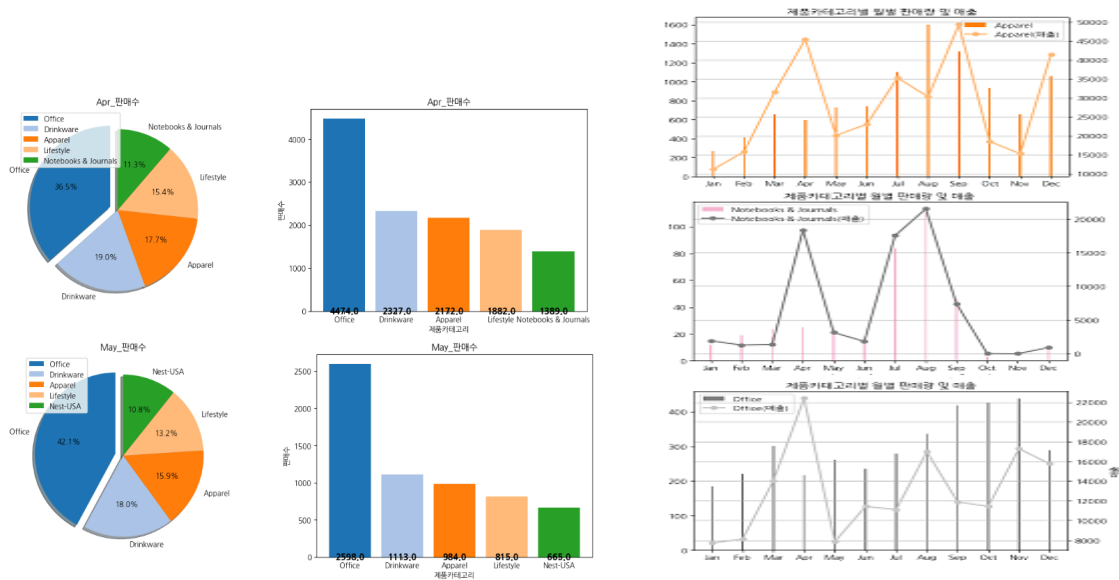
남성은 8,9월에 급격하게 소비량이 늘어나고, 여성은 4월에 한번 많은 소비를 하고 5월부터 천천히 소비량을 늘려나간다.



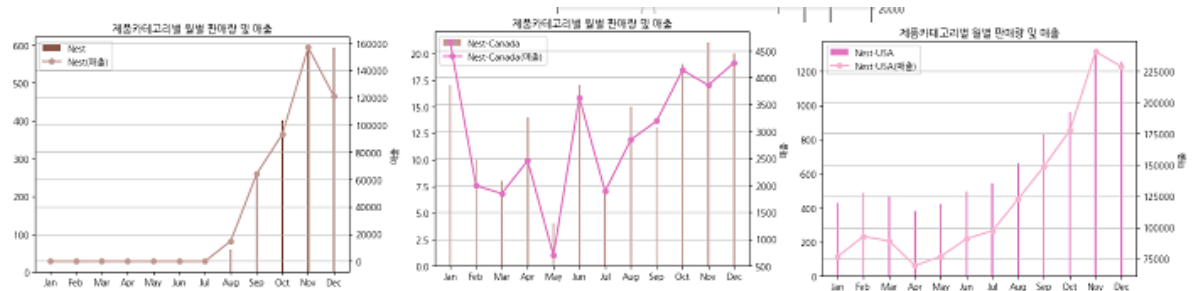
지역별 판매량이다. 높은 판매량과 매출을 가진 Chicago와 California지역을 보면 소비내역이 비슷하다. New Jersey는 Lifestyle / New_York 과 Washington DC지역은 Apparel의 비중이 높았다. 지역별로 달리 상품에 대한 마케팅을 진행해야한다.



