

프로그래머스 최종프로젝트 보고서

클러스터 적용 후 사용할 df

	고객ID	거래ID	거래날짜	제품ID	제품카테고리	수량	평균금액	배송료	쿠폰상태	구매금액	...	월	요일	성별	고객지역	가입기간	활인율	오프라인비용	온라인비용	GST	Cluster
0	USER_1358	Transaction_0000	2019-01-01	Product_0981	Nest-USA	1	153.71	6.5	Used	153.71	...	Jan	화요일	남	Chicago	1.000000	10	4500	2424.50	0.10	1
1	USER_1358	Transaction_0001	2019-01-01	Product_0981	Nest-USA	1	153.71	6.5	Used	153.71	...	Jan	화요일	남	Chicago	1.000000	10	4500	2424.50	0.10	1
2	USER_1358	Transaction_0013	2019-01-01	Product_0983	Nest-USA	2	122.77	6.5	Not Used	245.54	...	Jan	화요일	남	Chicago	1.000000	10	4500	2424.50	0.10	1
3	USER_1358	Transaction_0015	2019-01-01	Product_0984	Nest-USA	1	81.50	6.5	Clicked	81.50	...	Jan	화요일	남	Chicago	1.000000	10	4500	2424.50	0.10	1
4	USER_1358	Transaction_0016	2019-01-01	Product_0981	Nest-USA	1	153.71	6.5	Clicked	153.71	...	Jan	화요일	남	Chicago	1.000000	10	4500	2424.50	0.10	1
...
52519	USER_0175	Transaction_23678	2019-12-14	Product_0994	Nest	1	355.74	6.5	Used	355.74	...	Dec	토요일	여	California	3.916667	30	4000	3434.31	0.05	0
52520	USER_0175	Transaction_23678	2019-12-14	Product_0996	Nest	2	49.95	6.5	Clicked	99.90	...	Dec	토요일	여	California	3.916667	30	4000	3434.31	0.05	0
52521	USER_0972	Transaction_23922	2019-12-16	Product_0994	Nest	1	355.74	6.5	Clicked	355.74	...	Dec	월요일	여	New York	3.416667	30	4000	3116.98	0.05	0
52522	USER_0972	Transaction_23922	2019-12-16	Product_0996	Nest	5	49.95	6.5	Not Used	249.75	...	Dec	월요일	여	New York	3.416667	30	4000	3116.98	0.05	0
52523	USER_0847	Transaction_19111	2019-10-17	Product_0990	Nest	1	199.00	6.0	Used	199.00	...	Oct	목요일	여	California	2.250000	10	2500	1783.56	0.05	0

그 중 사용할 칼럼들 선정 (고객ID / 제품카테고리 / 수량 / 배송료 / 쿠폰상태 / 구매금액 / 초종말 / 월 / 요일 / 성별 / 고객지역 / 오프라인비용 / 온라인비용 / Cluster) + count를 위한 거래횟수 칼럼추가

	고객ID	제품카테고리	수량	평균금액	배송료	쿠폰상태	구매금액	초종말	월	요일	성별	고객지역	오프라인비용	온라인비용	Cluster	거래횟수
699	USER_1277	Nest-USA	1	122.77	6.5	Clicked	122.77	초	Jan	목요일	남	Chicago	4500	1576.38	3	1
700	USER_1277	Nest-USA	2	153.71	6.5	Used	307.42	초	Jan	목요일	남	Chicago	4500	1576.38	3	1
701	USER_1277	Nest-USA	1	122.77	6.5	Not Used	122.77	초	Jan	목요일	남	Chicago	4500	1576.38	3	1
702	USER_1277	Office	1	2.05	6.5	Clicked	2.05	초	Jan	목요일	남	Chicago	4500	1576.38	3	1
703	USER_1277	Office	1	2.05	6.5	Clicked	2.05	초	Jan	목요일	남	Chicago	4500	1576.38	3	1
...
52513	USER_1220	Lifestyle	2	2.80	6.0	Clicked	5.60	중	Apr	목요일	여	California	4000	1476.58	3	1
52514	USER_0564	Waze	1	5.59	6.0	Clicked	5.59	중	May	토요일	여	California	3000	1801.66	3	1
52515	USER_0633	Headgear	1	15.19	6.5	Not Used	15.19	초	Apr	일요일	여	Chicago	2500	2719.46	3	1
52516	USER_0186	Headgear	1	9.99	6.0	Clicked	9.99	중	Jul	금요일	여	New York	2500	923.40	3	1
52517	USER_0048	Headgear	1	10.63	6.0	Clicked	10.63	말	Jul	토요일	여	Chicago	2500	1151.70	3	1

2747 rows x 16 columns

분석해볼 두 데이터프레임 (Premium / Churn)

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 30371 entries, 293 to 50027
Data columns (total 16 columns):
#   Column      Non-Null Count  Dtype
---  -
0   고객ID      30371 non-null object
1   제품카테고리 30371 non-null object
2   수량        30371 non-null int64
3   평균금액    30371 non-null float64
4   배송료      30371 non-null float64
5   쿠폰상태    30371 non-null object
6   구매금액    30371 non-null float64
7   초종말      30371 non-null object
8   월          30371 non-null object
9   요일        30371 non-null object
10  성별        30371 non-null object
11  고객지역    30371 non-null object
12  오프라인비용 30371 non-null int64
13  온라인비용  30371 non-null float64
14  Cluster     30371 non-null int64
15  거래횟수     30371 non-null int64
dtypes: float64(4), int64(4), object(8)
memory usage: 3.9+ MB
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2747 entries, 699 to 52517
Data columns (total 16 columns):
#   Column      Non-Null Count  Dtype
---  -
0   고객ID      2747 non-null object
1   제품카테고리 2747 non-null object
2   수량        2747 non-null int64
3   평균금액    2747 non-null float64
4   배송료      2747 non-null float64
5   쿠폰상태    2747 non-null object
6   구매금액    2747 non-null float64
7   초종말      2747 non-null object
8   월          2747 non-null object
9   요일        2747 non-null object
10  성별        2747 non-null object
11  고객지역    2747 non-null object
12  오프라인비용 2747 non-null int64
13  온라인비용  2747 non-null float64
14  Cluster     2747 non-null int64
15  거래횟수     2747 non-null int64
dtypes: float64(4), int64(4), object(8)
memory usage: 364.8+ KB
```

커다란 함수를 만들어서 해당 데이터프레임 안에서 칼럼끼리의 상관관계를 분석할 수 있도록 만들었다 (GetInfo(dataframe, column)).

함수에 데이터프레임, 타겟칼럼을 집어넣으면 GetInfo(df, col)

1. 판매수량, 거래횟수, 구매금액의 합계값을 가진 데이터프레임
 2. 구매금액, 배송료, 오프라인비용, 온라인비용의 평균값을 가진 데이터프레임
 3. 타겟칼럼의 고유 고객 수에 대한 데이터프레임
 4. 타겟칼럼의 쿠폰상태 값(sum)을 가진 데이터프레임
 5. 각 제품카테고리별, 타겟칼럼의 매출, 판매수를 나타낸 데이터프레임
 6. 시간대별 타겟칼럼의 구매금액 값(sum)을 가진 데이터프레임
 7. 요일별 타겟칼럼의 거래횟수에 대한 데이터프레임
- 등이 있다

각 데이터프레임들을 통해 알 수 있는 인사이트들을 찾아보자

데이터 셋은 Premium 고객들로, 방문율도 높고 자주 구매하고, 높은 실적을 남기는 고객들이다.

일단 비교적 간단한 데이터프레임(1)을 예시로 들어 타겟칼럼에 대한 것을 설명하고 앞으로의 데이터프레임들 (2~7)에 대해서는 한가지 예시 (고객지역)를 들면서 결과를 설명하도록 할 것이다.

1. 판매수량, 거래횟수, 구매금액의 합계값

해당 데이터프레임은 그래프보단 숫자로 보는 것이 편할 것 같아 단순 데이터프레임만 출력하도록 했다.

	총 판매수량	총 거래횟수(중복포함)	구매금액
성별			
남	50824	11404	1097818.69
여	85265	18967	1781522.01

(target column = 성별)

	총 판매수량	총 거래횟수(중복포함)	구매금액
고객지역			
California	40642	8838	877981.04
Chicago	55059	11465	1080225.34
New Jersey	11583	2636	250909.92
New York	23074	5872	526673.54
Washington DC	5731	1560	143550.86

(target column = 고객지역)

	총 판매수량	총 거래횟수(중복포함)	구매금액
초중말			
초	41975	9031	866206.58
중	48649	10404	982719.97
말	45465	10936	1030414.15

(target column = 초중말)

	총 판매수량	총 거래횟수(중복포함)	구매금액
월			
Jan	5533	1132	110784.73
Feb	6100	1517	137081.28
Mar	8603	1909	158505.09
Apr	14561	1535	202430.06
May	7378	1847	126075.84
Jun	10699	1902	150411.58
Jul	11904	2568	185607.85
Aug	16641	3552	249466.09
Sep	16051	3690	316531.83
Oct	14966	3567	352156.82
Nov	12664	3524	455510.78
Dec	10989	3628	434778.75

(target column = 월)

	총 판매수량	총 거래횟수(중복포함)	구매금액
요일			
월요일	6604	2733	232936.39
화요일	5344	2589	234614.41
수요일	24104	5409	543356.98
목요일	23987	4913	486761.76
금요일	31290	5240	547161.71
토요일	24361	5001	428096.65
일요일	20399	4486	406412.80

(target column = 성별)

	총 판매수량	총 거래횟수(중복포함)	구매금액
제품카테고리			
Accessories	934	183	4724.14
Android	19	19	289.35
Apparel	19246	10072	337361.15
Bags	8210	1070	85254.07
Bottles	1171	145	3872.74
Drinkware	18876	1919	125181.05
Gift Cards	108	103	13651.54
Headgear	2200	431	33068.16
Housewares	1051	56	2118.50
Lifestyle	13682	1744	40851.59
Nest	2462	1904	450173.87
Nest-Canada	230	165	35493.92
Nest-USA	12756	8256	1511819.46
Notebooks & Journals	6643	376	75286.86
Office	47744	3608	156065.11
Waze	757	320	4129.19

(target column = 성별)

	총 판매수량	총 거래횟수(중복포함)	구매금액
쿠폰상태			
Clicked	68585	15408	1450302.76
Not Used	19631	4629	463003.35
Used	47873	10334	966034.59

(target column = 성별)

인덱스에 타겟 칼럼이 들어가고, 그에 따른 데이터들을 볼 수 있다. 이런 방식은 타겟을 바꿔가며 분석을 할 수 있다. 지금은 위에서 예고한대로 '고객지역' 관련해서 분석을 하겠다.

Chicago쪽(중서부)에서 거래가 가장 많았고 [55059], 그 다음은 California(서부) [40642] New Jersey [11583] + New York [23074] (북동부), 마지막으로 Washington DC (남부) [5731]로 판매 수량 순위가 매겨졌다. 한 번의 거래에서 판매되는 수량도 위의 순서대로였지만 [4.8 > 4.6 > 4.0 > 3.6] 오히려 판매된 물품 중에서 판매된 물건의 평균가는 역순을 기록했다 [19.6 < 21.6 < 22.4 < 25.0]

여기서 확실히 거래가 많이 이루어지는 지역 (Chicago)에서는 한번 거래할 때 많은 양을 하 다보니 물건을 저렴하게 구매하는 경향이 있었다

2. 구매금액, 배송료, 오프라인비용, 온라인비용의 평균값

	구매금액	배송료	오프라인비용	온라인비용
고객지역				
California	99.341598	10.676041	3019.891378	1918.275575
Chicago	94.219393	10.246071	2947.893589	1941.955157
New Jersey	95.185857	9.363930	2736.191199	1873.616855
New York	89.692360	9.497745	2836.665531	2008.520131
Washington DC	92.019782	9.678250	2855.833333	1922.453583

여기서의 '구매금액'은 총 금액을 거래횟수로 나눈 값으로, 위에서 계산한 물건당 평균가격과 다르다. 즉 Chicago를 예시로 들면 한번의 '거래'에서는 평균 94.2를 지불했지만, 하나의 '물건' 값은 19.6을 지불한 것이다. 배송료를 봤을 때, 물류창고에서부터 거리가 멀면 배송비가 좀 더 많이 나갈 것이라고 가정해보면 물류창고의 위치도 대략적으로 추정할 수 있다. (Ohio로 생각) California는 온라인비용이 적는데 비해 오프라인비용은 크고 Chicago는 전부 적절하게 비싸고, New Jersey는 모든 비용이 적다. New York은 California랑 반대로 오프라인비용보단 온라인비용이 많이 나갔다. Washington DC는 딱 중간이다. 따라서 California / New York 에서는 온/오프라인으로 광고하면 효과적으로 손님을 끌어들이 수 있고, 나머지 지역들은 판매방식에 따른 이득을 보진 못할 것 같다.

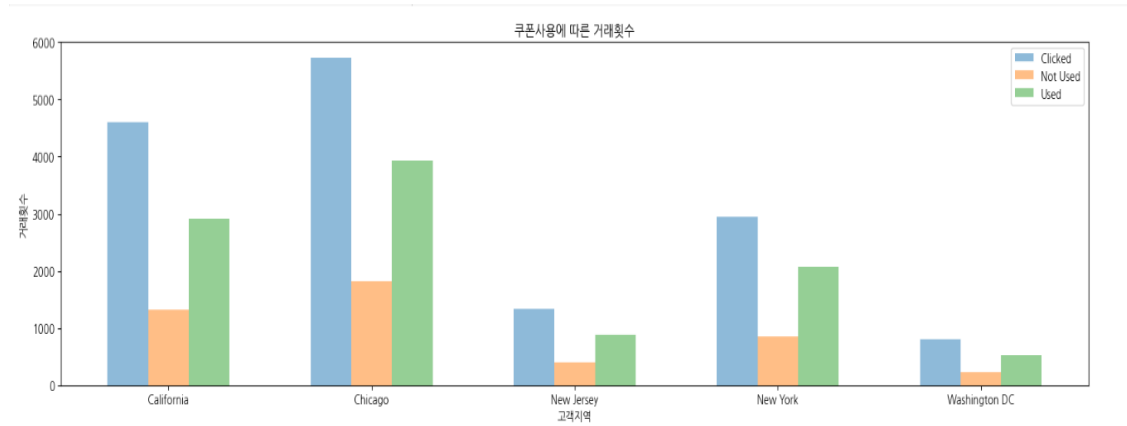
3. 타겟칼럼의 고유 고객수

	거래횟수
고객지역	
California	108
Chicago	107
New Jersey	33
New York	89
Washington DC	15

해당 지역에서 거래한 고유 유저 수도 거래량에 비례하지만, 고객 1명당 사용한 평균 금액은 각각 [8129, 10095, 7603, 5917, 9570]로 New York에서는 평균적으로 적은 사용금액을 가졌고 Washington DC도 눈에 띄게 높은 수치를 나타냈다. 이는 New York은 거래하는 사람이 많아서, Premium 고객층에서 그나마 낮은 분들이 계시고, Washington DC지역은 그 반대로 몇몇의 whale 유저들이 지역의 매출을 담당하고 있다고 볼 수 있다.

