

『이커머스 고객 세분화 분석 아이디어 경진대회』 제안서

① 팀 명

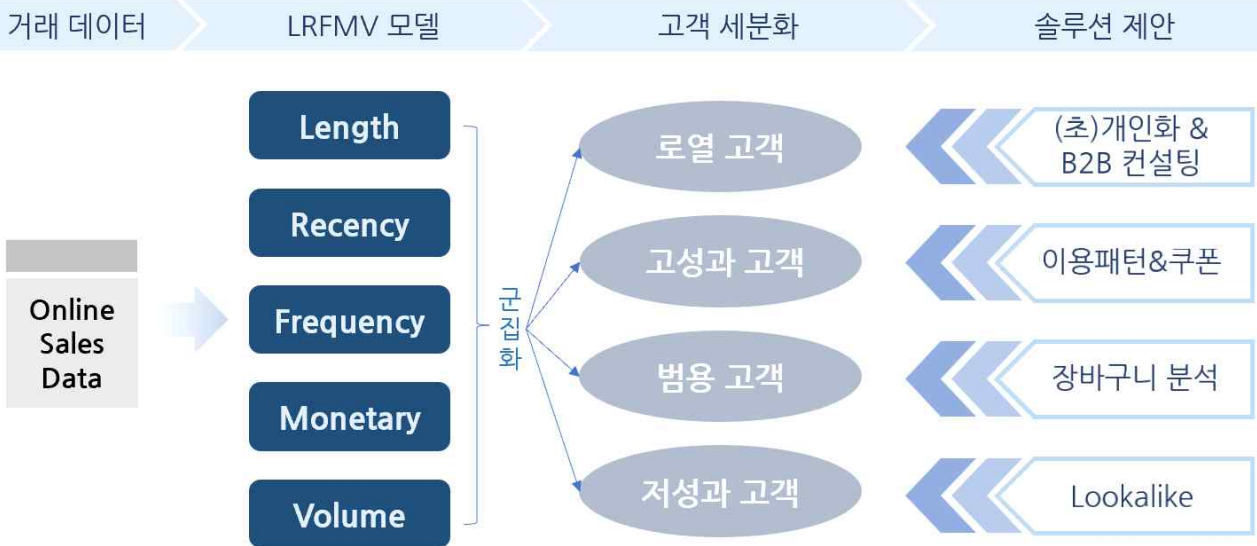
트리플송

② 제 목

LRFMV 모델을 이용한 고객 세분화
및 연관 분석 & 솔루션 제안

고객 세분화 분석 결과

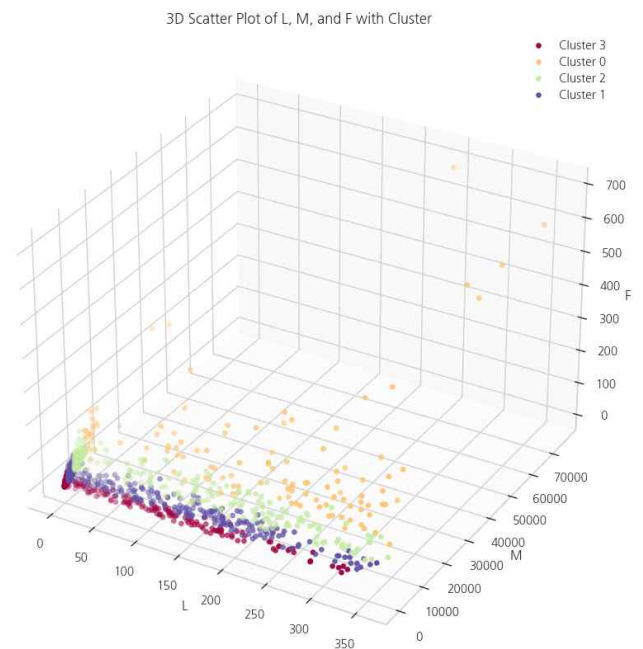
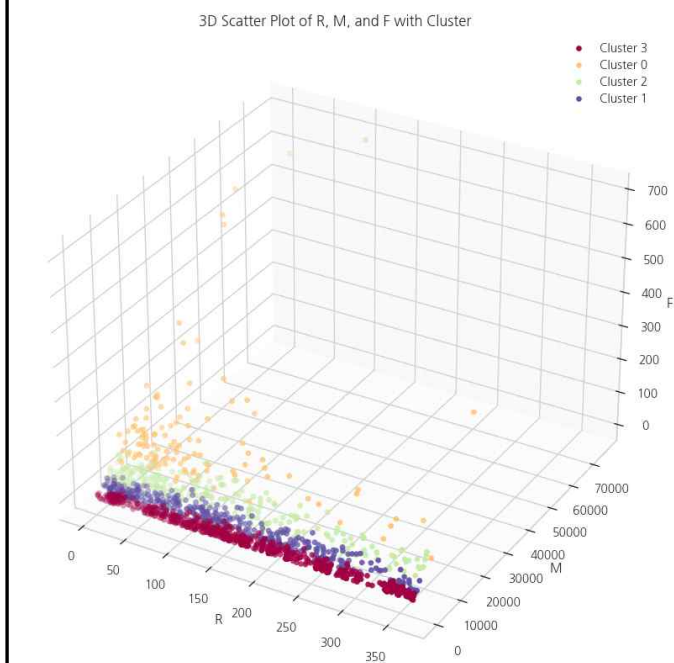
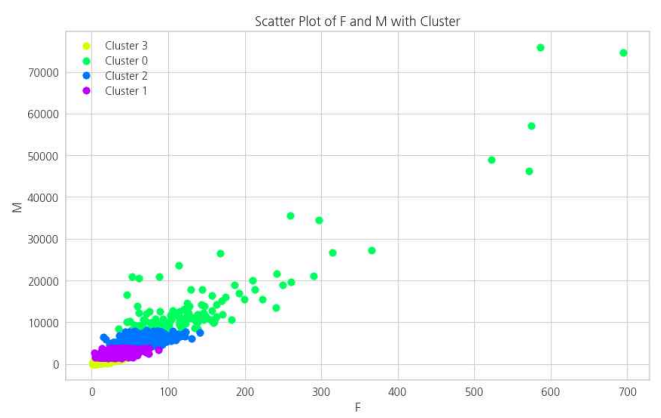
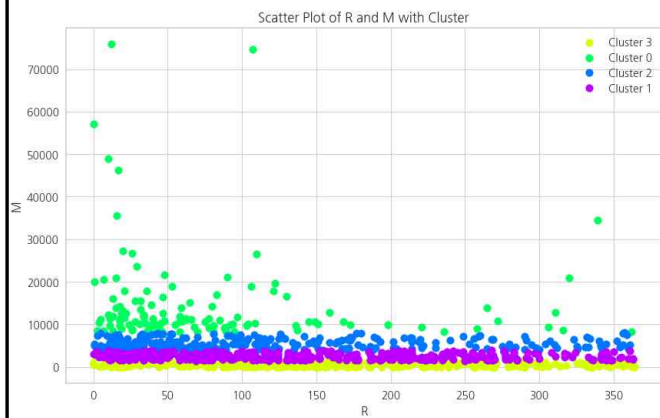
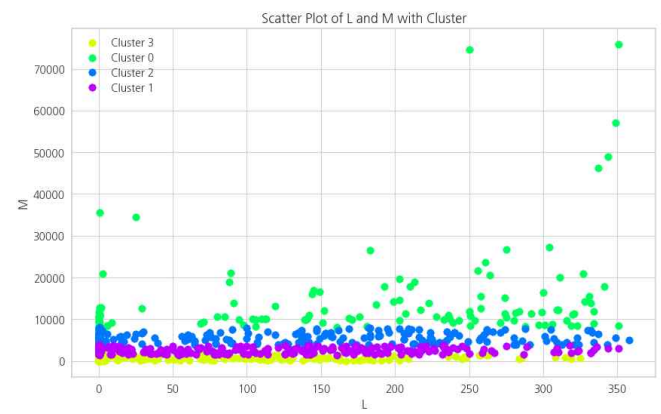
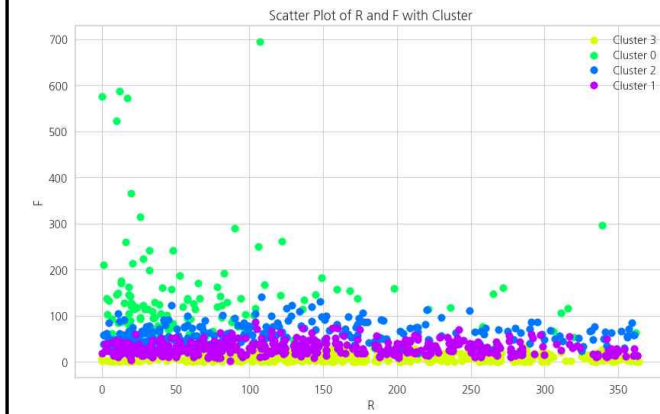
[분석 Flow Chart]