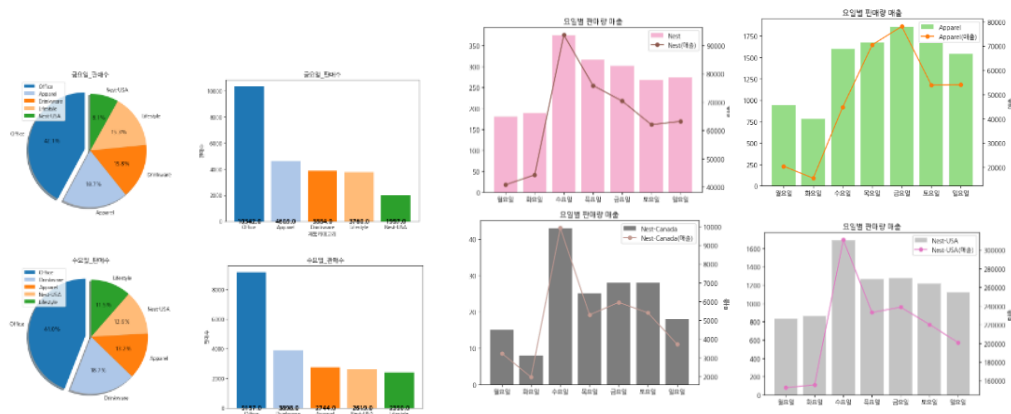
시간대(초중말)별 판매량이다. 중순에 소비가 늘어나는 것은 전체적으로도 판매량이 증가했지만, Apparel과 Office 제품에서 특히 늘었다. 즉, 중순에는 Apparel과 Office에 대한 소비가 많아지는 시기이므로 두 제품에 대한 마케팅을 진행해야한다.



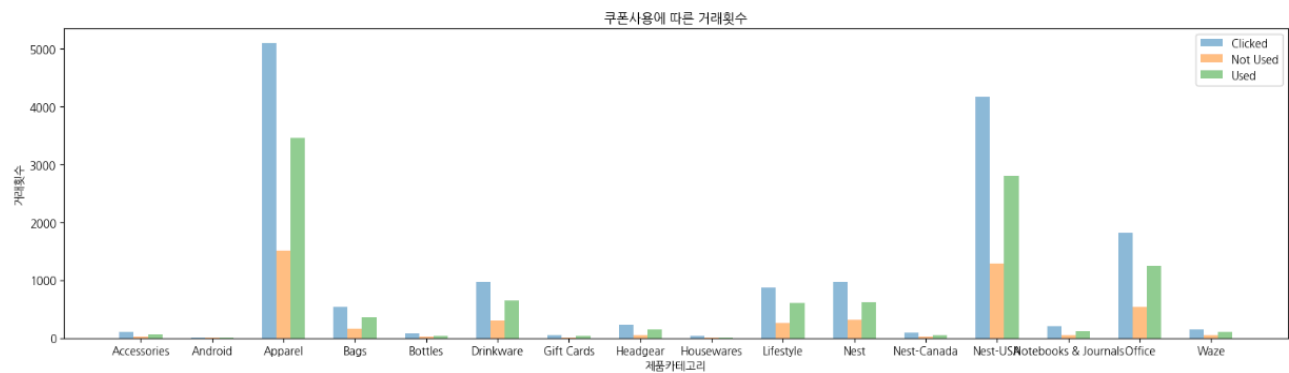
시간대(월)별 판매량이다. 4월, 8월달에는 전체적으로도 올랐지만 Apparel과 Notebooks & Journals의 판매량이 증가한걸 봐서 '학기의 시작'이 매출에 영향을 미친 것을 알 수 있었다. 또한 거꾸로, 5월달의 Office 판매량이 4월에 비해 판매량이 절반으로 줄어든 것을 봐서 4월달의 '연말 신고 기간'이 끝난 후, 새로운 예산을 절약하기 위함이라고 생각된다. 따라서 4월달의 소비가 늘어난 것에 대해 '연말 신고 기간'도 영향을 미친다고 추측할 수 있다.