4. 타겟칼럼의 쿠폰상태

이젠 그래프를 통해 알아보자

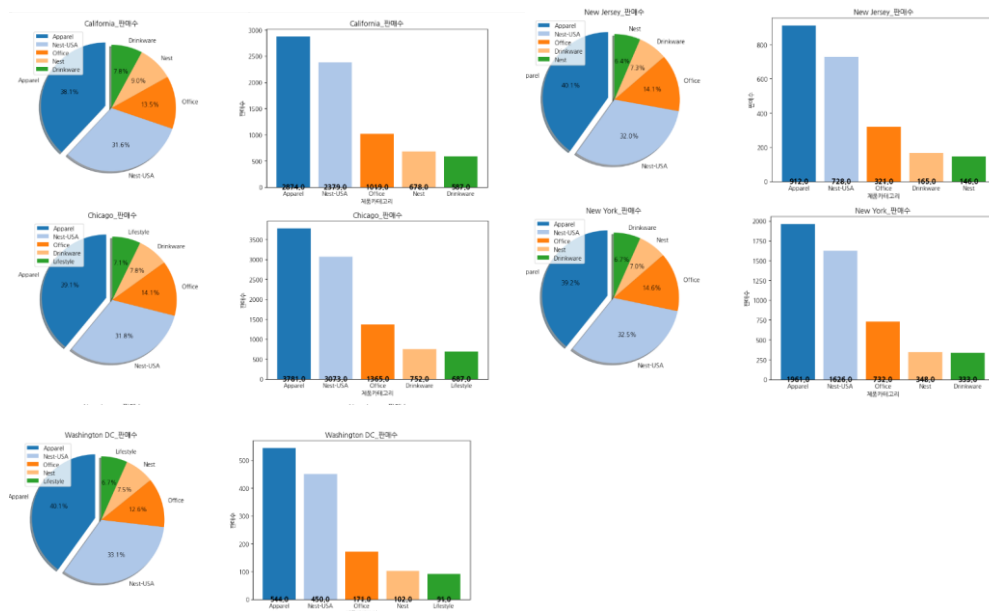


쿠폰사용에 따른 거래횟수 그래프이다.

모든 지역이 비슷한 양상을 보이고 있다 (Clicked > Used > Not Used) 일단 지금 단계에서는 더 이상 발견할 인사이트는 없는 것 같다.

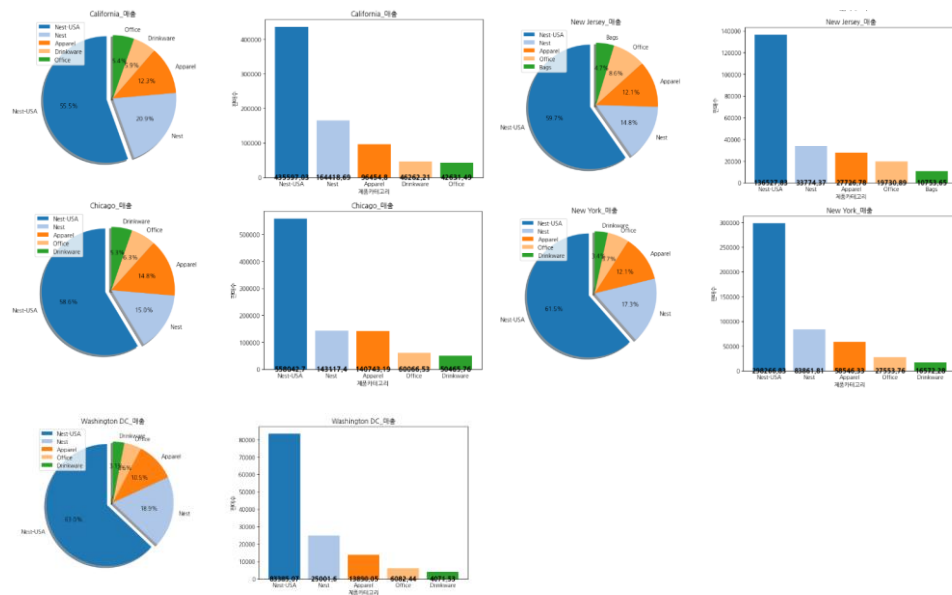
5. 타겟칼럼의 각 제품카테고리에 대한 판매수,매출

5-1. 판매수



지역별 제품 카테고리에 대한 판매 수 그래프이다. 가장 많은 판매량을 가진 top 5개의 데이터만을 사용해서 그래프로 나타내었다. 모든 지역에 공통적으로 Apparel, Nest-USA, Office를 많이 구매하였고, Nest, Drinkware, Lifestyle에 관한 소비가 지역마다 4~5등을 다르게 차지했다.

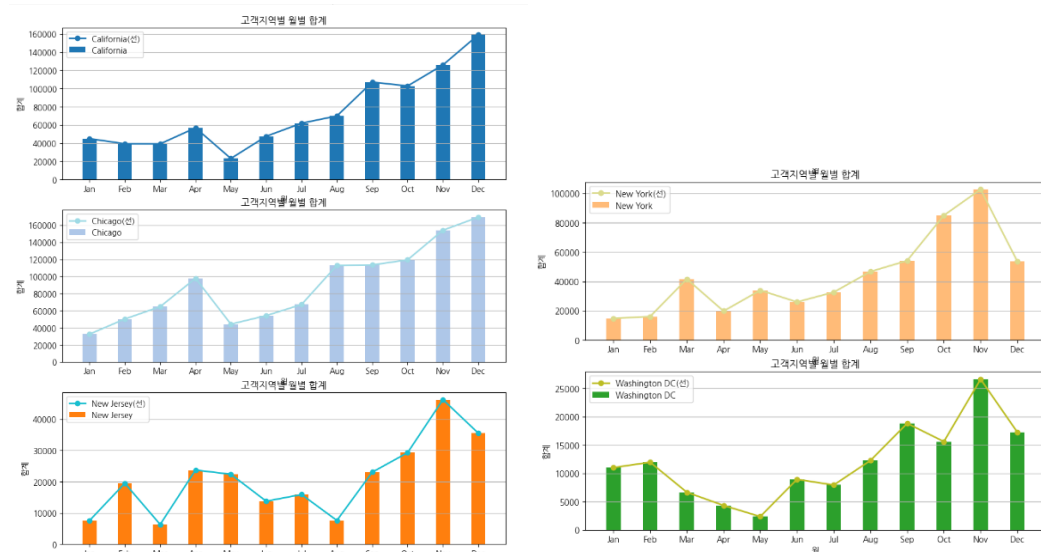
5-2. 매출



지역별 제품카테고리에 대한 매출 그래프이다. 특이하게도 가장 많은 판매량을 기록한 Apparel이 매출에 관해서는 3위를 차지하고 Nest-USA가 판매량은 2위, 매출은 압도적 1위 인 것을 볼 수 있었다. 매출 2위는 Nest로 해당 제품의 판매량이 상위 5개 안에 들지 않아도 매출로 상위권을 찍을 만큼 이익을 많이 낸다고 볼 수 있다. (적게 팔아도 매출이 높음) 그 외에는 Drinkware 또한 판매량 상위권에 속한만큼 매출도 상위권을 담당하는데, New Jersey에서는 Bags가 Drinkware대신 자리를 차지했다.

6. 타겟칼럼의 각 시간대에 대한 구매금액

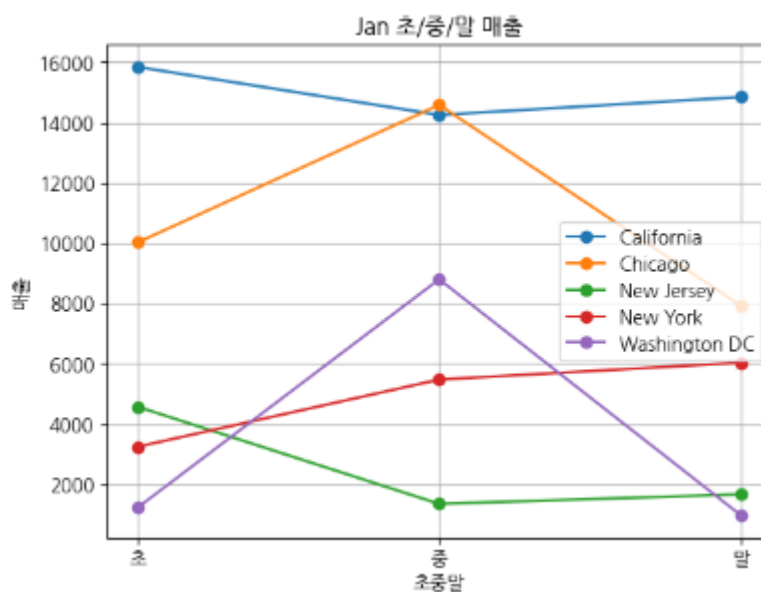
6.1 월별



지역별 월별 매출 추이다. Chicago나 California은 시간의 흐름에 따라 매출량이 증가했고, 5월달(May)에 한풀 꺾였다가 다시 증가했다. 전체적으로 봤을 때 11월(Nov)에서 매출이 높으므로 해당 기간에 마케팅을 진행하면 좋을 것 같다. 추가적인 정보들은 딱히 알 수 없으므로 다른 타겟칼럼을 활용해야 할 것 같다. (제품카테고리)

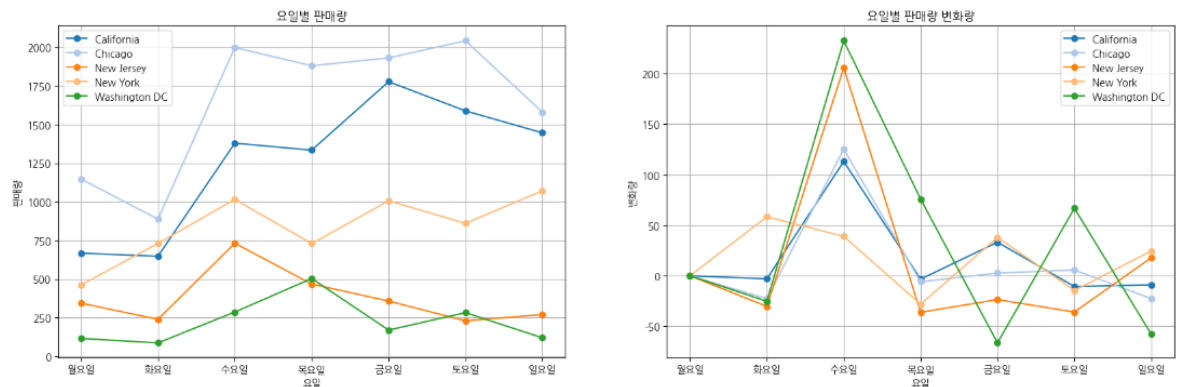
6-2. 초중말

여기는 1일~10일을 상순(초) 11일~20일을 중순(중) 21일 이후를 하순(말)로 나눠서 분기별로 판매량을 확인하기 위해 추가로 만든 칼럼을 활용했다. 해당 사진은 1월 달을 확인했다.



California는 시기에 상관없이 비슷한 소비를 하는 것으로 보이고 Chicago와 Washington DC는 중순에 상대적으로 많은 소비가 이루어지는 것을 볼 수 있다. New York과 New Jersey는 특이하게도 반대 성향을 가지는데 이는 점점 소비가 증가하고 감소하는 것이다. 물론 1월만 봐서 확답은 할 수 없고 다음 데이터들도 봐야 한다.

7. 타겟칼럼의 각 요일에 대한 거래횟수



지역별 요일별 판매량이다. 지역마다 다르지만 대부분 수요일~금요일 사이에 많은 판매가 이루어지므로 해당 기간 동안 마케팅을 하여 소비를 유도하는 전략을 세울 수 있을 것이다.

같은 방식으로 Premium → Churn으로 데이터셋을 바꿔서 진행했다.

1.

고객지역	총 판매수량	총 거래횟수(중복포함)	구매금액
California	3126	834	48977.89
Chicago	3408	790	49073.85
New Jersey	754	251	14864.32
New York	2425	749	45254.24
Washington DC	353	123	5626.04

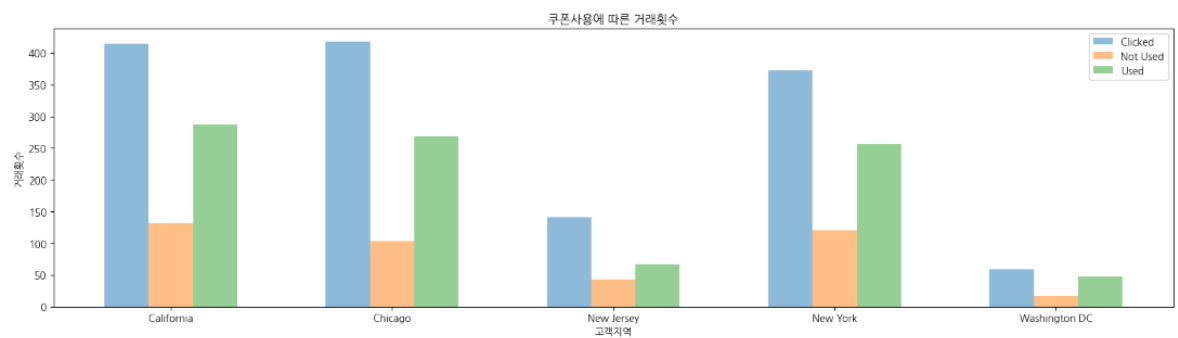
2.

고객지역	구매금액	배송료	오프라인비용	온라인비용
California	58.726487	9.965456	2606.714628	1802.798585
Chicago	62.118797	10.705241	2708.354430	1800.566405
New Jersey	59.220398	8.462988	2757.370518	1814.569442
New York	60.419546	11.300988	2553.404539	1765.458732
Washington DC	45.740163	12.222683	3191.056911	1572.565772

3.

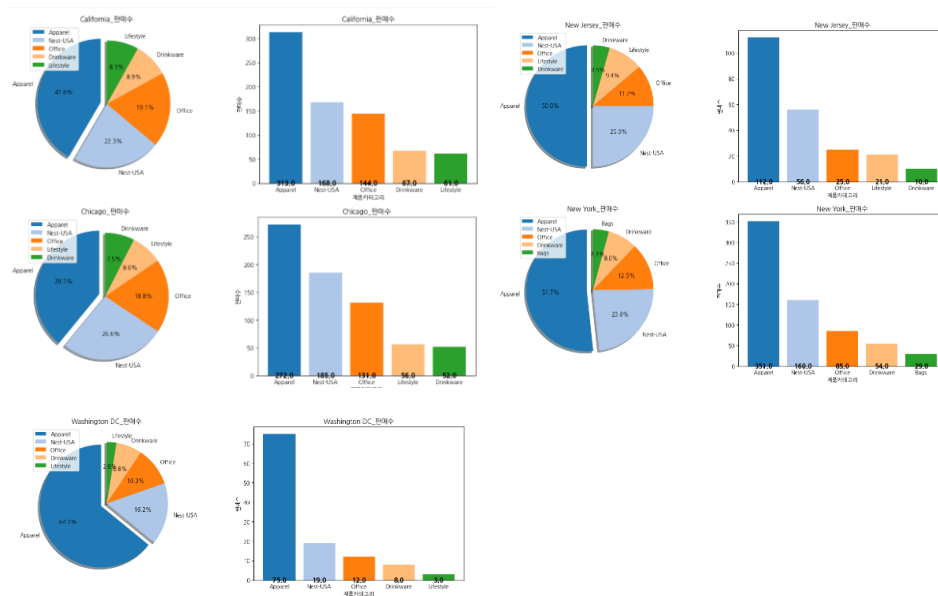
고객지역	거래횟수
California	834
Chicago	790
New Jersey	251
New York	749
Washington DC	123

4.

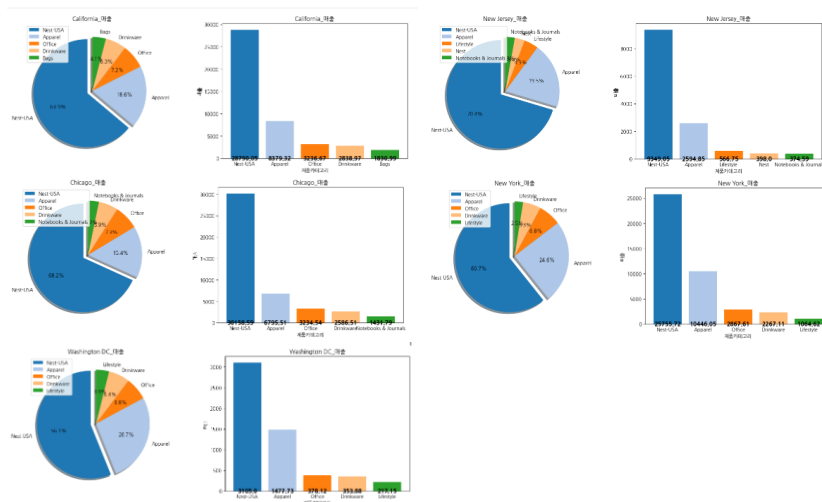


5.