고객세분화를 위해 RFM보다 개선된 버전의 LRFM에서 이익과 수량 간의 관계까지 반영할 수 있는 LRFMV model을 사용하였다. LRFMV는 각각 Length(길이 : 고객의 첫 번째 방문과 마지막 방문 사이의 일수), Recency(최근성 : 고객이 마지막으로 방문한 이후의 날짜 구간), Frequency(빈도 : 고객이 구매한 횟수), Monetary(금액 : 고객이 구매한 금액), Volume(거래량 : 고객이 하루 동안 구매한 평균 상품 수)을 의미한다. 군집분석을 수행할 때 LRFMV를 이용하여 이익극대화 측면의 고객 가치를 분석할 수 있다. 각 LRFMV 항목들의 높음/낮음에 따라 K-medoids 방법을 통해 이상치의 영향을 덜받는 군집 4개를 도출할 수 있었다. 각 군집들의 특성과 분류 명명은 다음과 같다.

Cluster	L	R	F	M	V	고객 충성도	유지 비용	고객 분류
0	↑	↓	↑	↑	↑	높음	낮음	로열 고객
1	↓	↑	↓	↓	↑	높음	높음	고성과 고객
2	↑	↓	↑	↑	↓	낮음	낮음	범용 고객
3	↓	↑	↓	↓	↓	낮음	높음	저성과 고객

● 군집들이 분류된 분포를 시각화



군집 0) 로열 고객

- 고객 유지 기간이 길며 최근에 구매하였고, 자주 많이 소비하면서 이익이 큰 고객
- 기업 입장에서 가장 이상적인 고객
- 해당 고객 군집에 대한 데이터가 많이 쌓여있으며 최소한 현상 유지의 필요성이 있는 군집

군집 1) 고성과 고객

- 고객 유지기간이 짧고 구매가 이루어진지 시간이 경과했기 때문에 자주 많이 소비하지는 않은 고객이지만 이익이 큰 고객
- 이익이 많이 발생하는 고객이기 때문에 다시 고객이 유입될 수 있도록 해야 함
- 고객이 자발적으로 서비스를 찾게 되도록 유인책이 필요

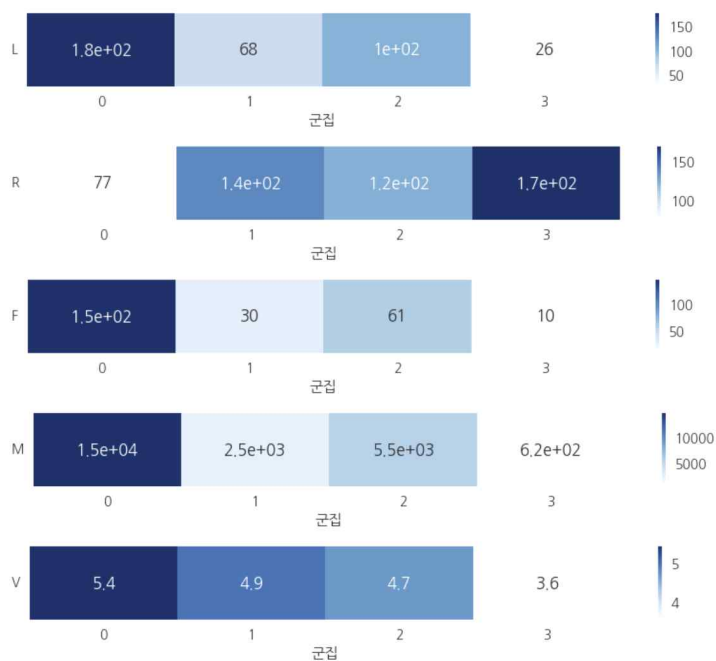
군집 2) 범용 고객

- 고객 유지 기간이 길며 최근에 구매하였고, 자주 많이 소비하면서 이익이 작은 고객
- 큰 이익이 되지는 않지만 기업 수익의 기반을 이루고 있는 고객
- 고객이 자주 찾아와주기 때문에 더 많은 소비가 이루어질 수 있도록 유도할 필요가 있음

군집 3) 저성과 고객

- 고객 유지기간이 짧고 구매가 이루어진지 시간이 경과했기 때문에 자주 많이 소비하지는 않은 고객이면서 이익이 작은 고객
- 짧게 서비스를 경험했다 이탈한 고객
- 관련된 데이터가 매우 적고 신규 고객과도 같은 군집

군집별 LRFMV의 평균 수치는 아래와 같다.