특이하게도 8월부터 12월까지는 판매량은 줄지만 매출은 늘어난다. 이익을 많이 낼 수 있는 'NEST' 계열의 제품들의 판매량을 확인해보니 판매량이 증가한것으로 보였다



시간대(요일)별 판매량이다. 수요일부터 살펴보면 판매수가 적지만 매출이 금요일 수준으로 높은 것은 역시 NEST-USA 제품의 영향이 크다. 추측하건데 NEST 제품들은 설치가 필요로하는 제품들이기 때문에 설치기사가 집을 방문해야한다. 따라서 물건이 주말에 도착해야하므로 수요일날 주문을 하는 것으로 보인다. 금요일은 수익률이 높은 Apparel 제품이 많이 판매되기 때문에 매출도 높은 것으로 추정된다.

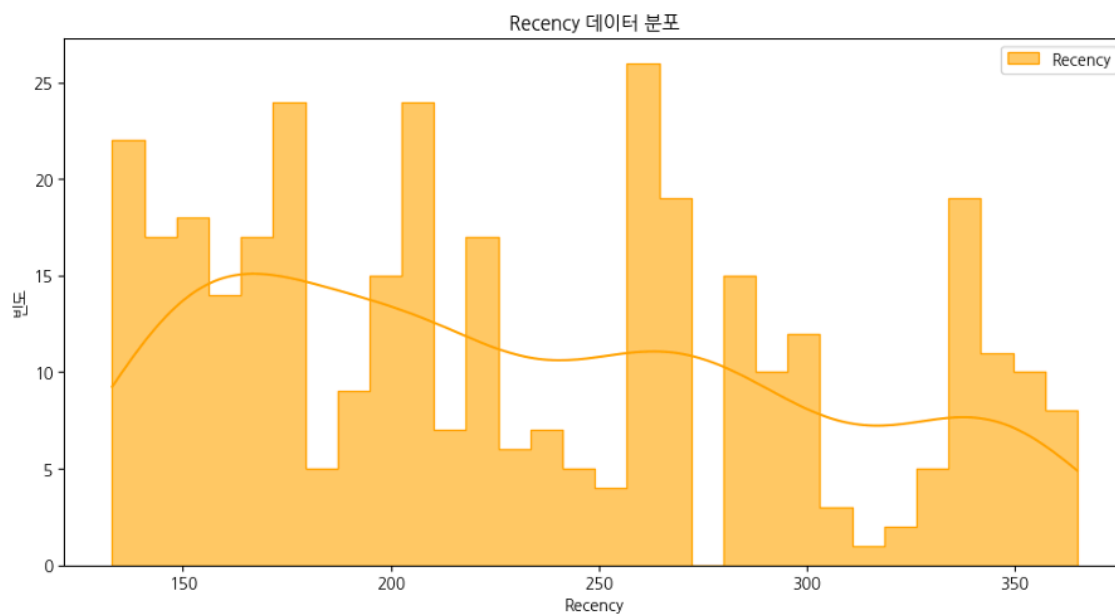


쿠폰상태별 판매량이다. Clicked에서 Used로 전환이 중요하므로, 고객이 실제로 쿠폰을 써서 구매한 물건과, 쿠폰을 쓰려고 했으나 어떤 이유로 인해 쿠폰을 쓰지 않고 구매한 물건을 살펴봐야한다. 살펴보니 제품목록은 동일했고, 해당 제품들에 대하여 쿠폰의 범용성을 늘려서 고객 충성도를 올려야한다. 예를 들면 요즘 쿠폰은 '특정 금액 이상 구매시', '특정 브랜드 구매시' 같은 조건들이 붙어있기 때문에 쿠폰이 있다고해도 못쓸 수 있으므로, 그런 제한 없이 유연하게 쓸 수 있는 쿠폰을 제공한다면 쿠폰 사용량이 늘 것으로 예상된다. 달리 쿠폰을 안쓰는 고객들에게는 마케팅을 더 하여 충성도를