5-1.

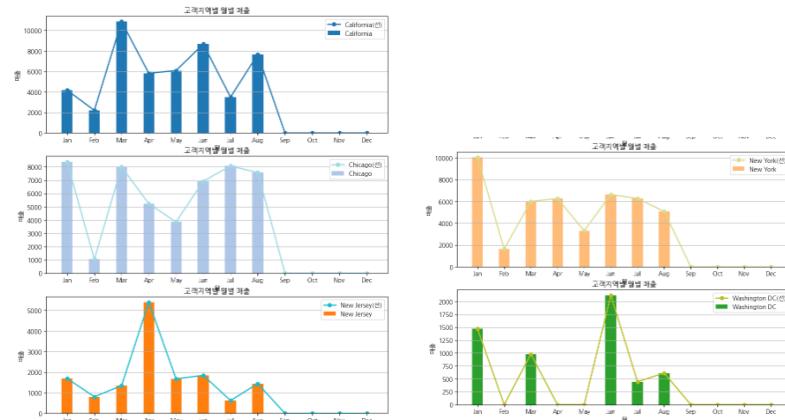


5-2.

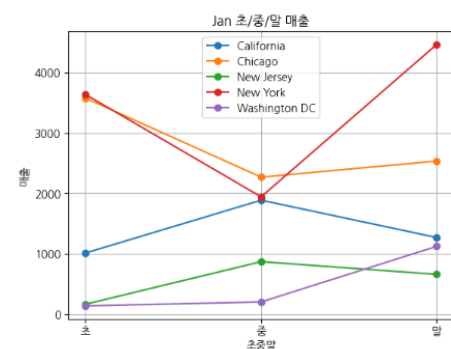


6. .

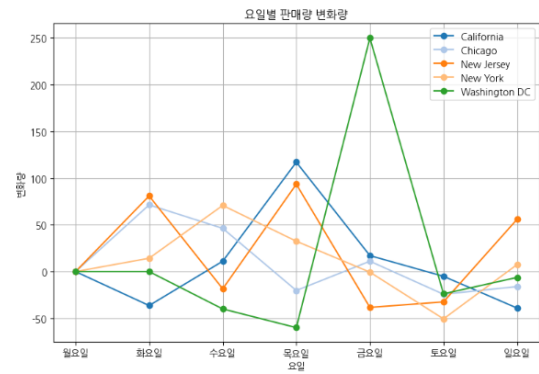
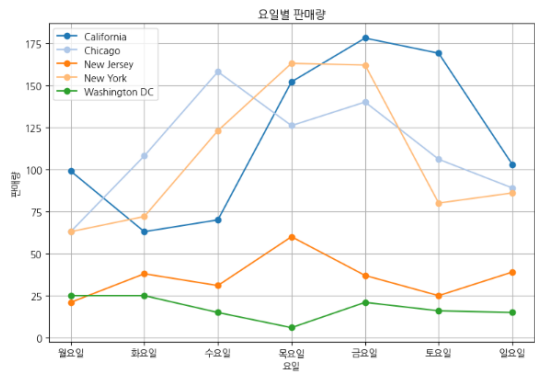
6-1.



6-2.



7. .



한계

아무래도 모든 데이터의 합산값을 가져가므로 유저별 맞춤 추천이 어렵다.

클러스터가 너무 골고루라서 차별점이 크진 않을 것 같다. (고유 고객 수?)

칼럼을 통일시켜야 할듯하다. 거래횟수 <> 매출 (그냥 둘 다 쓰자)

도메인 지식이 부족하여 해석에 한계를 느낀다

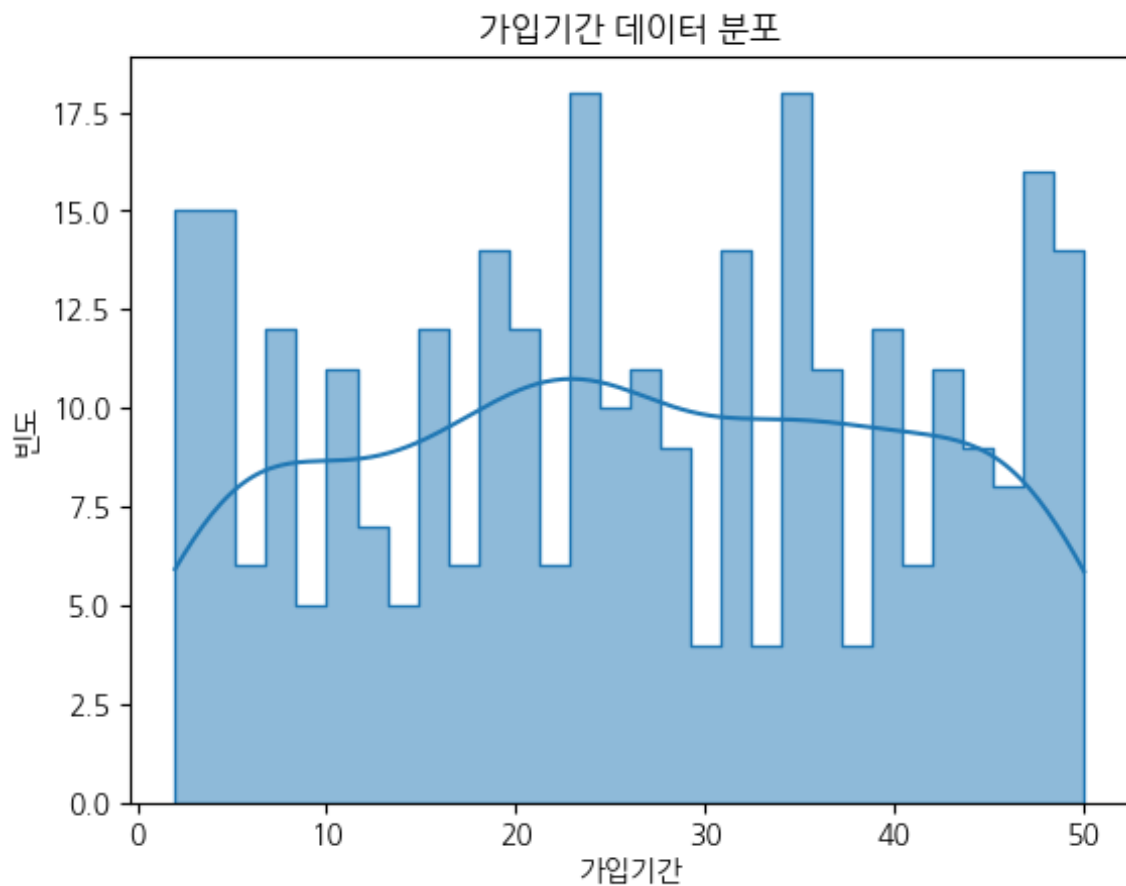
[일반고객]

일반고객의 경우 자주 방문하지만 높은 실적을 남기진 않는 고객이다.

높은 실적을 낼 수 있는 방안을 세워야 함

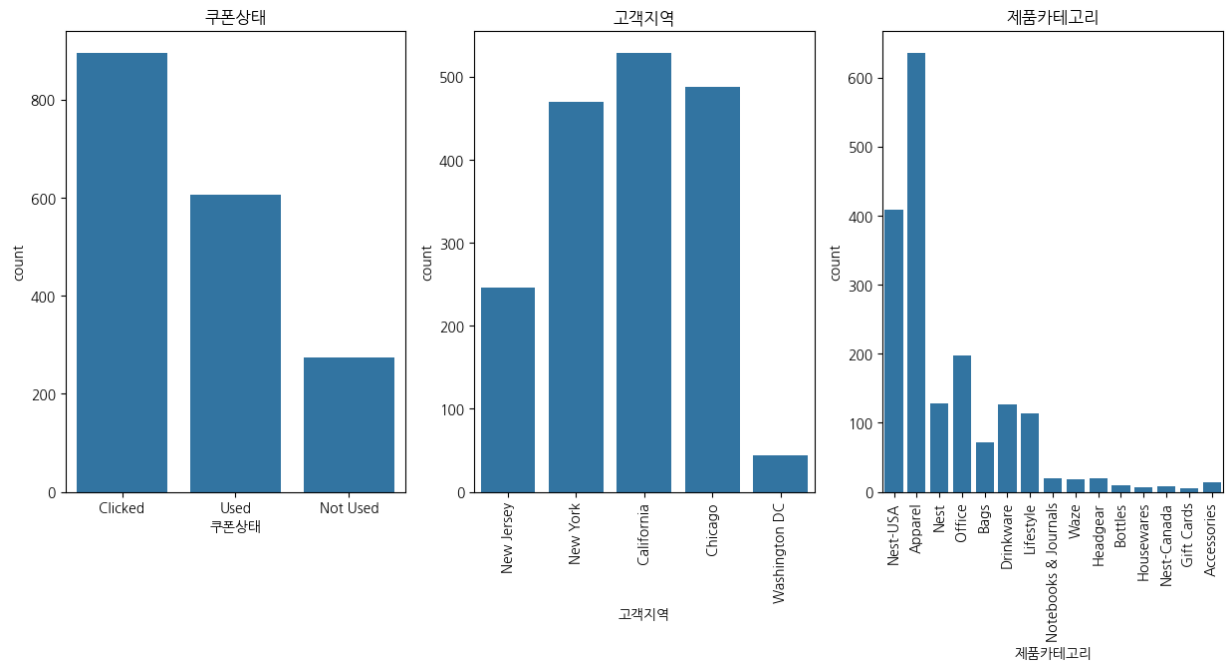
- 제품 구매 빈도 증가 유도
- 고가 제품 구매 유도

일반고객의 가입기간 별 구매빈도 분포를 확인하면 비교적 골고루 분포되어 있으나, 특정 기간에는 감소하는 경향을 띤다.



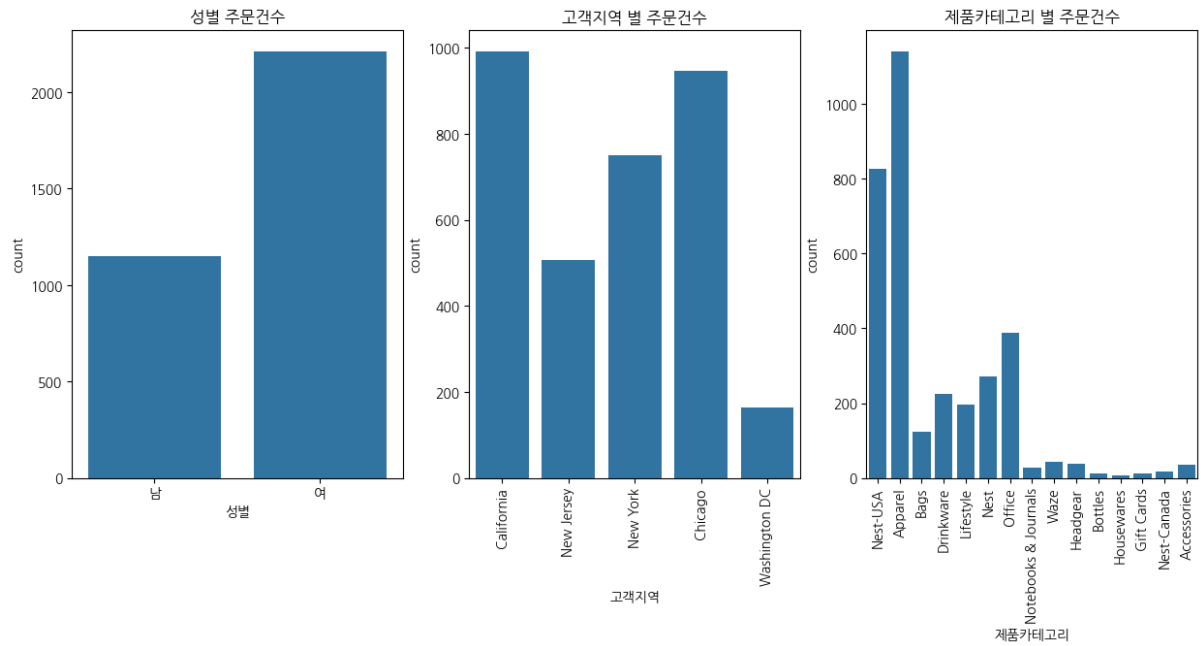
[특정 기간에 따른 일반고객]

10 ~ 14 개월, 21~23 개월, 29~40 개월의 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.

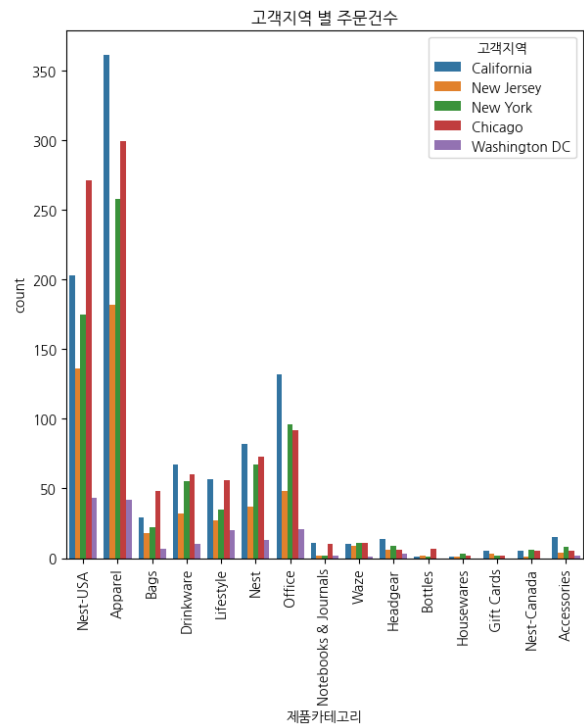
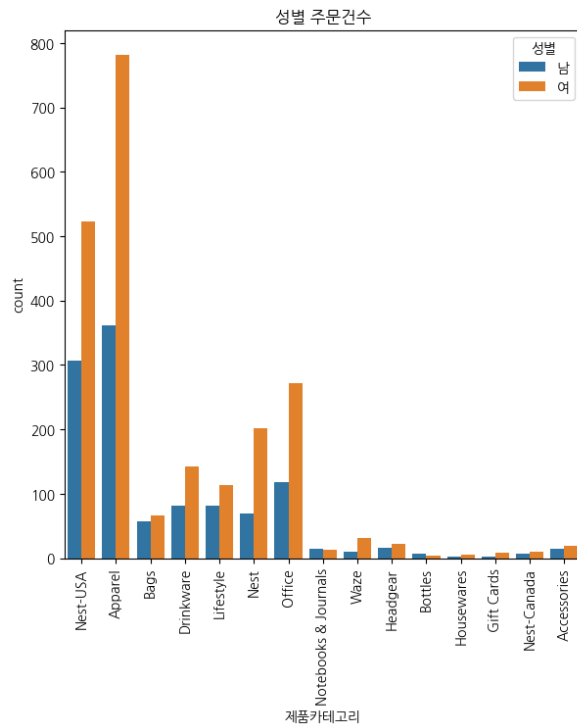
Nest 제품은 스마트 홈 제품 라인으로 마트 스피커, 스마트 디스플레이, 스트리밍 기기, 온도조절기, 연기 감지기, 라우터 및 보안 시스템(스마트 도어벨, 카메라, 스마트 락 등)이 있다. 이 제품들은 상대적으로 고가인 제품으로 이 제품에 관심이 많은 고객의 경우 “쿠폰”을 제공해 구매를 유도할 수 있도록 한다. (할인율이 높은 제품 알림으로 고객의 제품 구매 유도)



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.