군집 0) 로열 고객

- 고객 유지 기간이 길며 최근에 구매하였고, 자주 많이 소비하면서 이익이 큰 고객
- 기업 입장에서 가장 이상적인 고객
- 해당 고객 군집에 대한 데이터가 많이 쌓여있으며 최소한 현상 유지의 필요성이 있는 군집

군집 1) 고성과 고객

- 고객 유지기간이 짧고 구매가 이루어진지 시간이 경과했기 때문에 자주 많이 소비하지는 않은 고객이지만 이익이 큰 고객
- 이익이 많이 발생하는 고객이기 때문에 다시 고객이 유입될 수 있도록 해야 함
- 고객이 자발적으로 서비스를 찾게 되도록 유인책이 필요

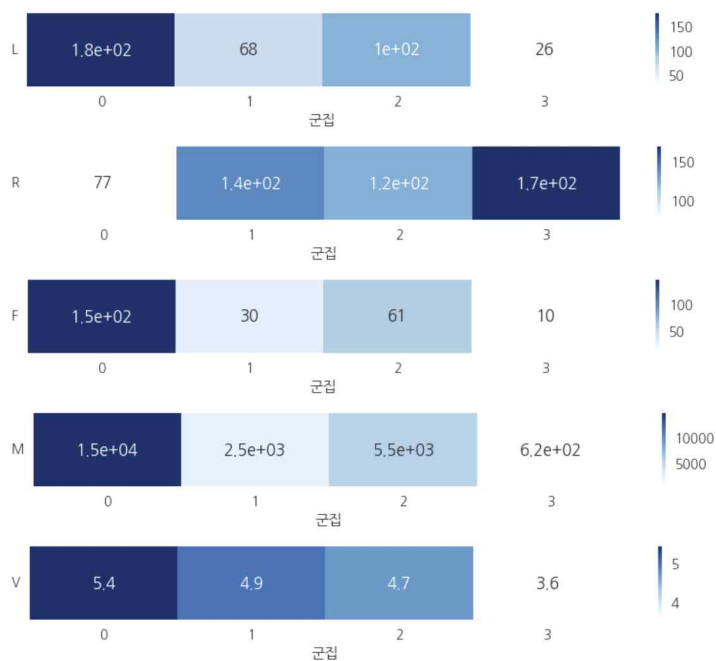
군집 2) 범용 고객

- 고객 유지 기간이 길며 최근에 구매하였고, 자주 많이 소비하면서 이익이 작은 고객
- 큰 이익이 되지는 않지만 기업 수익의 기반을 이루고 있는 고객
- 고객이 자주 찾아와주기 때문에 더 많은 소비가 이루어질 수 있도록 유도할 필요가 있음

군집 3) 저성과 고객

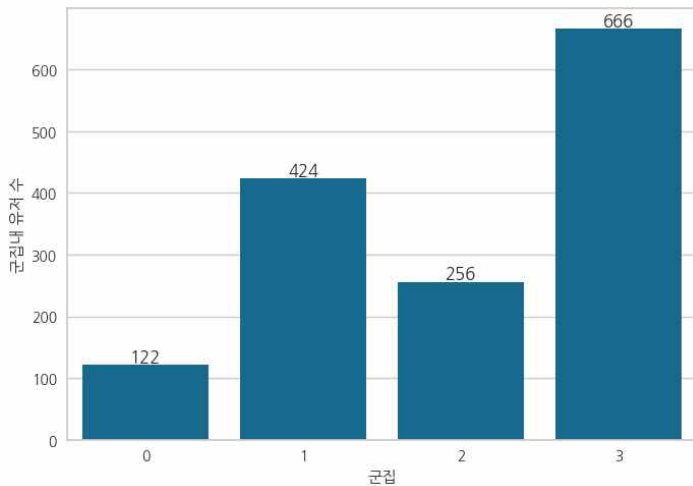
- 고객 유지기간이 짧고 구매가 이루어진지 시간이 경과했기 때문에 자주 많이 소비하지는 않은 고객이면서 이익이 작은 고객
- 짧게 서비스를 경험했다 이탈한 고객
- 관련된 데이터가 매우 적고 신규 고객과도 같은 군집

군집별 LRFMV의 평균 수치는 아래와 같다.



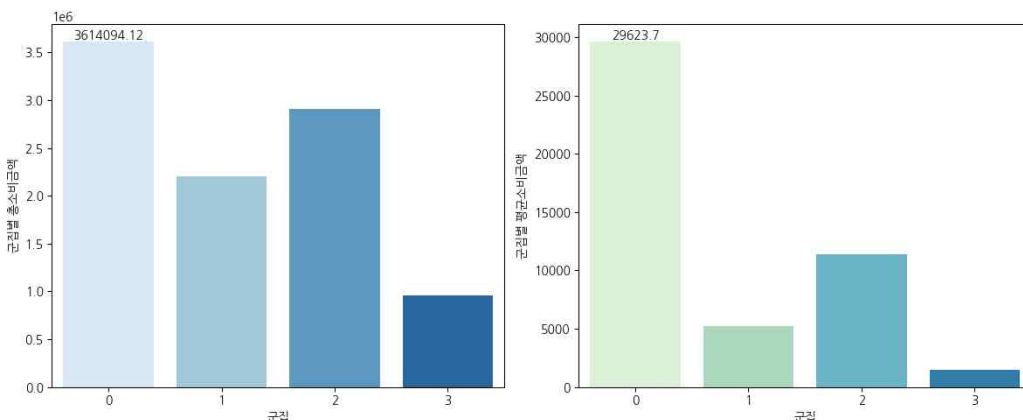
이 밖에 군집별 특징은 다음과 같다.

1. 군집 내 유저(고객) 수



군집 3 (저성과 고객) > 군집 1 (고성과 고객) > 군집 2 (범용 고객) > 군집 0 (로열 고객) 순으로 고객 수가 많았다. 군집 3의 고객이 가장 많은 부분을 차지한 것은, 2019년 1년 간의 데이터만 사용했기 때문으로 판단되며, 향후 더 많은 기간의 데이터로 세분화한다면 서서히 낮아질 것으로 예상된다.

2. 군집별 총 소비금액 & 평균 소비금액



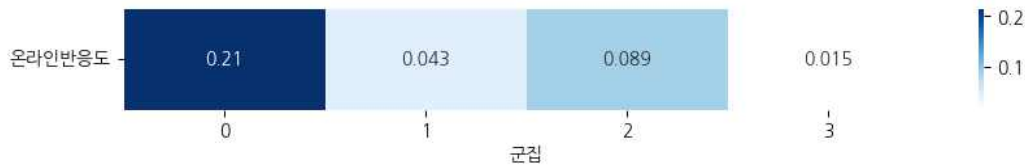
군집별 총 소비금액은 군집별 고객 수와 반대의 순서로, 군집 0 (로열 고객) > 군집 2 (범용 고객) > 군집 1 (고성과 고객) > 군집 3 (저성과 고객) 으로 나타났다. 평균 소비금액도 동일한 양상을 보였으며 군집 0 (로열 고객)이 수는 적지만 많은 부분의 거래금액을 차지하고 있음을 알 수 있다.

3. 군집별 마케팅 반응도

마케팅 반응도는 온라인 반응도와 오프라인 반응도 2개로 나누어 파생변수를 생성하였다.