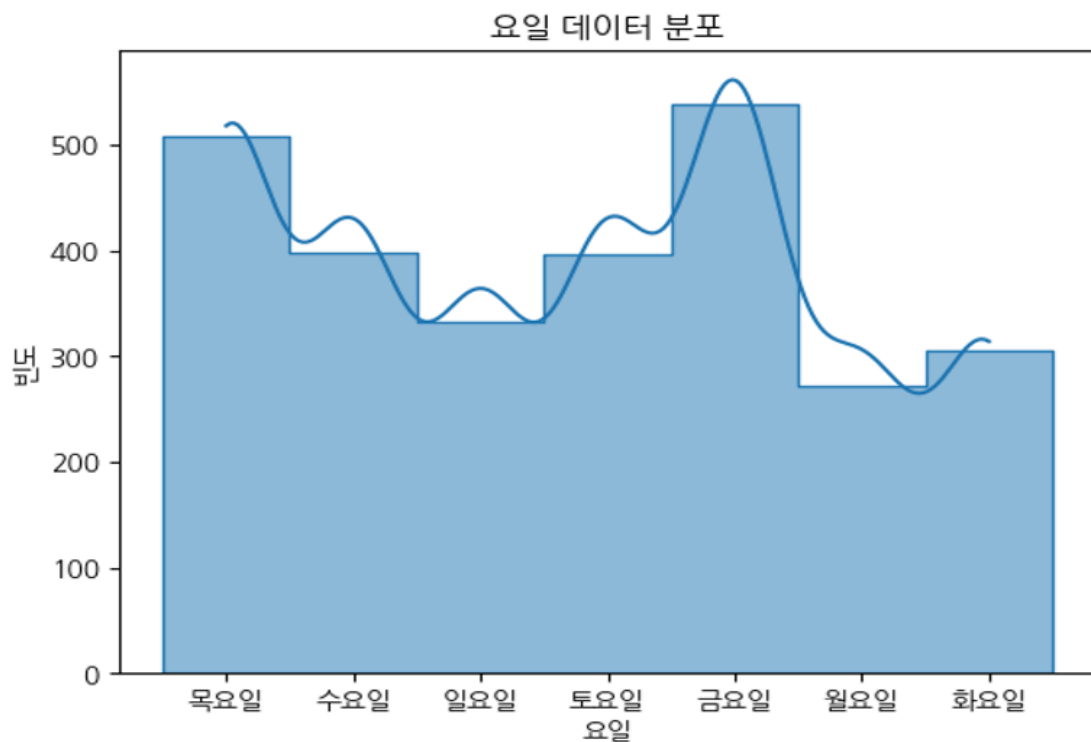
늘리는 방법을 사용하면 좋을 것 같다.