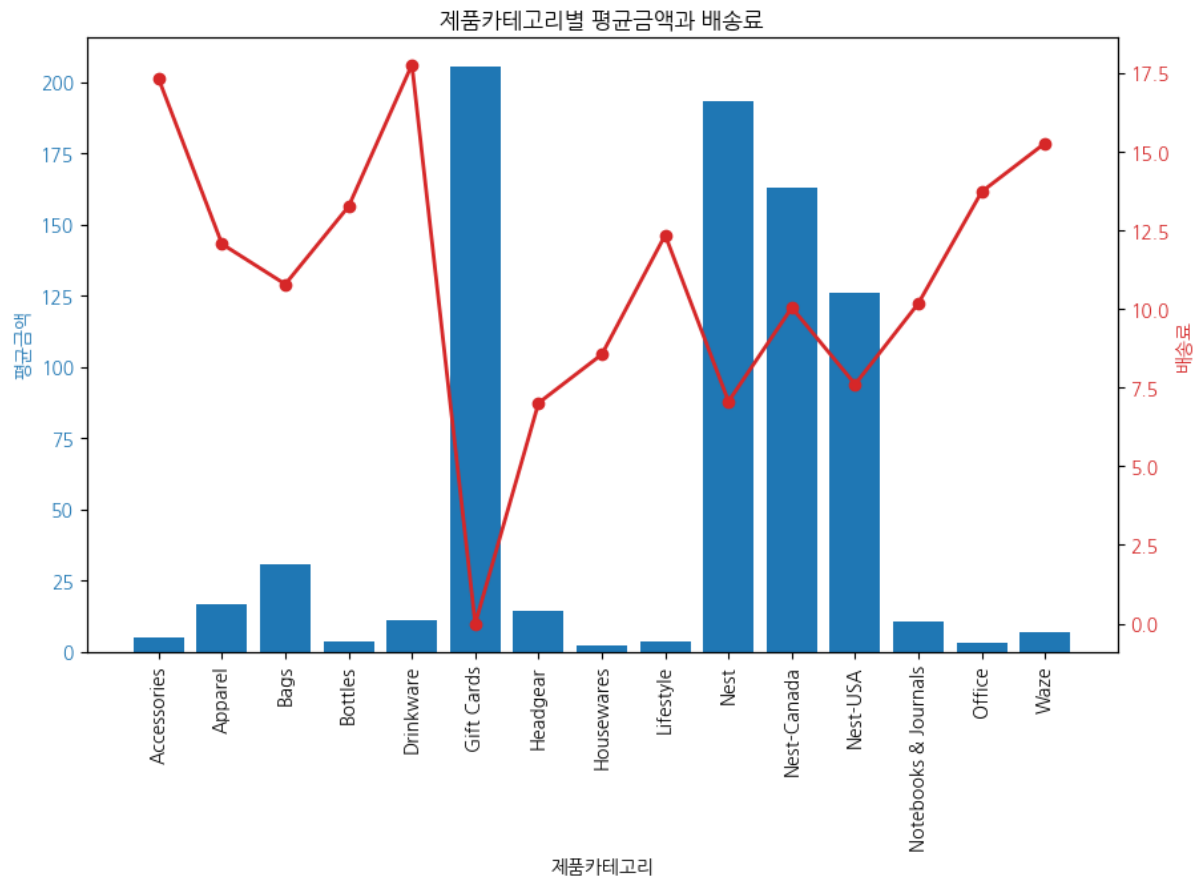
고객지역 별 주문건수는 캘리포니아와 시카고가 비슷하게 높으며 워싱턴 DC의 경우 많이 적은 것을 확인할 수 있다. 지역별 특징이 있는지 확인해봐야 한다. (워싱턴 DC의 경우 상권 발달이 잘 되어 있어 이커머스를 활용하지 않는 것인지 등)



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.

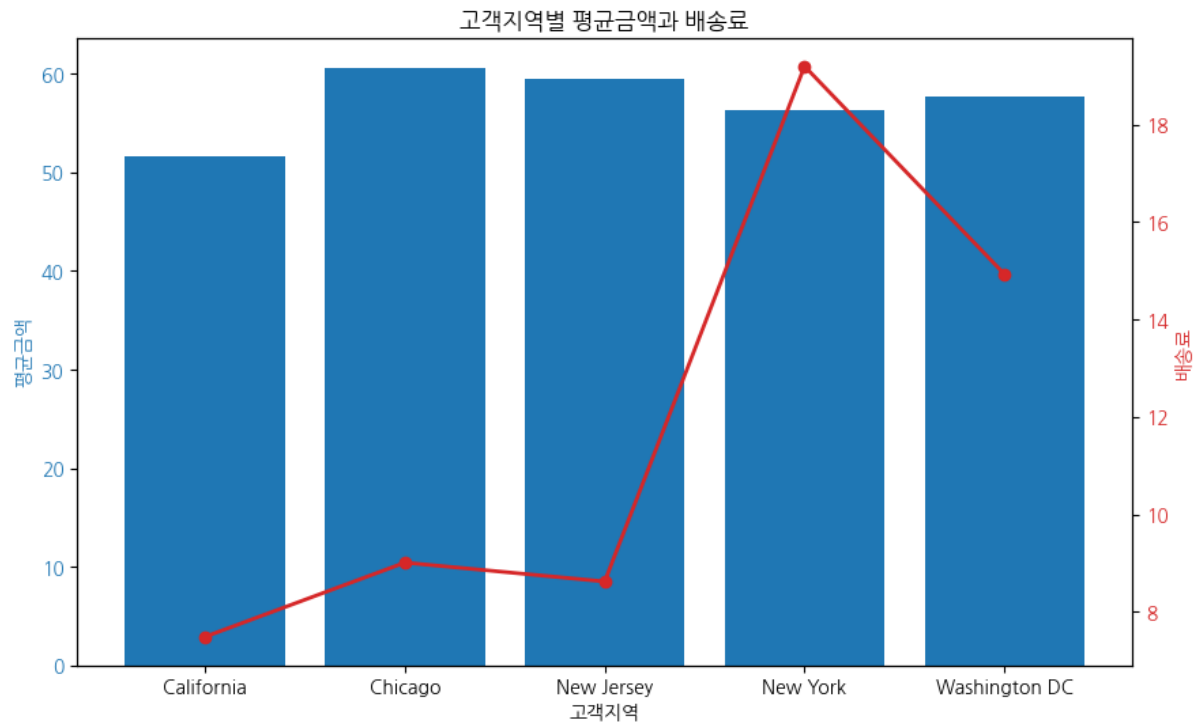
또한 남성의 경우 'Notebooks & Journals'와 'Bottles' 카테고리에서는 여성보다 높은 것으로 관찰된다. 이 카테고리의 제품에 다양성 뿐만 아니라 남성 고객이 선호할 포인트를 찾아 마케팅 전략을 세울 필요가 있어 보인다. 특히 Notebook 의 경우 고가 제품으로 마케팅 전략을 잘 세운다면 높은 실적을 세울 수 있을 것으로 보인다.



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 12 월 등)에는 배송료 '무료' 또는 '할인' 등의 이벤트를 진행하면 좋을 것으로 보인다.

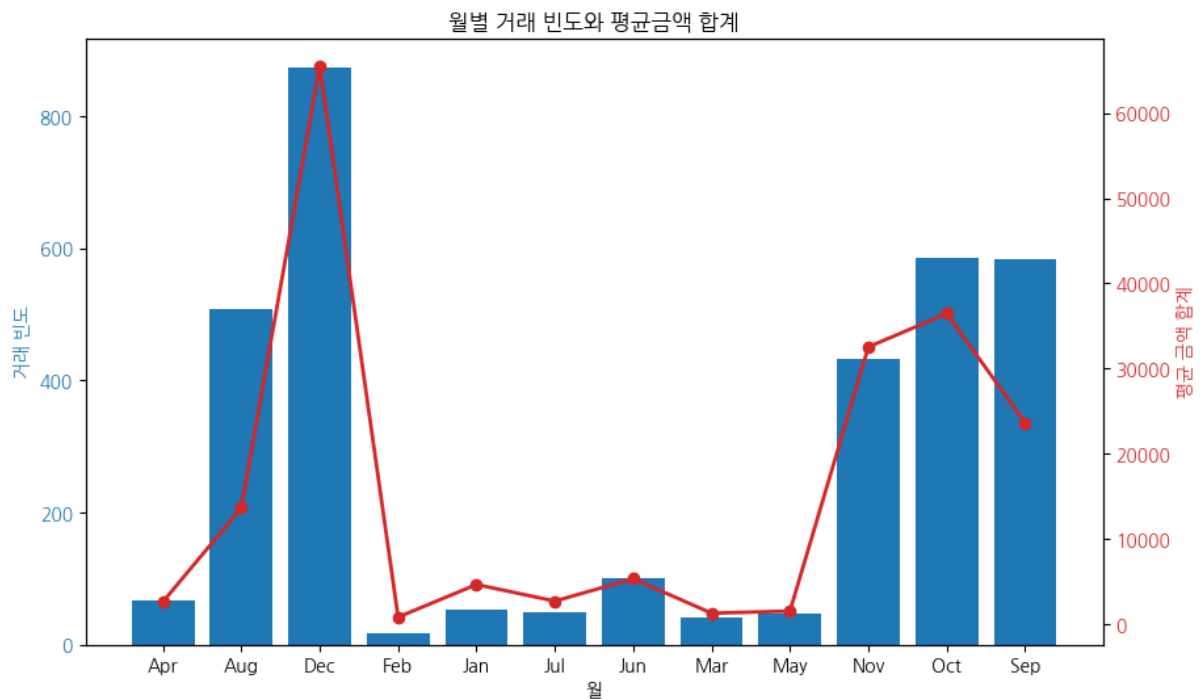
또한 Apparel 제품의 경우 주문건수가 굉장히 많은 제품이지만 배송료가 비싼 점이 고객의 제품 구매에 있어 걸림돌이 될 수도 있다는 점 또한 인지해야할 필요가 있다. (상대적으로 저가의 제품이 배송료가 높다는 점 인지해야 함)



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.

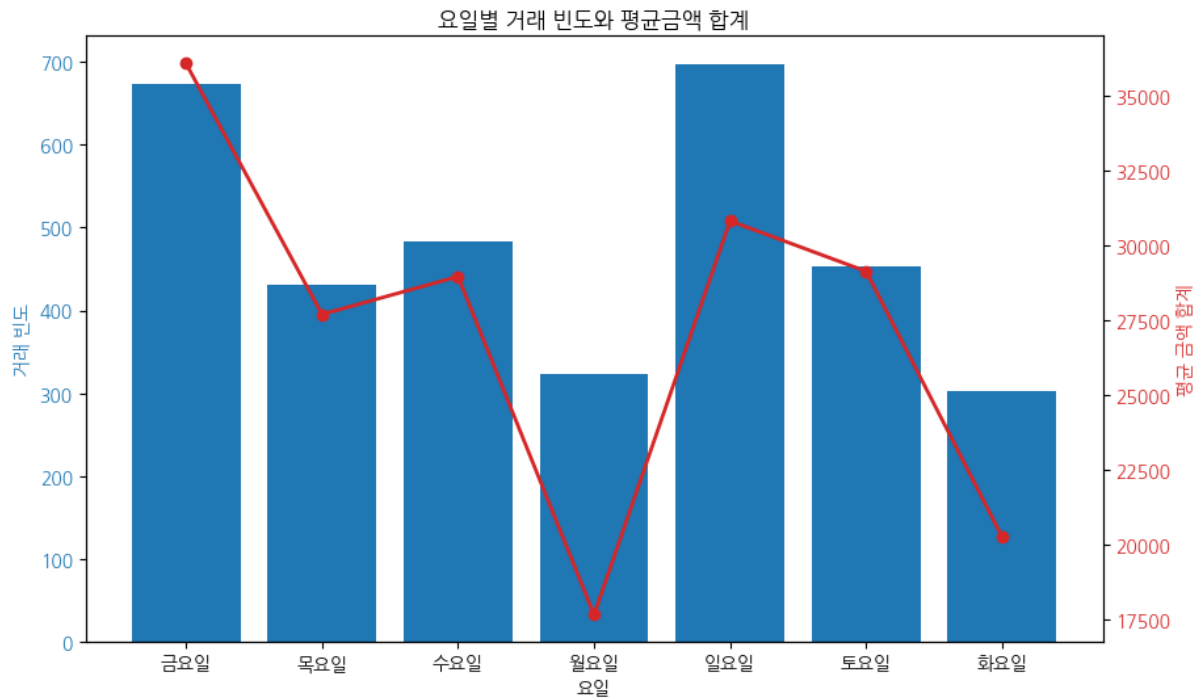
	제품카테고리	월	판매량
0	Accessories	Dec	13
1	Apparel	Dec	278
2	Bags	Sep	33
3	Bottles	Sep	5
4	Drinkware	Sep	63
5	Gift Cards	Oct	8
6	Headgear	Sep	12
7	Housewares	Aug	2
8	Lifestyle	Sep	72
9	Nest	Dec	119
10	Nest-Canada	Dec	6
11	Nest-USA	Dec	280
12	Notebooks & Journals	Aug	16
13	Office	Dec	96
14	Waze	Dec	16

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



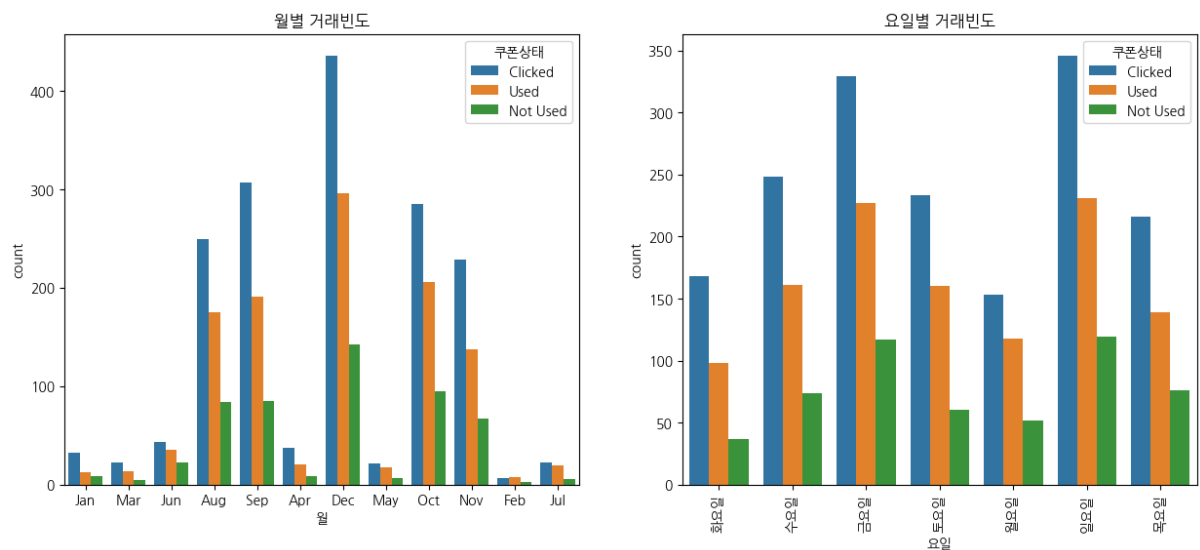
8~12월 거래 빈도가 높다. 12월과 11월의 경우 크리스마스와 블랙프라이데이가 있어 높은 것으로 파악된다. 8~12월의 경우 거래빈도가 높으므로 해당 달의 경우 평균 금액을 높이기 위해 고가의 제품을 많이 전시하면 좋을 것 같다.