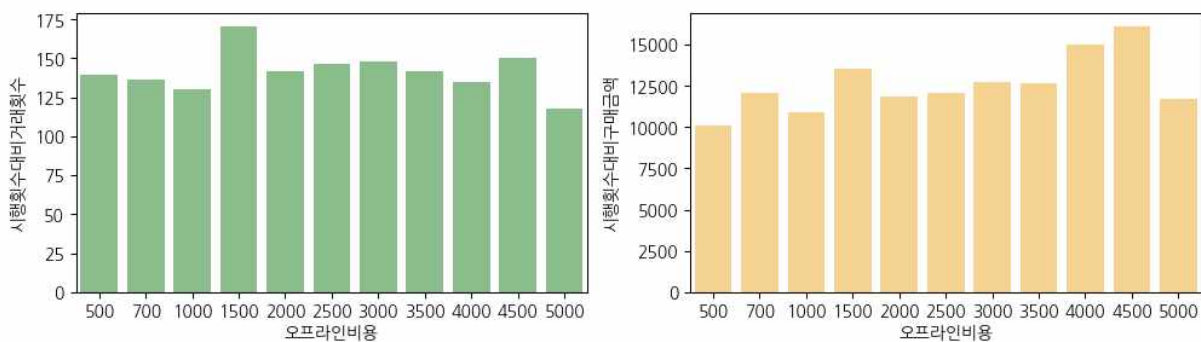
온라인 마케팅 비용은 매일 바뀌며 온라인 마케팅 특성상 고객들 반응이 빠를 것이라 판단하여 하루만에 마케팅 효과가 나타난다고 가정하였다. 따라서 온라인 마케팅에 반응하는 정도를 의미하는 '온라인 반응도' 변수는 다음과 같이 계산하여 고객별로 산출하였다.

$$\text{온라인 반응도} = \frac{\text{고객별 구매일자의 온라인 비용 총합}}{\text{전체 온라인 비용 총합}} \rightarrow \text{MinMax Scaling (0-1)}$$

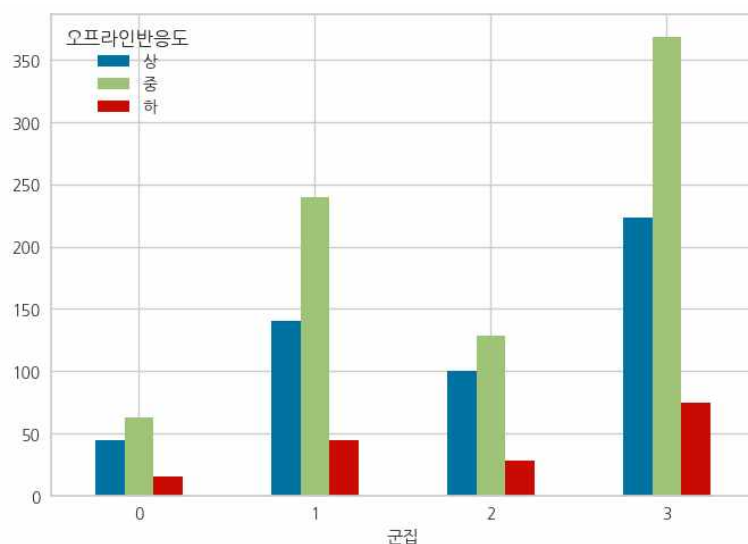


Clustering 이후, 군집별 온라인 반응도를 살펴보았을 때 로열고객인 군집 0이 가장 반응이 컸으며 그 다음으로 는 군집 2 (범용 고객), 군집 1 (고성과 고객), 군집 3 (저성과 고객) 순으로 나타났다.

오프라인 마케팅 비용은 일주일 간격으로 달라진다는 것을 확인하였고, 각 오프라인 마케팅 비용별 거래횟수와 구매금액을 그래프로 그려보았을 때 다음과 같았다.



오른쪽 그래프에서 대체적으로 오프라인마케팅 비용이 높을수록 구매금액이 높아지는 추세를 볼 수 있었다. 따라서 오프라인 마케팅 비용을 상(3500~5000) / 중(1500~3000) / 하(500~1000)구간으로 나누어 고객ID마다 오프라인마케팅 비용별 시행횟수 대비 구매금액 비율을 산출하였다. 그리고 이 비율의 비교를 통해 각각 오프라인 마케팅 비용이 상 / 중 / 하 구간에서 어떤 때 가장 많이 반응하는지 할당해주었다.

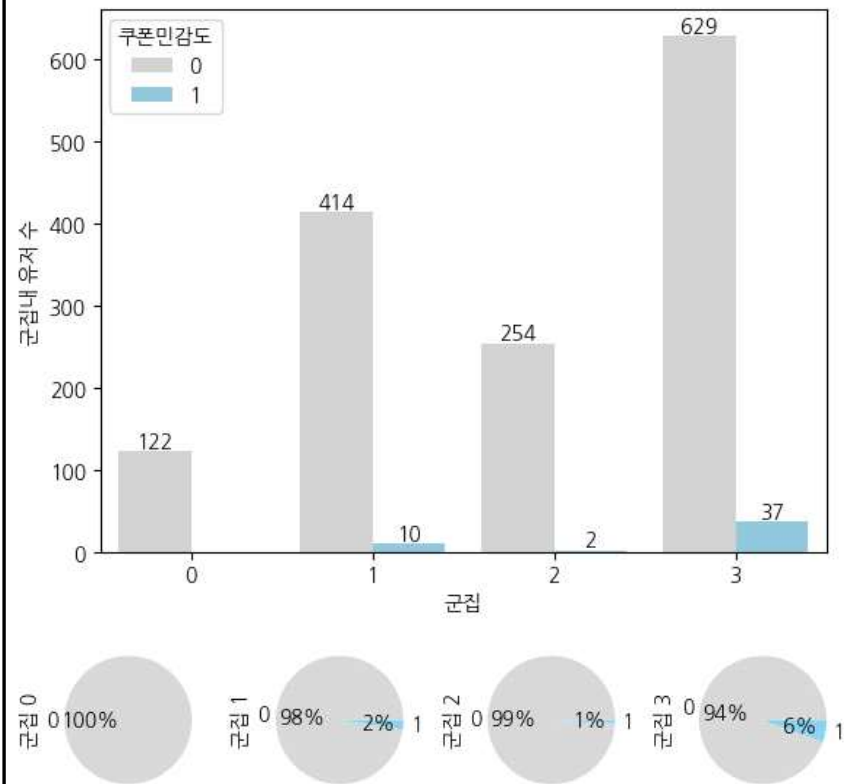


이렇게 해서 나눈 오프라인 반응도는 군집별 막대그래프를 그려보았을 때 큰 차이가 나지는 않았다.

마케팅 반응도의 두 변수(온라인 반응도, 오프라인 반응도)는 고객이 온라인/오프라인 마케팅에 모두 반응을 한다는 가정 하에 생성해냈기 때문에 정확한 마케팅 반응도를 측정한다고 하기에는 한계가 있다. 추후 이커머스의 온라인/오프라인 마케팅 소요 시간을 반영하거나 해당 마케팅 비용의 상세 내역을 참고하는 식으로 보완할 수 있을 것이다.