IV. 이탈 위험 고객

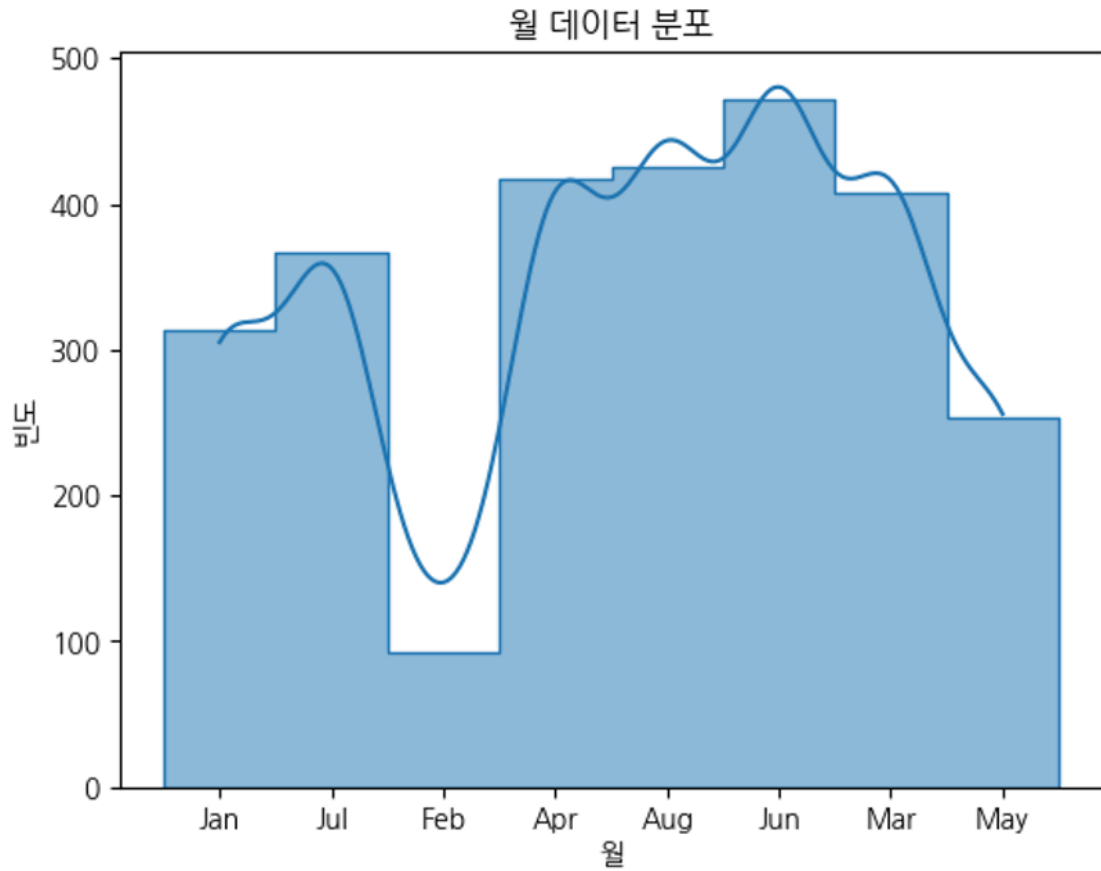
이탈위험의 고객은 이커머스 거래를 장기간 하지 않는 경우로, 이들의 최근 거래 기간과 이들이 거래한 제품에 대해 분석해야 한다.



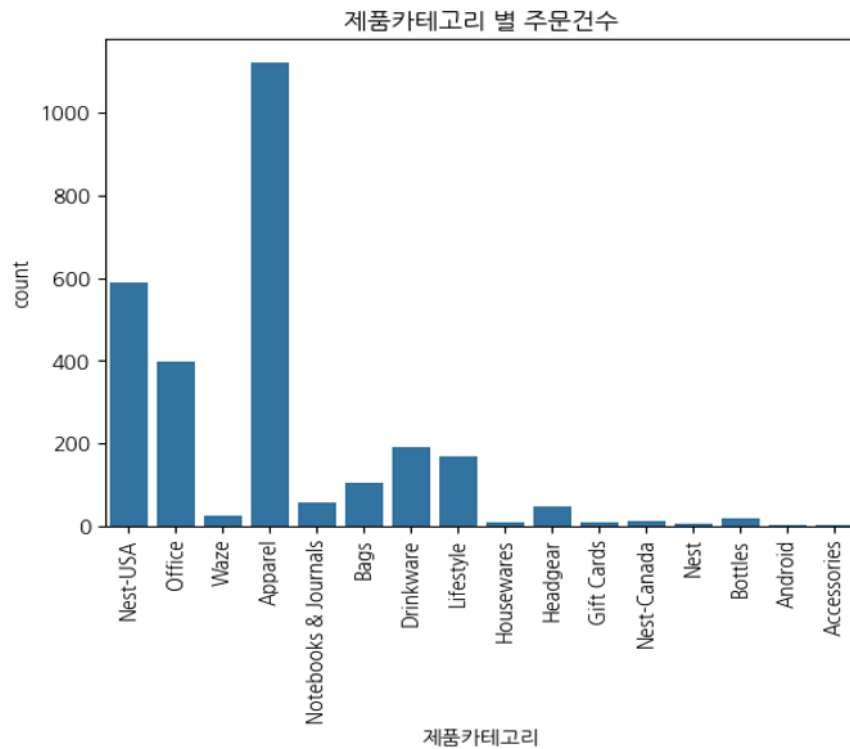
최근 거래 일이 일년 정도 되가는 고객이 상대적으로 많은 편이며, 제일 작은 값이 133일이다. 4개월동안 이 고객들은 거래를 하지 않은 것으로 확인된다.