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



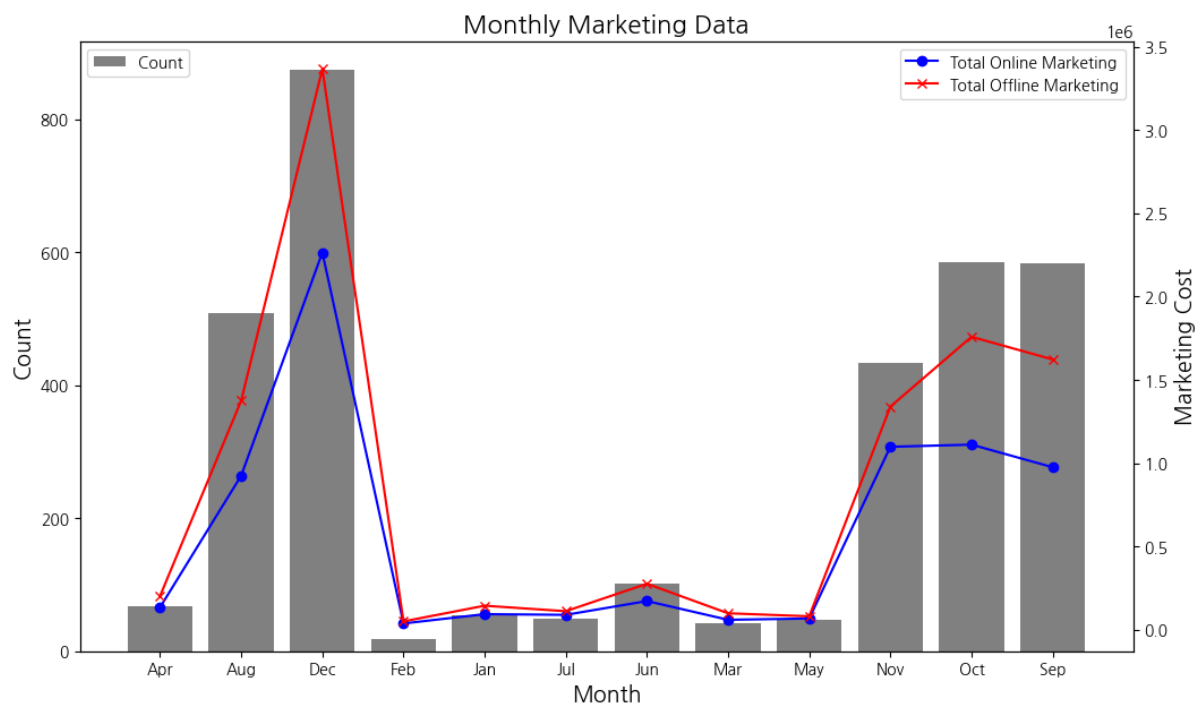
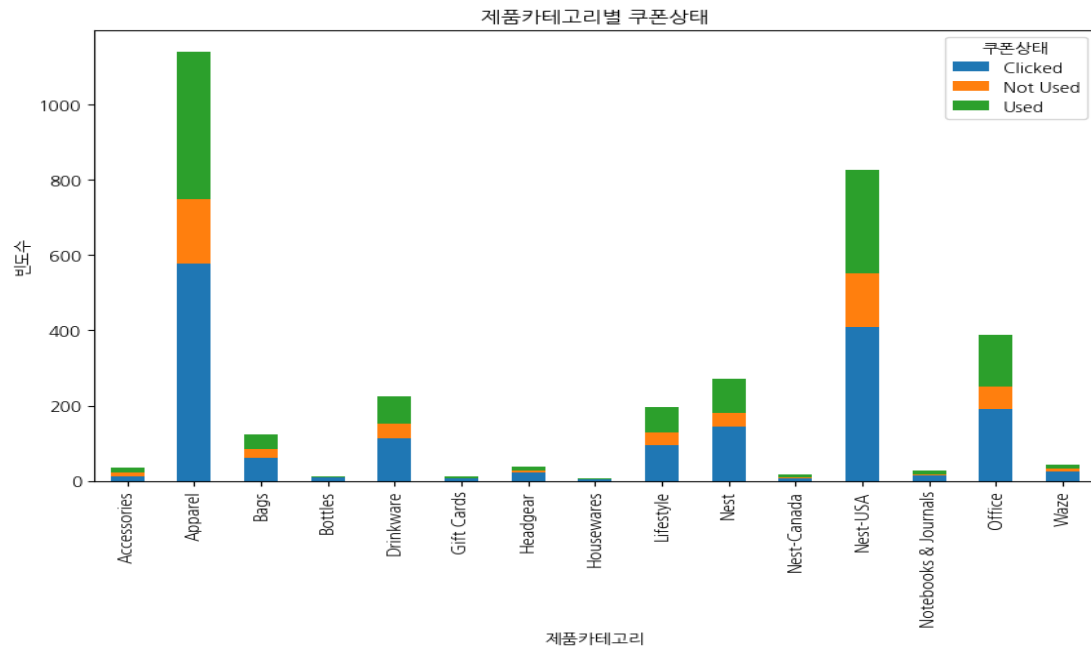
금요일과 일요일 거래량이 많은 것을 확인할 수 있다. 금요일과 일요일 고객이 선호하는 Apparel, Nest 카테고리의 제품을 더욱더 활발하게 홍보할 수 있도록 한다. 특히 금요일은 평균 금액 합계가 거래빈도수보다 높은 것은 것을 확인할 수 있다. 거래빈도수와 금액 사이의 관계가 비례해 보인다. 이때는 고가의 제품도 많이 노출 시키면 좋을 것 같다.

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.



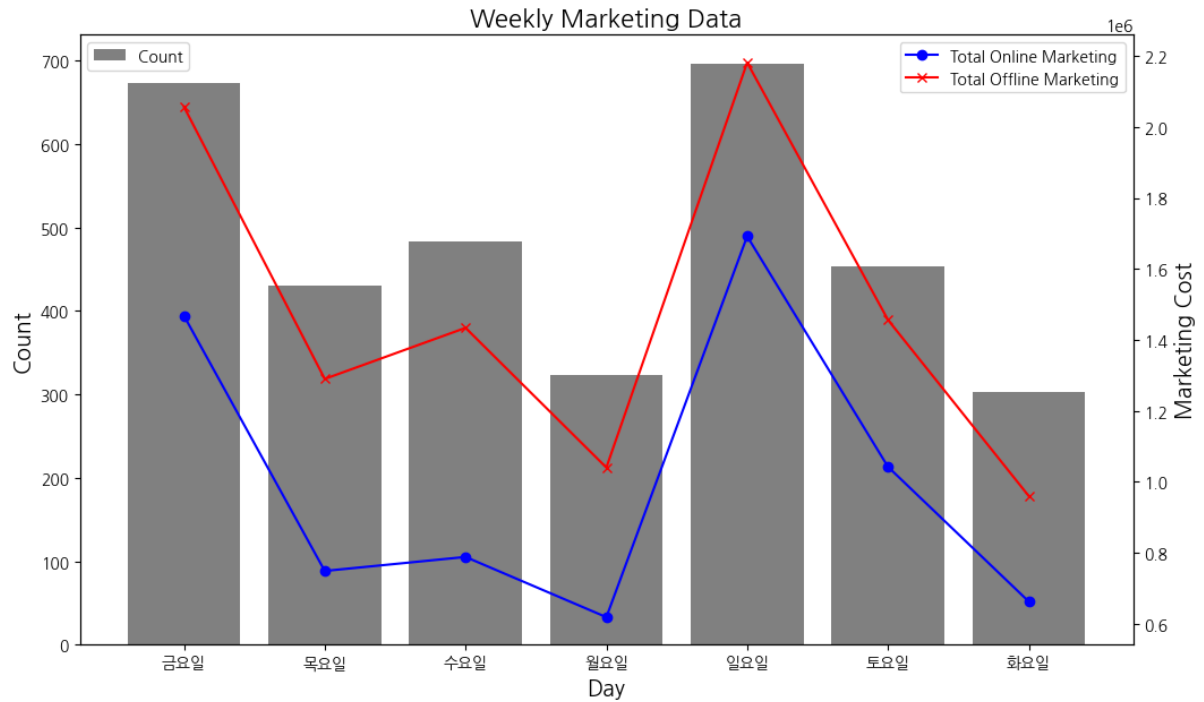
대부분 쿠폰 사용량이 적다. 거래가 많은 8,9,10,11,12 월은 쿠폰 사용률을 높여 더 많은 제품을 구매할 수 있도록 유도하며, 위의 달은 특히 금요일과 일요일날 행사를 많이 하면 좋을 것으로 보인다. 또한 11 월과 12 월은 달의 특성상 고가의 제품에 할인 쿠폰을 많이 제공하면 좋을 것 같다.

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

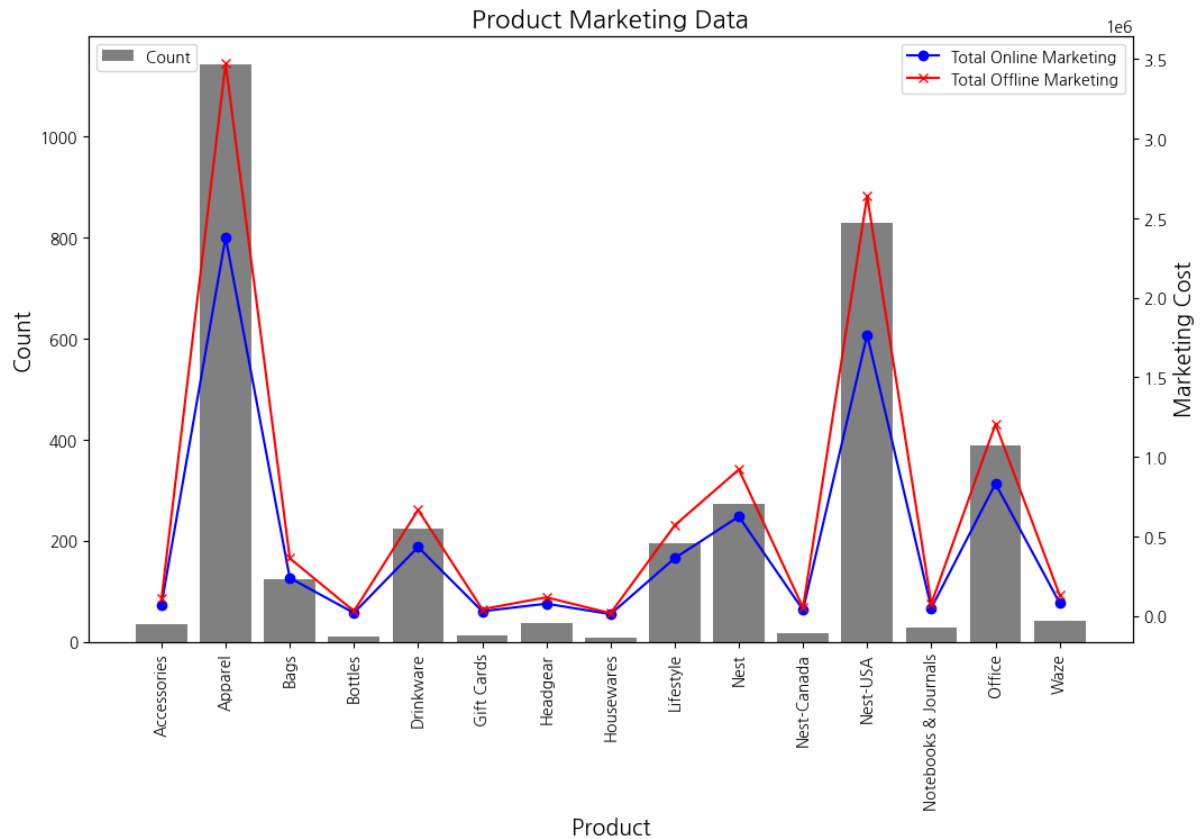


앞서 봤듯이 8~12 월 거래 빈도 수, 매출 합계 모두 높았다. 뿐만 아니라 해당 달은 마케팅 비용 둘다 높은 것을 확인할 수 있다.

11~12 월의 경우 마케팅 비용을 줄이고 쿠폰 사용량을 늘릴 수 있도록 하고 11,12 월에 사용하는 마케팅 비용을 다른 달에 투자하는 것이 좋아 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.