4. 군집별 쿠폰민감도

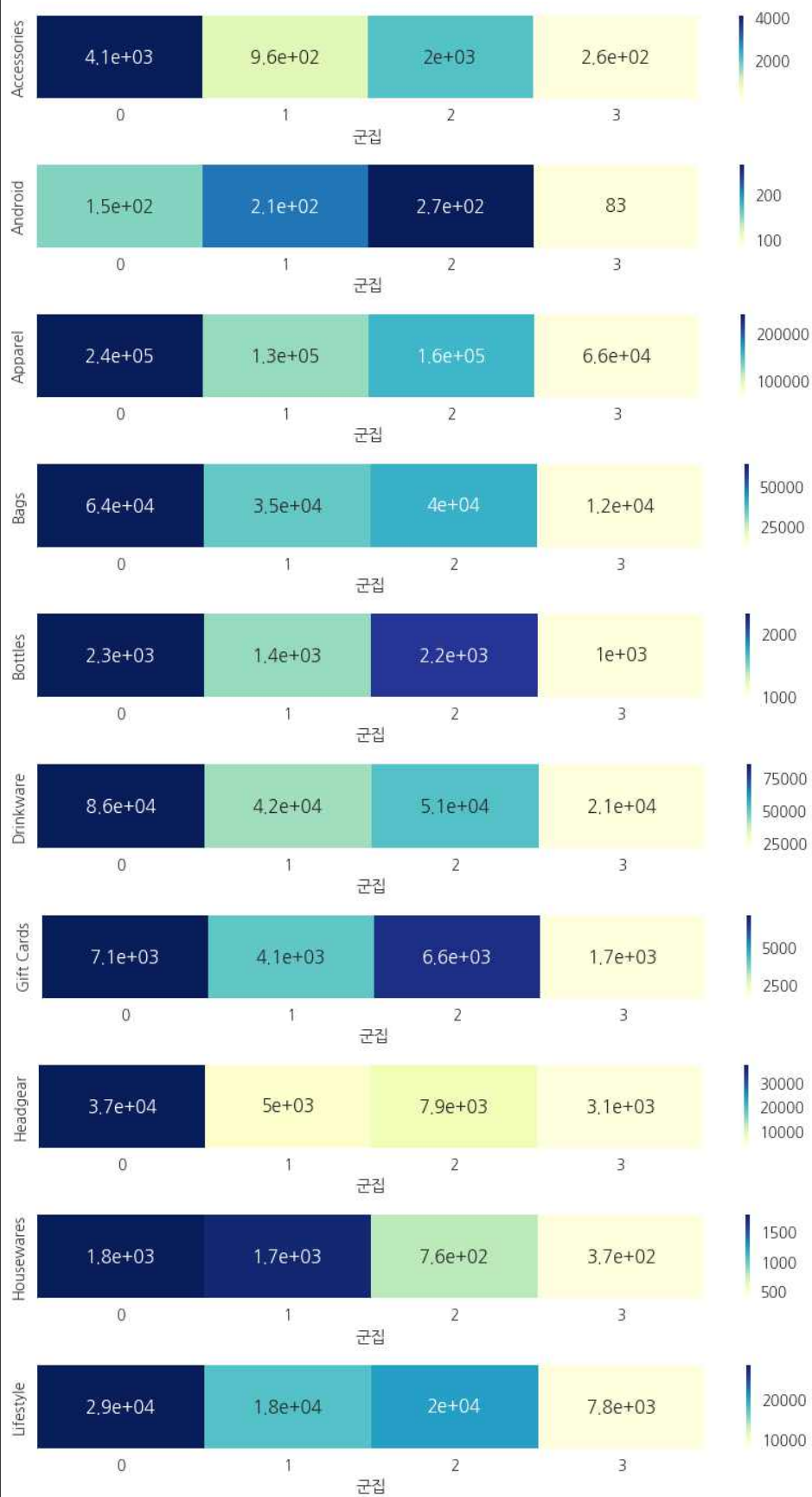
쿠폰민감도는 12개월을 4개월씩 나눠 Time Series Clustering을 적용하였다. 쿠폰 데이터를 보았을 때 각각 10% 쿠폰 : 1월,4월,7월, 10월 / 20% 쿠폰 : 2월, 5월, 8월, 11월 / 30% 쿠폰 : 3월, 6월, 9월, 12월에 이루어짐을 알 수 있었고, 각 쿠폰 할인율별 추세에 따라 군집이 될 수 있도록 하였다. 군집 0은 쿠폰할인율이 늘어나도 사용률이 늘지 않는 민감도가 낮은 고객이라고 할 수 있고, 군집 1은 쿠폰할인율이 늘어났을 때 사용률도 함께 늘어나는 추세를 보여 민감도가 높은 고객이라고 할 수 있다. 전체 고객 중에서 쿠폰민감도가 낮은 군집 0의 고객이 1,419명으로 대부분을 차지했으며 평균 쿠폰 사용률은 14%에 그쳤다. 쿠폰민감도가 높은 군집 1의 고객은 49명으로 91%의 높은 쿠폰 사용률을 보였다.