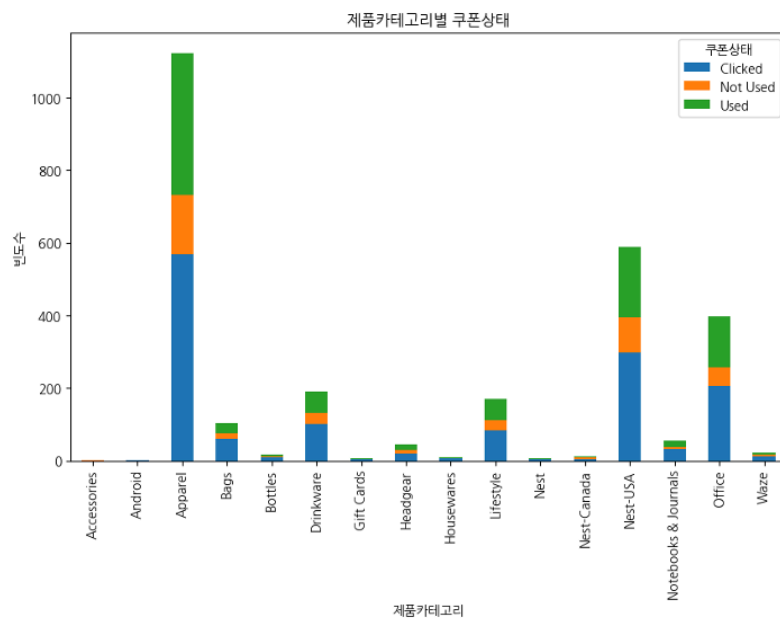
일반고객이 일요일, 월요일에 구매를 많이 하는 반면 이탈위험 고객의 경우 구매량이 적어지는 경향을 보인다.