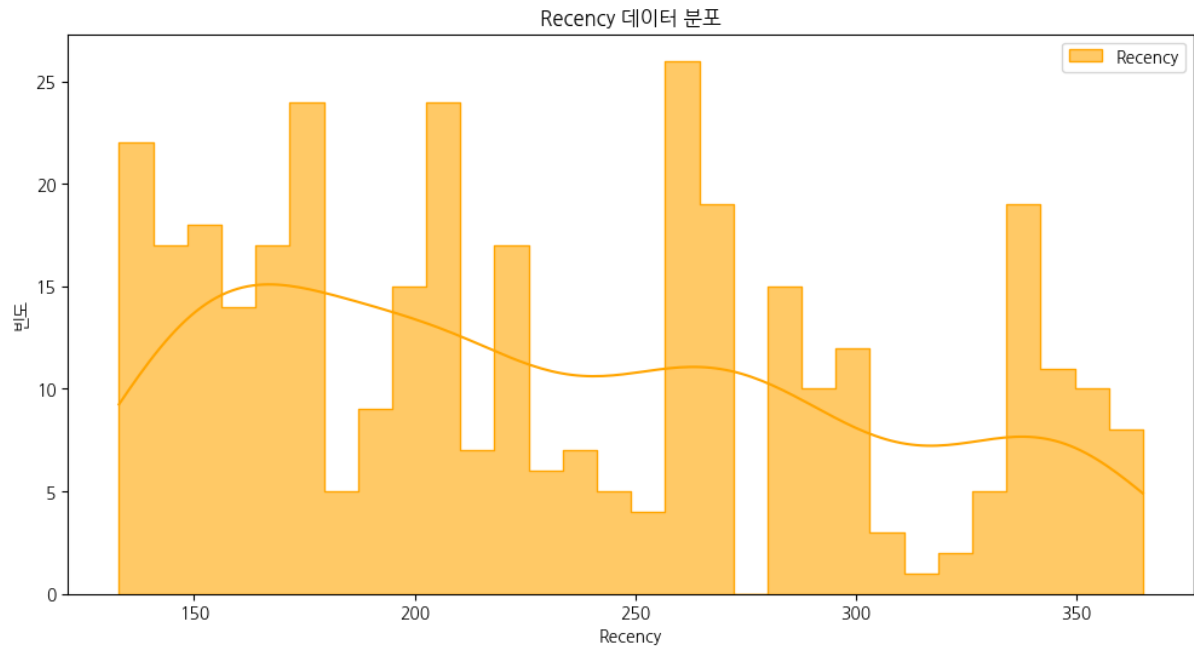
또한 금요일과 일요일, 8~12월에 거래량이 뿐만 아니라 마케팅 비용 또한 높다. 11월, 12월의 경우 블랙프라이데이, 크리스마스 등 특별한 날이 있기 때문에 고객들이 제품을 많이 구매한다. 이 경우 많은 마케팅 비용을 투자하지 않아도 거래량이 계속 많을 것으로 생각한다. 따라서 이 달은 마케팅 비용을 줄이고 다른 달에 투자하는 것이 효율적일 것 같다. 날이 좋은 달은 오프라인 마케팅을 통해 고객이 단순히 시각적으로 즐기는 것 뿐만 아니라 체험을 통해 즐길 수 있도록 하여 거래 빈도를 높일 수 있도록 한다. 따라서 계절별 오프라인,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

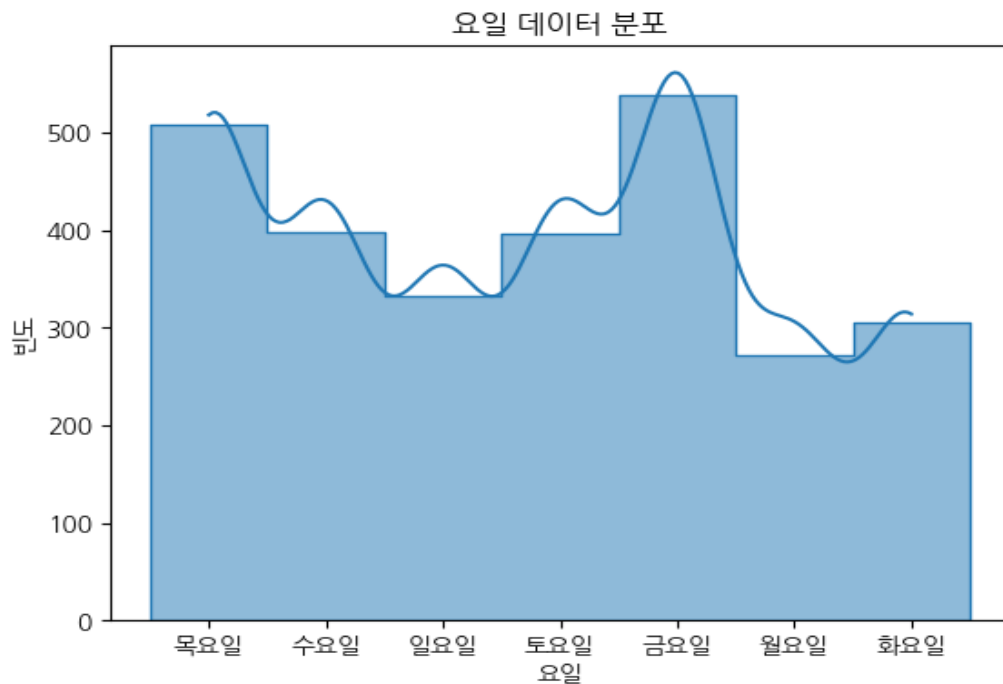
고가의 쿠폰 사용량은 적은 편이다. 이는 고객이 현명한 구매를 하지 못하는 것을 나타낸다. 뿐만 아니라 배송료 또한 제품 구매 금액에 비해 높은 경우가 많다. 이러한 경우 고객이 배송료와 제품 둘 다 비싸게 느껴 구매를 하지 않는 경우가 생길 수도 있다. 고객이 쿠폰을 사용할 수 있도록 유도해야 하며, 배송료의 경우 특정 금액 이상 절감 또는 무료 행사를 진행해 구매할 수 있도록 유도해야 한다. 또한 11,12월의 경우 거래 빈도수가 많으므로 이 달은 배송료 무료 혜택 등을 통해 고객이 더욱더 많은 제품을, 자주 살 수 있도록 한다.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

[이탈위험 고객]

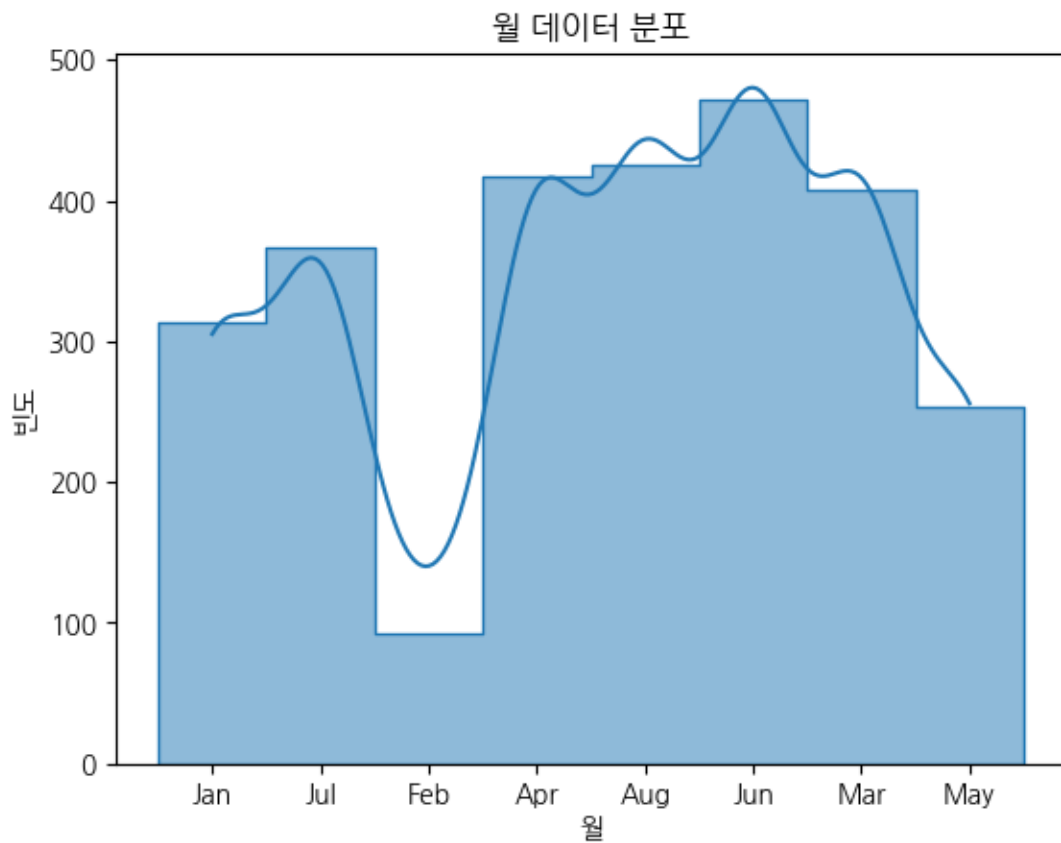
이탈위험의 고객은 이커머스 거래를 장기간 하지 않는 경우로, 이들의 최근 거래 기간과 이들이 거래한 제품에 대해 분석해야 한다.



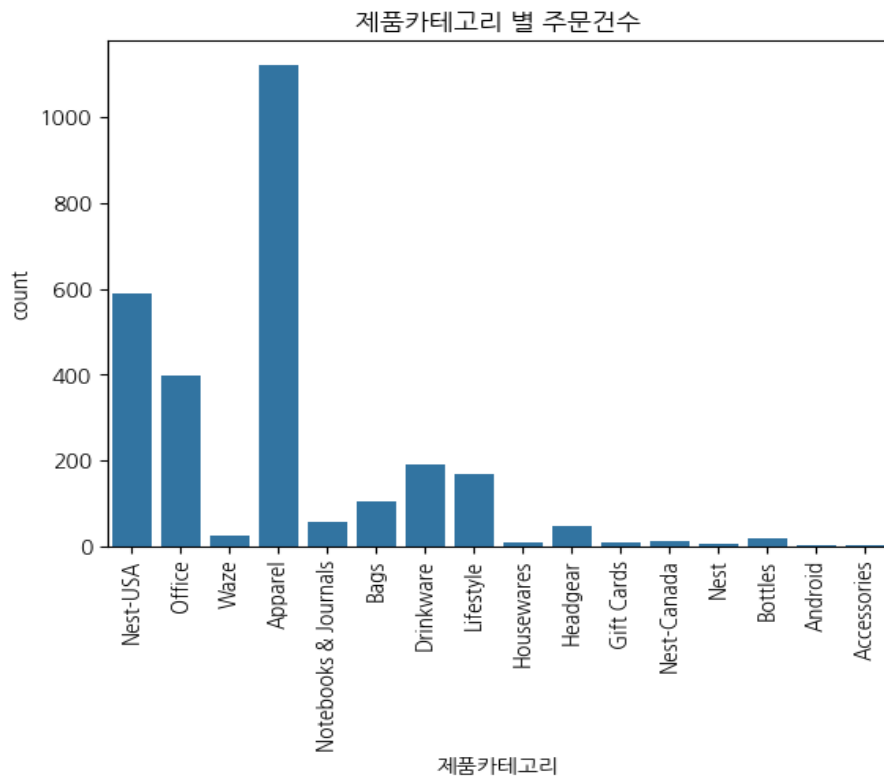
최근 거래 일이 일년 정도 되가는 고객이 상대적으로 많은 편이며, 제일 작은 값이 133 일이다. 4개월동안 이 고객들은 거래를 하지 않은 것으로 확인된다.



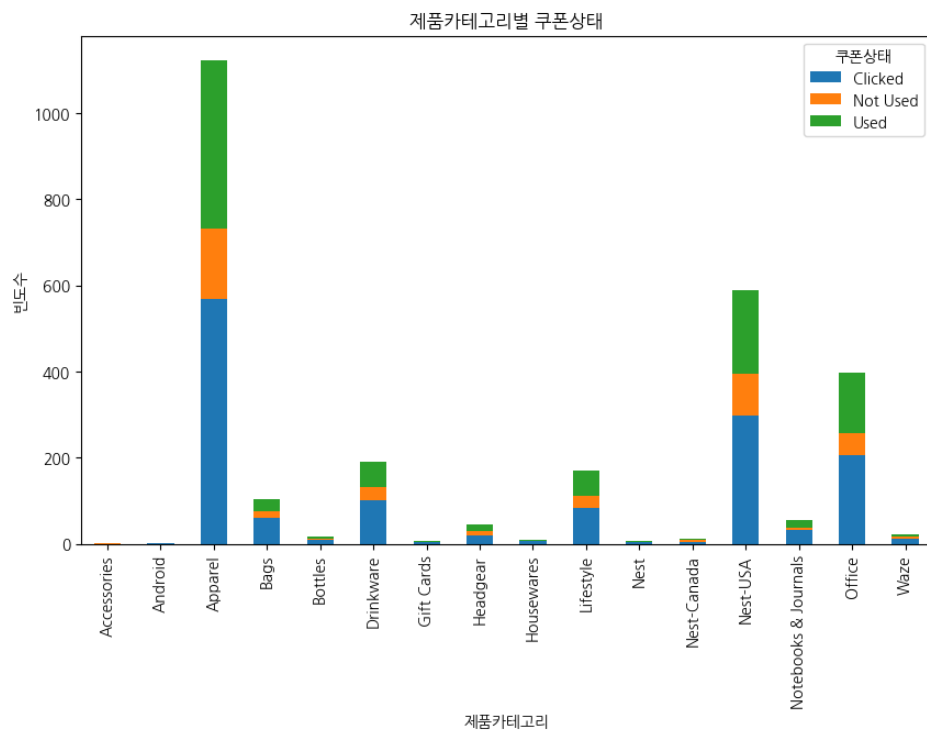
일반고객이 일요일, 월요일에 구매를 많이 하는 반면 이탈위험 고객의 경우 구매량이 적어지는 경향을 보인다.



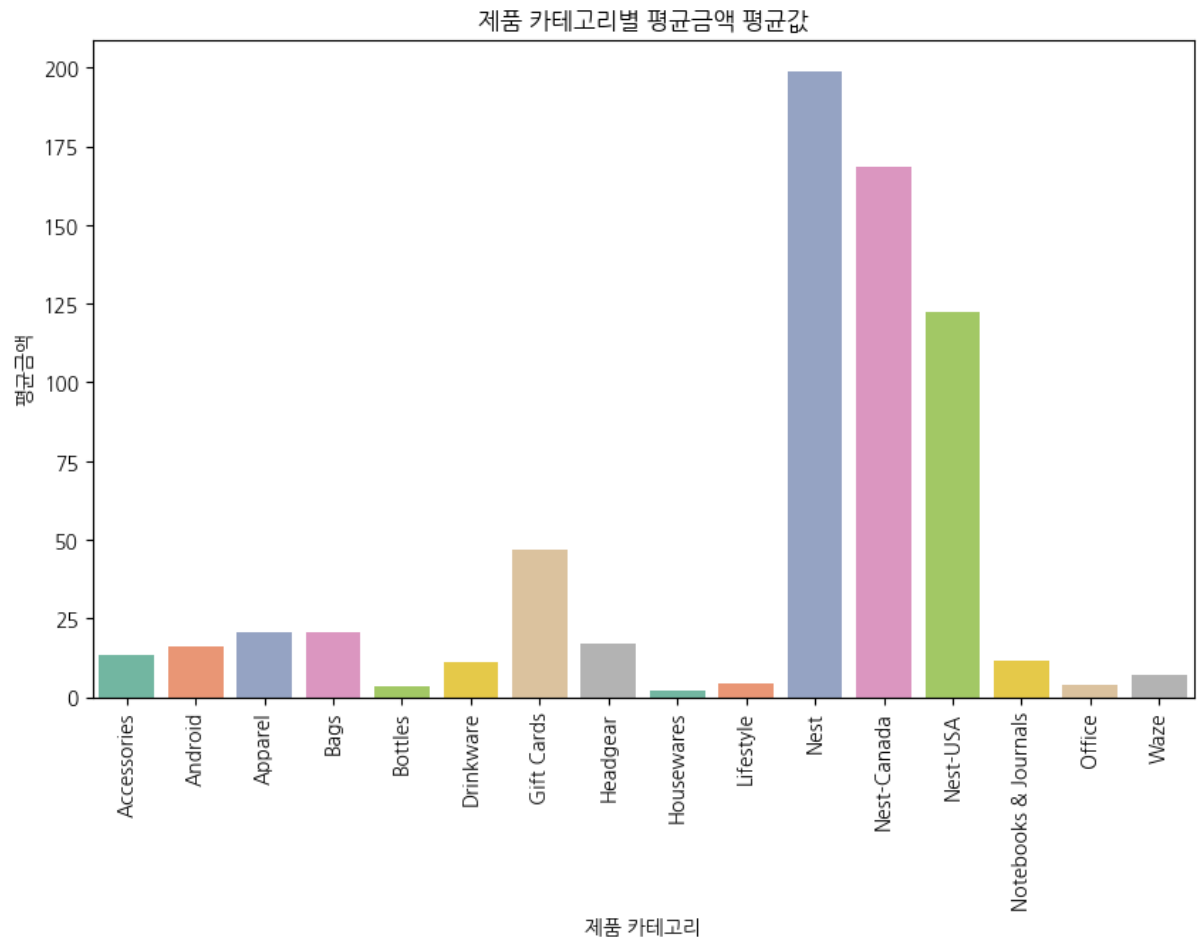
이탈위험 고객의 경우 3,4,5,6,8월에 구매를 많이 한 것으로 보인다. 특이점은 11월, 12월에 구매를 하지 않았다. 이들에게 11월, 12월은 블랙프라이데이, 크리스마스 등 특별한 날이 포함된 달이므로 고객의 구매량이 증가하는 달이다. 고객의 구매량이 증가하는 만큼 행사 상품도 많고, 할인도 많이 들어가는 경우, 마케팅을 많이 하기 때문에 이탈위험 고객의 마음을 사로잡을 마케팅 전략이 필요해 보인다.



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 타깃으로 마케팅을 세워 이탈위험 고객을 붙잡는 방안을 찾을 필요가 있다.