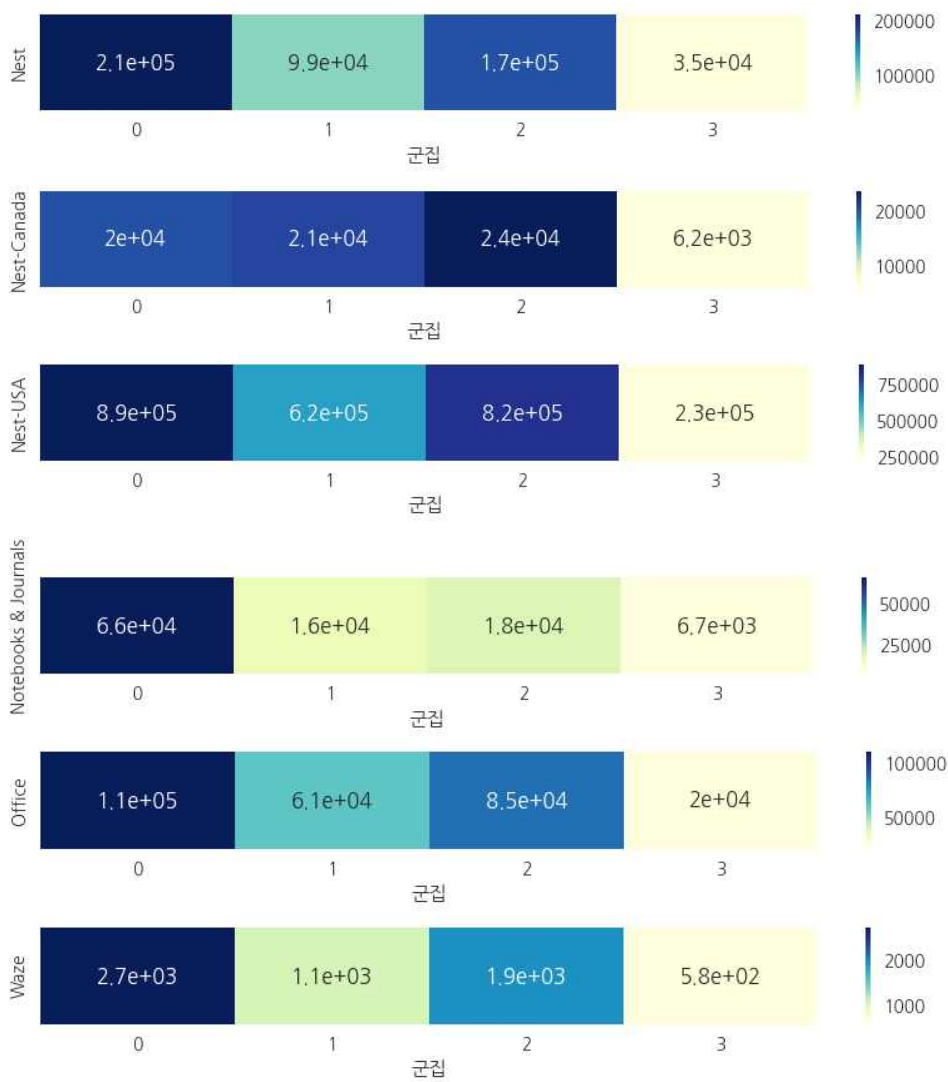


군집별 쿠폰민감도를 보았을 때, 군집 3 (저성과 고객)에 민감도가 높은 고객이 가장 많았으며 이에 해당하는 고객들은 더 좋은 쿠폰을 제공하는 타 경쟁사로 쉽게 넘어갈 수 있음을 예상할 수 있었다.

5. 군집별 카테고리 소비금액

군집별 카테고리 소비금액은 총 구매금액이 가장 많았던 군집 0 (로열 고객)에서 대부분 많이 나왔으며 Android와 Nest-Canada의 경우에는 군집 2 (범용 고객)가 가장 많이 나왔다. 군집 3 (저성과 고객)의 경우에는 모든 카테고리에서 소비금액이 미미하다는 것을 확인할 수 있다.

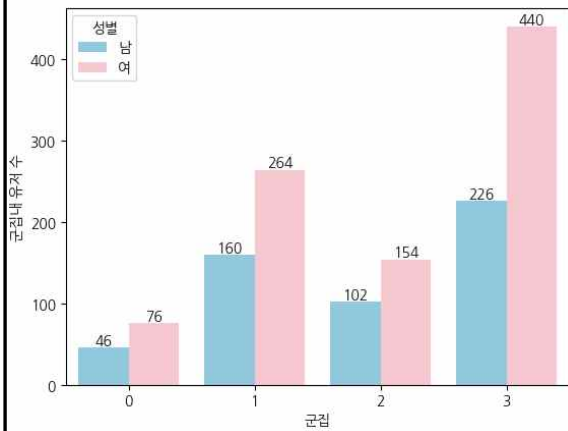




6. 군집별 고객정보

군집에 고객정보 데이터를 결합시켜보았을 때 군집마다 성별, 고객 지역, 가입 기간에 대한 차이가 크지 않음을 알 수 있었다.