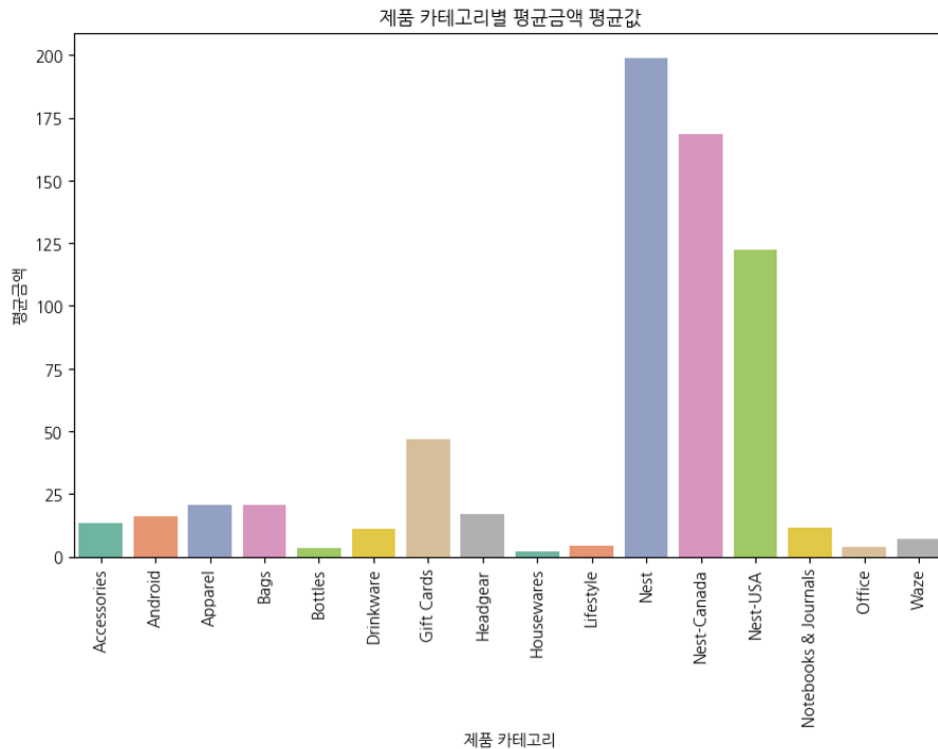
이탈위험 고객의 경우 3,4,5,6,8월에 구매를 많이 한 것으로 보인다. 특이점은 11월, 12월에 구매를 하지 않았다. 이들에게 11월, 12월은 블랙프라이데이, 크리스마스 등 특별한 날이 포함된 달이므로 고객의 구매량이 증가하는 달이다. 고객의 구매량이 증가하는 만큼 행사 상품도 많고, 할인도 많이 들어가는 경우, 마케팅을 많이 하기 때문에 이탈위험 고객의 마음을 사로잡을 마케팅 전략이 필요해 보인다.



이탈위험 고객 또한 Apparel 제품을 가장 많이 구매했고, 뒤를 이어 Nest, Office 제품을 많이 구매했다.



이탈위험 고객이 가장 많이 구매한 제품인 Apparel, Nest, Office 그 외 제품 모두 쿠폰 사용률이 낮은 것을 확인할 수 있다.



이탈위험 고객이 가장 많이 구매한 제품 Apparel, Nest의 경우 가격대가 높은 품목에 해당한다. 이 제품을 구매할 경우 총 구매 금액이 증가하기 때문에 이 제품들을 타겟으로 마케팅을 세워 이탈위험 고객을 붙잡는 방안을 찾을 필요가 있다.

[결론]

이탈위험 고객의 경우 최대 1년, 최소 4개월 가량 거래량이 없는 것으로 확인된다. 이들이 주된 구매 제품은 Apparel, Nest인 것을 확인했다. 특히 Nest의 경우 제품의 가격이 고가에 해당되기에 이를 많이 구매하는 이탈위험 고객의 구매를 유도할 필요가 있다.

이탈위험 고객의 경우 특이하게 11,12월에 거래한 내역이 없는 것을 확인할 수 있다. Nest 제품의 경우 고가 제품에 해당하기에 이 기간을 노리면 좋은 가격에 좋은 제품을 구매할 수 있다는 인식을 이탈위험 고객에게 심어줄 필요가 있다. 또한 이 제품의 경우 고객의 쿠폰 사용률도 낮기 때문에 이 점도 고객이 인지할 수 있도록 해 구매를 유도할 수 있다. Apparel 제품 또한 겨울 시즌이 제품이 고가인 경우가 많기 때문에 위와 같은 방법을 제시할 필요가 있다.

이 고객의 경우 봄-여름에 구매를 많이 했으므로 마케팅을 할 때 이점을 참고하면 좋을 거 같다. 특히 오프라인 마케팅의 경우 날씨와 계절감을 이용하여 팝업스토어 마케팅을 진행하면 좋을 것으로 보인다. 이 마케팅은 고객이 체험 뿐만 아니라 좋은 추억이 될 수 있도록 하여 장기간 이 제품을 생각하고 구매할 수 있도록 유도할 수 있을 것 같다.