[결론]

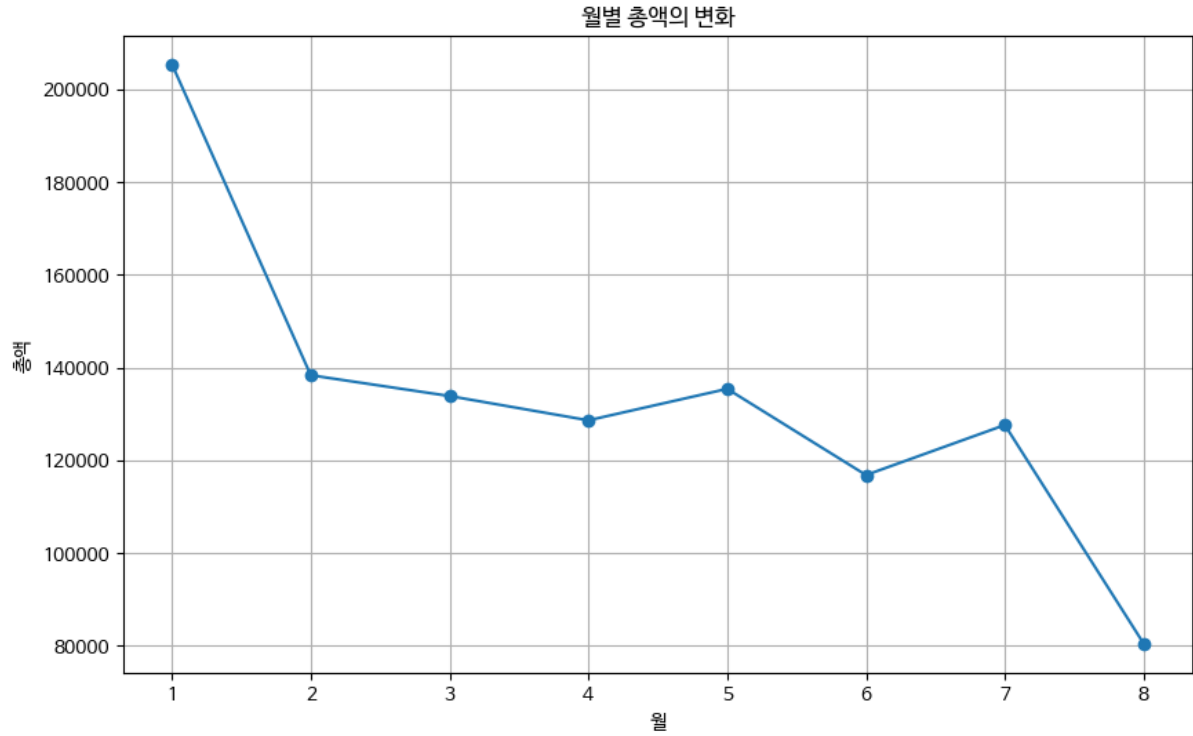
이탈위험 고객의 경우 최대 1년, 최소 4개월 가량 거래량이 없는 것으로 확인된다. 이들이 주된 구매 제품은 Apparel, Nest인 것을 확인했다. 특히 Nest의 경우 제품의 가격이 고가에 해당되기에 이를 많이 구매하는 이탈위험 고객의 구매를 유도할 필요가 있다.

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.

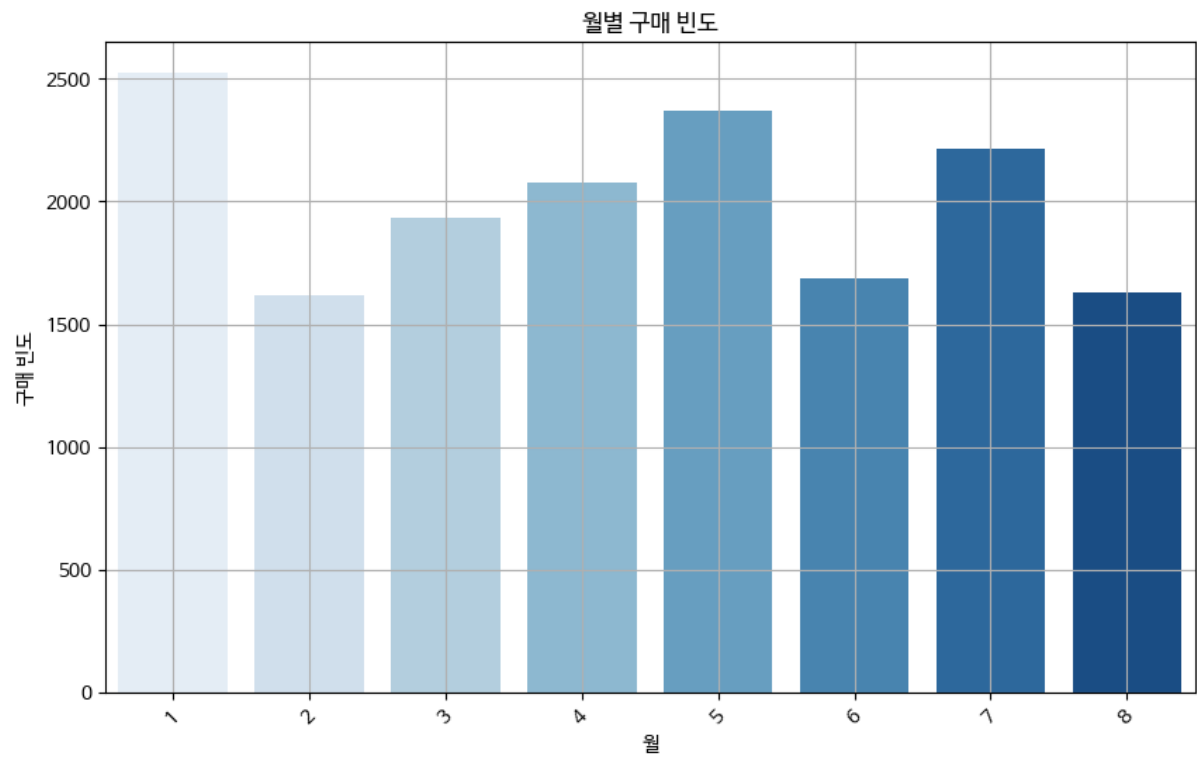
이 고객의 경우 봄-여름에 구매를 많이 했으므로 마케팅을 할 때 이점을 참고하면 좋을 거 같다. 특히 오프라인 마케팅의 경우 날씨와 계절감을 이용하여 팝업스토어 마케팅을 진행하면 좋을 것으로 보인다. 이 마케팅은 고객이 체험 뿐만 아니라 좋은 추억이 될 수 있도록 하여 장기간 이 제품을 생각하고 구매할 수 있도록 유도할 수 있을 것 같다.

재구매 유도 고객: 높은 방문율과 구매율을 가졌지만 방문이 뜸해진 고객

1. 월별 구매액 및 구매 빈도 분석

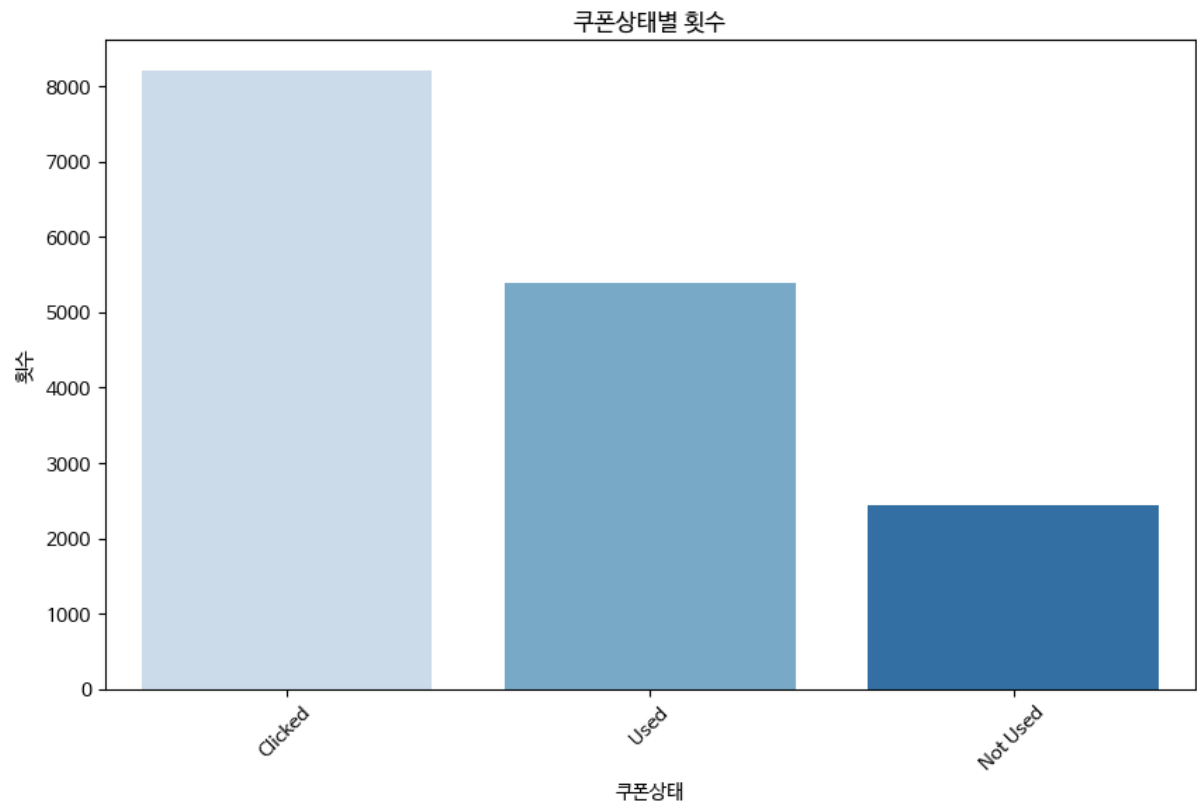


2월, 3월, 4월, 6월, 8월에 사용자의 구매 금액이 감소하는 것을 알 수 있다. 이중 2월 6월 8월인 경우 사용자들이 구매한 금액이 크게 감소한 것을 알 수 있다.

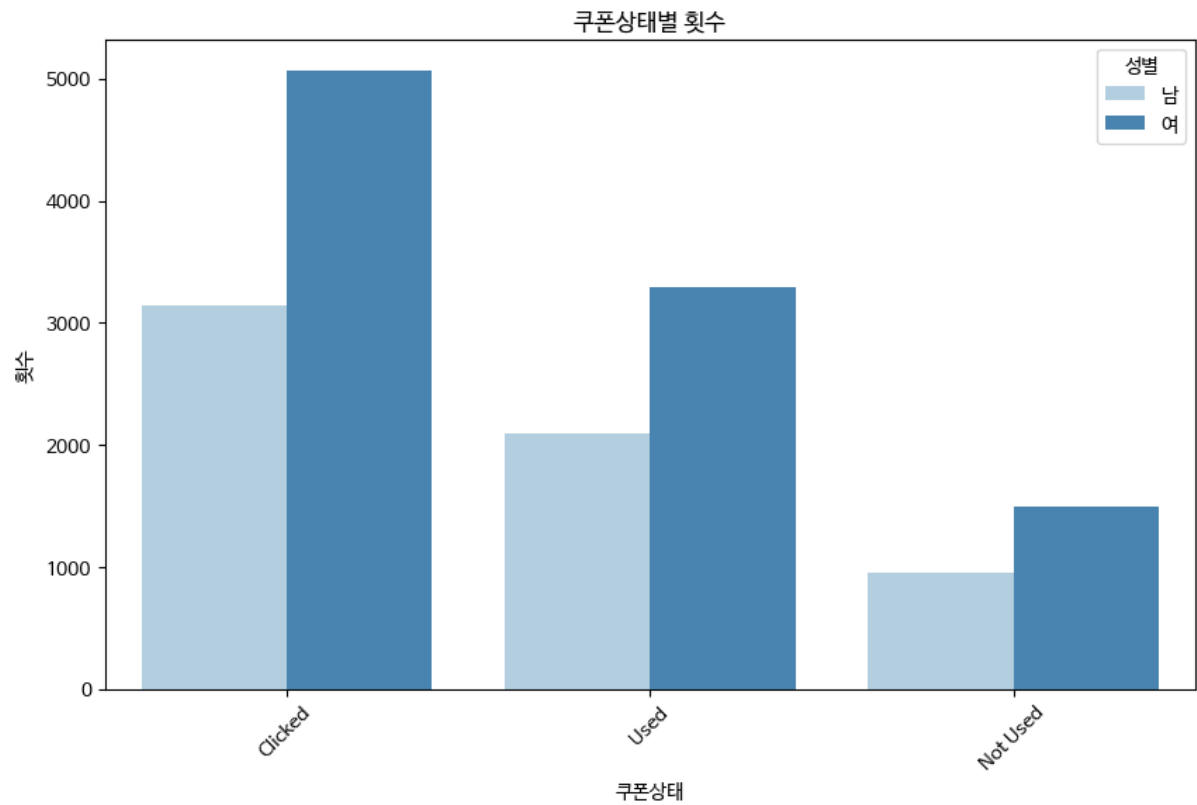


구매 빈도를 보았을 때 2월, 6월, 8월의 구매 빈도가 다른 월에 비해 낮은 것으로 보아 판매하는 제품의 종류가 크게 다르지 않을 것으로 판단된다. (만약 금액에서 차이가 나지만 구매 빈도에서 차이가 나지 않는다면 제품의 종류가 달라졌을 것이다.)

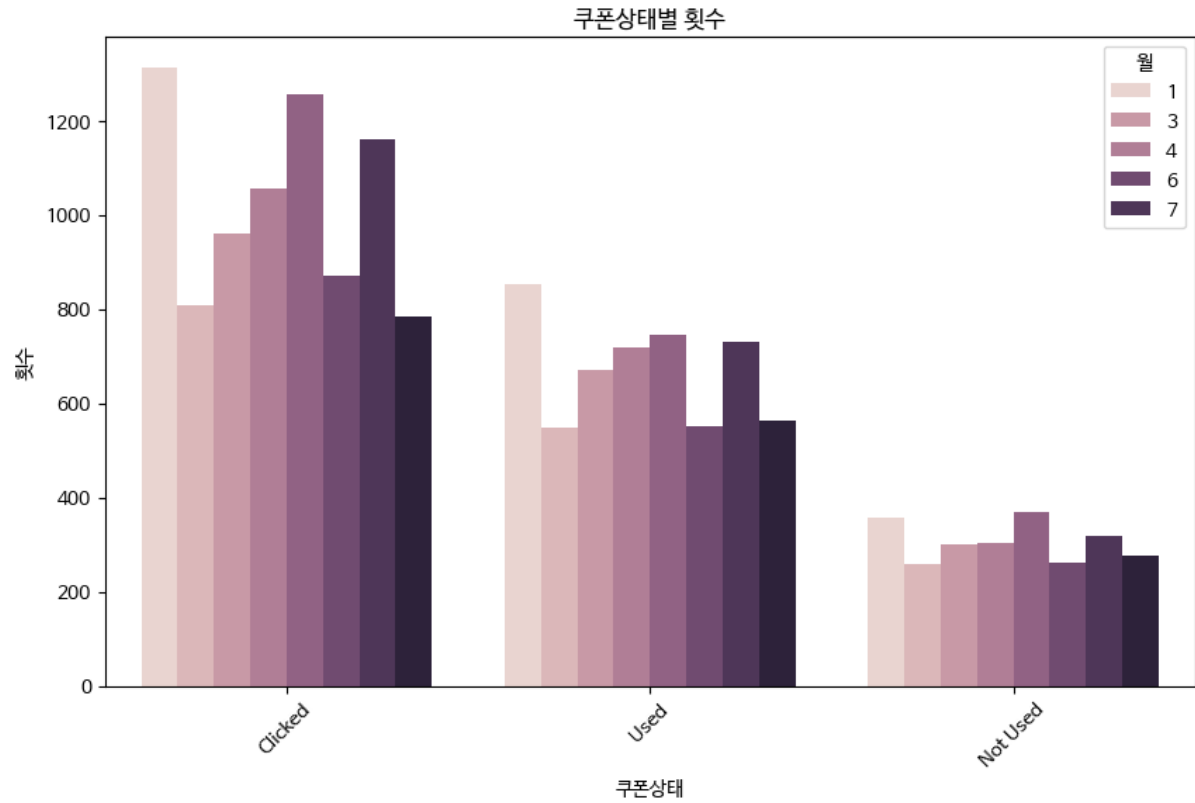
2. 쿠폰 사용 여부 분석



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.