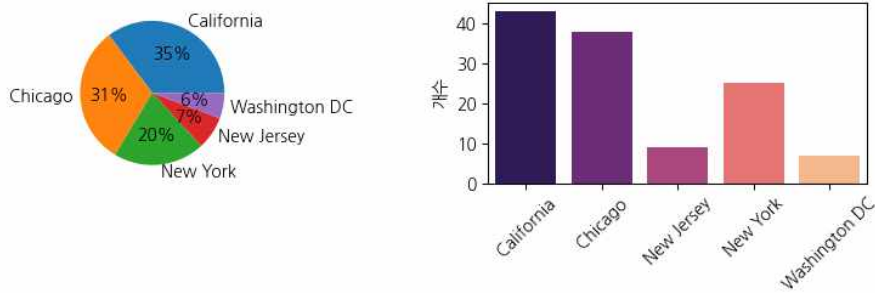
6-1. 성별



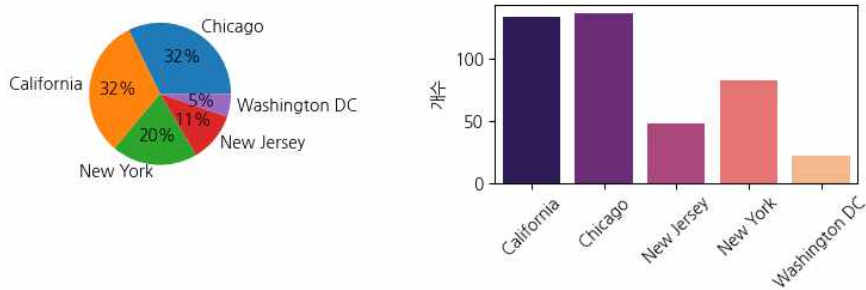


6-2. 고객 지역

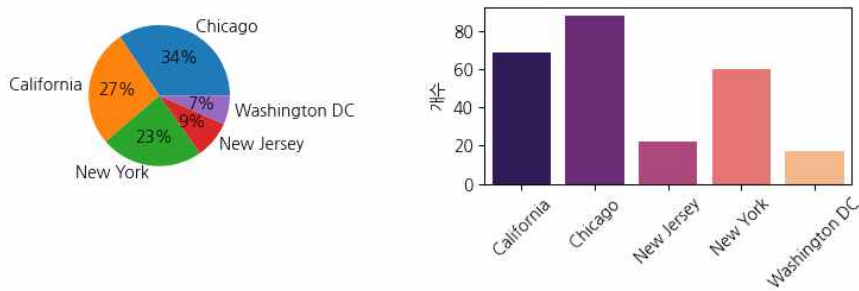
----- 군집 : 0 -----



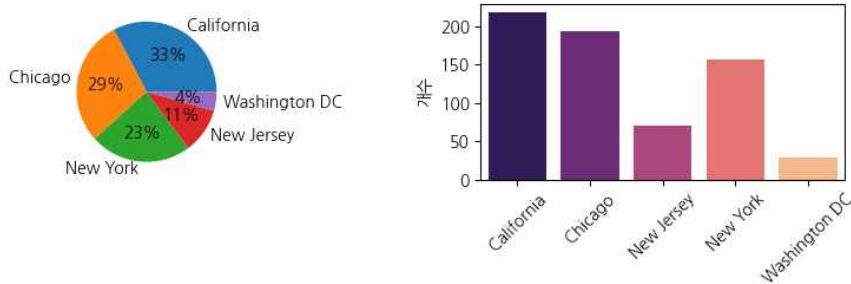
----- 군집 : 1 -----



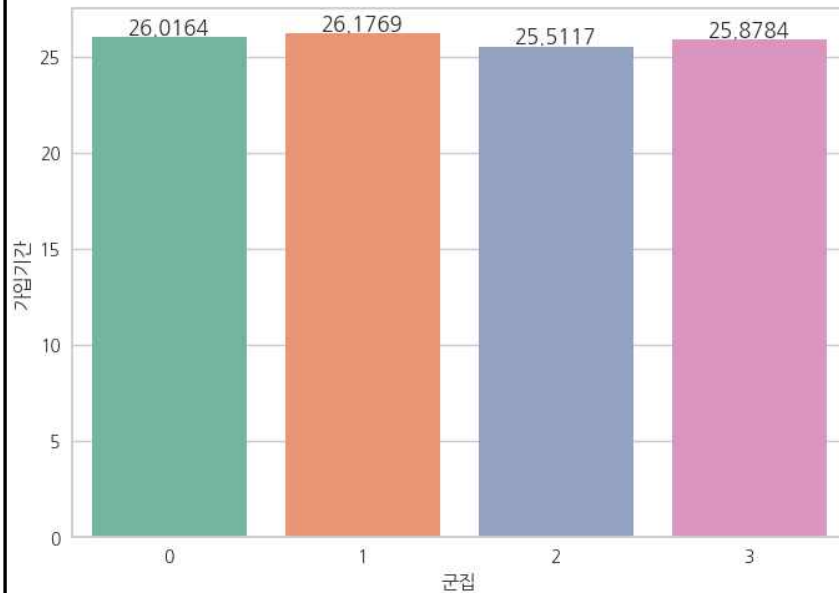
----- 군집 : 2 -----



----- 군집 : 3 -----



6-3. 가입기간



솔루션 제안

[고객세분화 분석 결과 기반 비즈니스 솔루션 제안]

고객세분화 분석 결과, 소비자 그룹을 L, R, F, M, V에 따라 크게 로열 고객, 고성과 고객, 범용 고객, 저성과 고객으로 나누었으며 각 그룹에 대한 비즈니스 솔루션 아이디어를 제안하였다.

1. Cluster 0 - 로열 고객

: L↑ R↓ F↑ M↑ V↑

로열 고객은 가입 기간이 길며 최근 자주 많은 비용의 제품을 구매하며 이익을 많이 벌어들여주는 고객으로 정의하였다.

[비즈니스 솔루션 아이디어]

- 초개인화 마케팅
 - LLM(Large Language Model) 기반 AI 챗봇 등을 통해 기업이 고객과의 소통을 확장
ex) 고객이 질문할 사항에 대해 미리 답변 생성 등
- 여러 외부 데이터를 결합하여 고객이 처한 상황을 예측하고 이에 맞는 상품 제안
 - 해당 고객이 좋아하는 취향과 구매패턴을 시계열로 분석하여 더욱 편리한 쇼핑경험을 제공함으로써 궁극적인 락인효과(Lock-in Effect) 추구
- B2B 컨설팅
 - 로열 고객에 대한 방대한 데이터를 분석하여 각 입점사별 매출 확대를 위한 정보 제공 가능
 - 이커머스 입점 브랜드 내 앱/웹에서 고객에게 우위를 차지하기 위해 광고 운영 컨설팅 진행
 - 입점 브랜드 수수료 정책 및 입점 브랜드사 대상으로 커뮤니티 운영

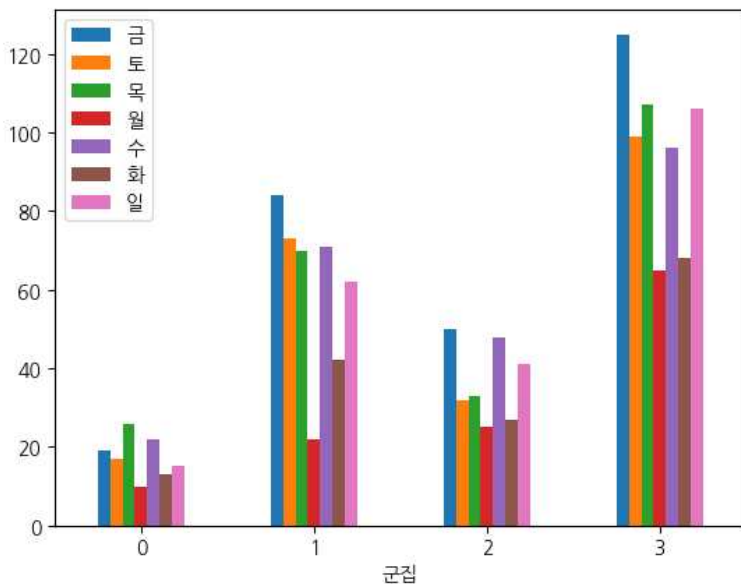
2. Cluster 1 - 고성과 고객

: L↓ R↑ F↓ M↓ V↑

고성과 고객은 가입 기간이 짧고 자주 구매하지 않으며 저비용 구매 고객이지만, 이익을 많이 벌어다주는 고객으로 정의하였다.

[비즈니스 솔루션 아이디어]

● 이용패턴 분석을 기반으로 쿠폰을 제공하는 유인 전략



○ Cluster 0을 제외한 Cluster 1, 2, 3은 모두 금요일이 가장 높은 이용건수 분석(Cluster 0의 경우 금요일이 이용건수 비율 3위 차지)

○ 반대로, 월요일, 화요일은 Cluster 0, 1, 2, 3 모두에서 비교적 낮은 이용건수 기록

- [Idea1] 상대적으로 이용건수 비율이 낮은 월요일, 화요일에 방문 유도 쿠폰 발행

- [Idea2] 이용건수 비율이 높은 수요일, 목요일, 금요일에 할인 쿠폰 발행하여 구매전환율 상승 유도

→ Cluster 1 고객은 본 이커머스에 대한 흥미를 갖게 하는 것이 중요

● Cluster 0 고객과의 차이점?

○ Recency, Frequency가 반대 방향으로, 이를 반대로 움직이게 하려면 고객이 흥미를 가지고 자주 들어오게 만드는 것이 중요

○ 그렇게 된다면, Cluster 0 고객으로 Shift 할 가능성이 가장 큰 고객군이라 생각

3. Cluster 2 - 범용 고객

: L↑ R↓ F↑ M↑ V↓

범용 고객은 가입 기간이 길며 최근 자주 많은 비용의 제품을 구매하였지만, 상대적으로 이익을 적게 벌어다주는 고객으로 정의하였다.