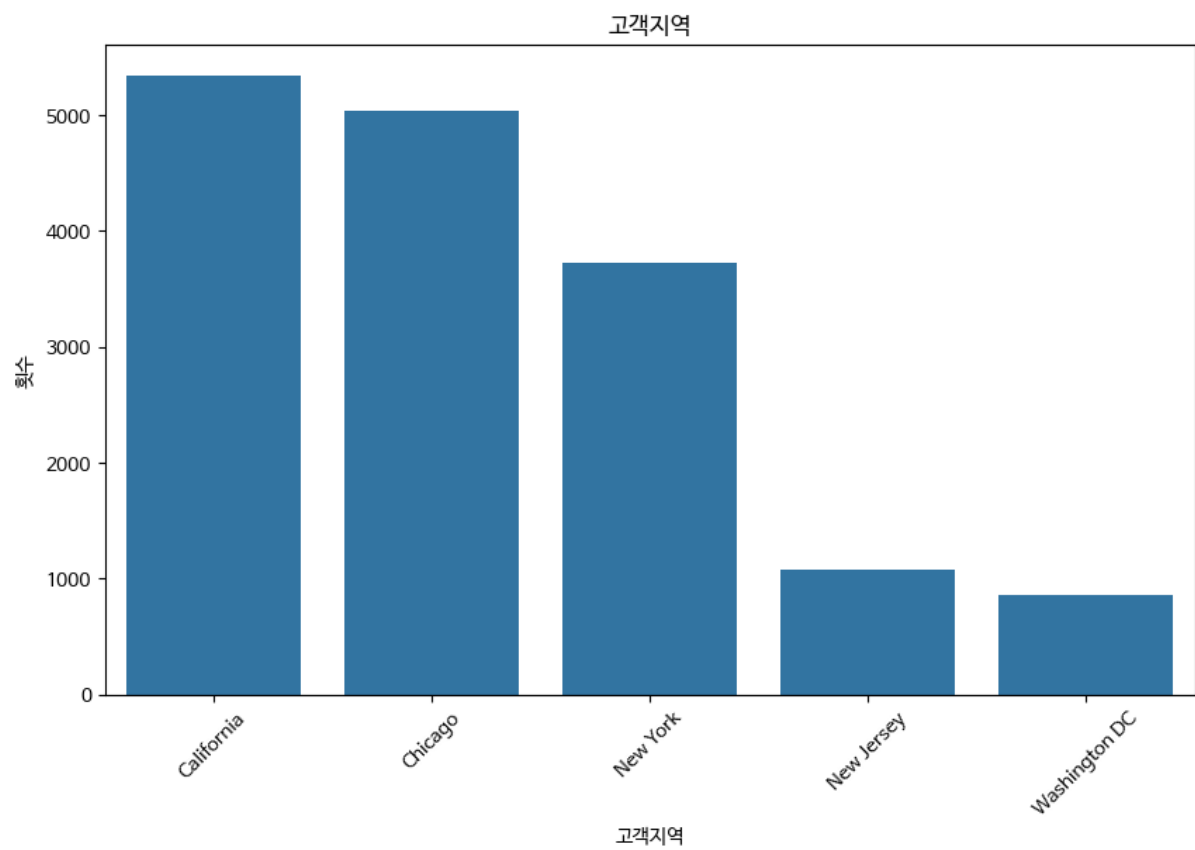


월별로 시각화했을 때 구매 빈도가 낮았던 2월, 6월, 8월에 다른 월에 비해 전체적으로 낮다는 것을 알 수 있다.

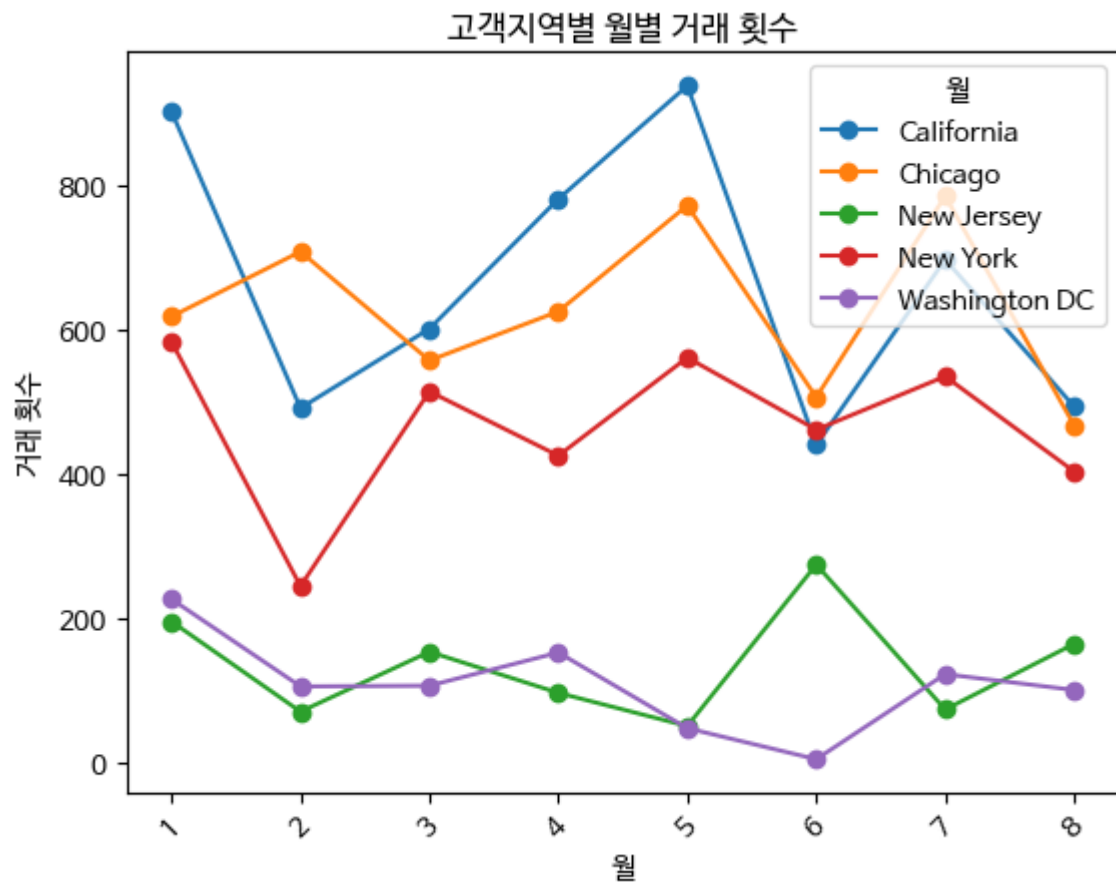
월	
1	10.0
2	20.0
3	30.0
4	10.0
5	20.0
6	30.0
7	10.0
8	20.0

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

3. 지역별 분석

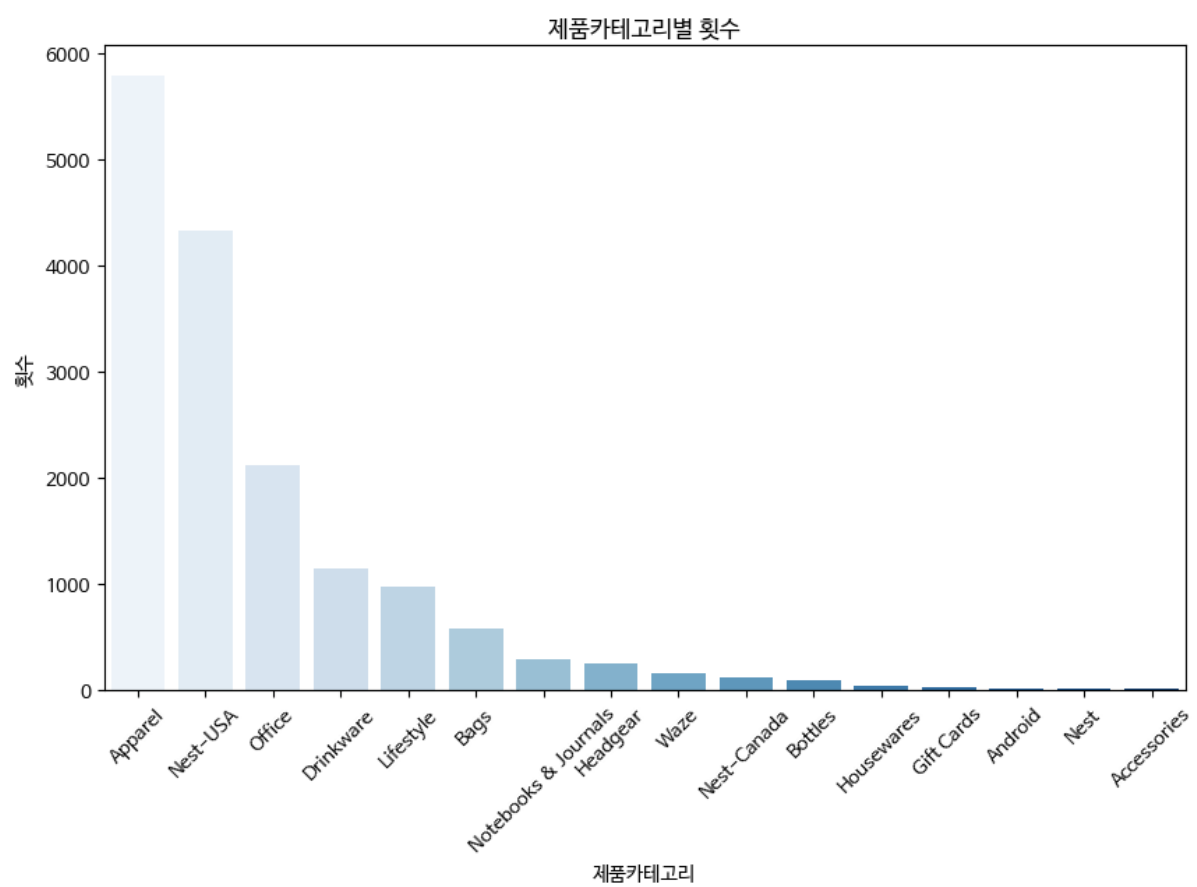


5개의 지역이 있고 그 중 **California, Chicago, New York**에서 고객 사용 빈도가 다른 두 지역보다 많다는 것을 알 수 있다.



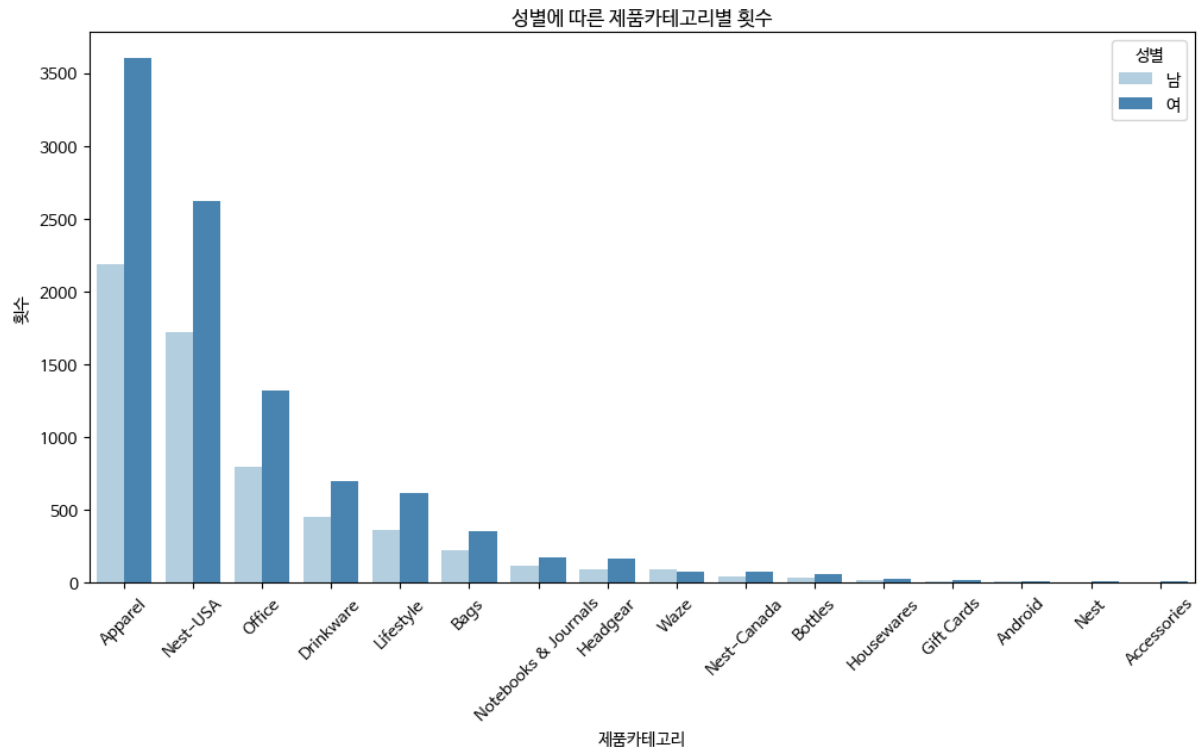
월별 지역 거래 빈도를 보았을 때 거래량이 많은 3개의 지역의 거래횟수가 감소할 때 거래액이 감소했다는 것을 알 수 있다. 예외적으로 2월의 **Chicago**는 거래횟수가 증가했지만 나머지 두 지역의 거래량이 크게 감소했기 때문에 2월 전체의 구매빈도 및 거래액이 감소했다는 것을 알 수 있다.

4. 제품카테고리 분석

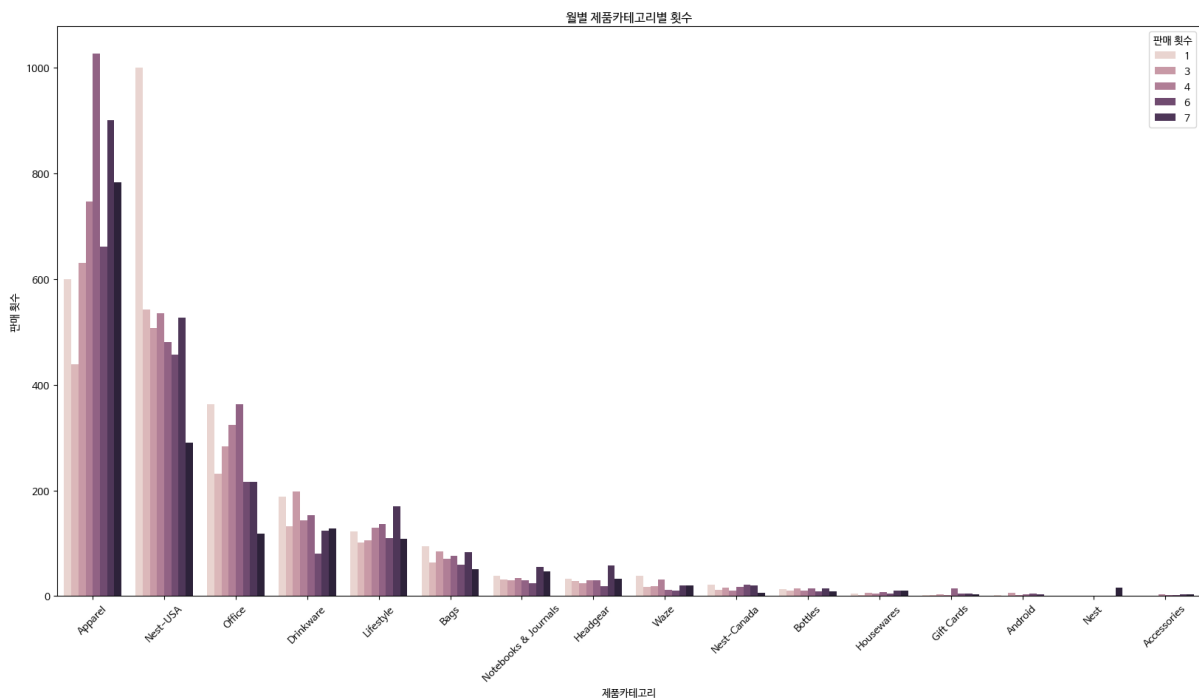


거래빈도가 가장 많은 카테고리 3개는 Apparel, Next-USA, Office이다. 이 3개의 카테고리가 전체

거래량의 76.3%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 거래량이 많은 카테고리들에서 2월, 6월, 8월의 거래량이 낮다는 것을 알 수 있다.