[비즈니스 솔루션 아이디어]

- 장바구니 분석을 통해 추천 제품 제공
- Cross & Up Selling 전략을 통해 추가 제품 구매 유도
 - Cross & Up Selling : 이미 고른 제품보다 더 비싸고 더 좋은 버전을 사게 하는 판매기법인 Cross Selling 전략과 관련된 다른 상품을 판매하는 기법인 Up Selling 전략 진행

-장바구니 분석 결과 :

1) 전체 카테고리에 대한 결과: (신뢰도 최소 0.9 이상 rules)

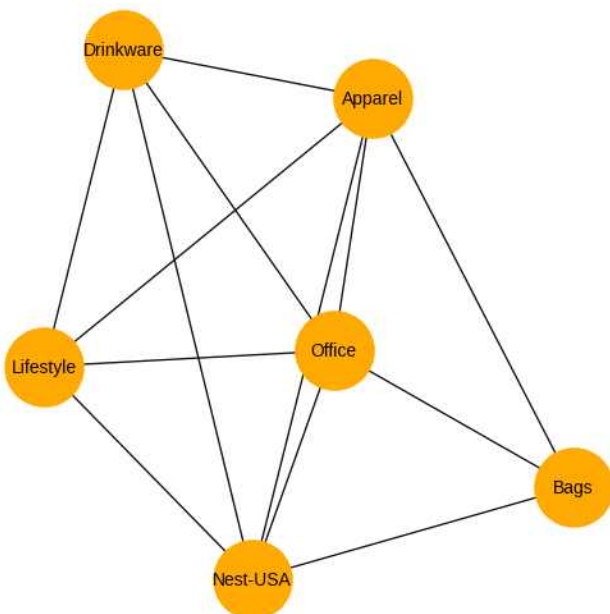
- lift가 모두 1이상이므로, 상품을 아래와 같은 순서의 조합으로 산 것이 필연적이라고 할수있음. 신뢰도가 모두 0.9 이상이므로, 같이 구매하는 발생확률이 아주 높은 편임. 따라서 아래 rule들은 2군집 고객 대상으로 적용 가능한 rule들임.
- 인사이트: 기존의 전체군집에서 Apparel, Drinkware, Lifestyle, Nest-USA,Office가 전체 규칙을 장악
- 군집2에서는 기존과 달리 Bags 와 연관된 규칙이 높은 신뢰도로 나왔음.
 - Bags → Apparel
 - (Bags, Nest-USA) → Apparel (이게 밑에거보다 신뢰도,향상도 높음. 그러나 둘다 높음)
 - (Bags, Apparel) → Nest-USA

→ 가방 사는 고객에게는 의류 추천

→ 가방, 의류를 산 고객에게는 Nest-USA 추천

→ 가방, Nest-USA 산 고객에게는 의류 추천

가방을 산 고객에게 구매 순서에 따라 다음 상품을 공격적으로 마케팅하여, 살 확률이 높은 카테고리를 고객에게 알리고, 구매 여정을 도움.



Drinkware → Apparel (기존 상위 rule과 동일)

Lifestyle → Apparel (기존 상위 rule과 동일)

Drinkware → Nest-USA (기존 상위 rule과 동일)

(Drinkware, Apparel) → Nest-USA (기존 상위 rule과 동일)

그 이외는 기존의 전체 군집에 장바구니 분석 했을때와 거의 유사.

Drinkware를 구매한 고객에게 특히 다음 추천 규칙의 상품들을 추천해주기

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Bags)	(Apparel)	0.378016	0.819035	0.343164	0.907801	1.108379	0.033555	1.962776	0.157210
1	(Drinkware)	(Apparel)	0.495979	0.819035	0.449062	0.905405	1.105454	0.042838	1.913060	0.189266
2	(Lifestyle)	(Apparel)	0.434316	0.819035	0.391421	0.901235	1.100362	0.035701	1.832272	0.161235
3	(Bags)	(Nest-USA)	0.378016	0.819035	0.345845	0.914894	1.117039	0.036236	2.126340	0.168454
4	(Drinkware)	(Nest-USA)	0.495979	0.819035	0.454424	0.916216	1.118654	0.048200	2.159907	0.210444
5	(Bags, Nest-USA)	(Apparel)	0.345845	0.819035	0.320375	0.926357	1.131034	0.037117	2.457316	0.177104
6	(Bags, Apparel)	(Nest-USA)	0.343164	0.819035	0.320375	0.933594	1.139871	0.039312	2.725122	0.186816
7	(Nest-USA, Drinkware)	(Apparel)	0.454424	0.819035	0.420912	0.926254	1.130909	0.048723	2.453887	0.212171
8	(Drinkware, Apparel)	(Nest-USA)	0.449062	0.819035	0.420912	0.937313	1.144412	0.053114	2.886825	0.229044
9	(Office, Drinkware)	(Apparel)	0.416890	0.819035	0.388740	0.932476	1.138506	0.047292	2.680008	0.208633
10	(Nest-USA, Lifestyle)	(Apparel)	0.387399	0.819035	0.361930	0.934256	1.140679	0.044637	2.752575	0.201321
11	(Apparel, Lifestyle)	(Nest-USA)	0.391421	0.819035	0.361930	0.924658	1.128960	0.041343	2.401901	0.187698
12	(Office, Lifestyle)	(Apparel)	0.369973	0.819035	0.343164	0.927536	1.132475	0.040143	2.497319	0.185672
13	(Office, Nest-USA)	(Apparel)	0.564343	0.819035	0.521448	0.923990	1.128146	0.059231	2.380823	0.260732
14	(Office, Apparel)	(Nest-USA)	0.571046	0.819035	0.521448	0.913146	1.114904	0.053741	2.083545	0.240263
15	(Office, Drinkware)	(Nest-USA)	0.416890	0.819035	0.388740	0.932476	1.138506	0.047292	2.680008	0.208633
16	(Office, Lifestyle)	(Nest-USA)	0.369973	0.819035	0.337802	0.913043	1.114780	0.034781	2.081099	0.163425
17	(Office, Nest-USA, Drinkware)	(Apparel)	0.388740	0.819035	0.367292	0.944828	1.153587	0.048901	3.279993	0.217810
18	(Office, Drinkware, Apparel)	(Nest-USA)	0.388740	0.819035	0.367292	0.944828	1.153587	0.048901	3.279993	0.217810
19	(Office, Nest-USA, Lifestyle)	(Apparel)	0.337802	0.819035	0.319035	0.944444	1.153119	0.042364	3.257373	0.200524
20	(Office, Apparel, Lifestyle)	(Nest-USA)	0.343164	0.819035	0.319035	0.929688	1.135101	0.037972	2.573727	0.181204

2) 중저인기 카테고리는 구매 빈도수 자체가 낮고, 따라서 장바구니 규칙에서 등장 빈도가 낮아 활용하지 못하는 단점이 발생. 이를 보완하고자 중저인기 카테고리만 따로 제한하여 장바구니 분석을 하고, lift(=향상도)가 1 이상인 rule이 있는지 탐색

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Headgear)	(Bags)	0.265963	0.505041	0.150744	0.566787	1.122260	0.016422	1.142531	0.148413
1	(Bags)	(Headgear)	0.505041	0.265963	0.150744	0.298479	1.122260	0.016422	1.046351	0.220100
2	(Bags)	(Notebooks & Journals)	0.505041	0.232357	0.140182	0.277567	1.194568	0.022833	1.062579	0.329073
3	(Notebooks & Journals)	(Bags)	0.232357	0.505041	0.140182	0.603306	1.194568	0.022833	1.247710	0.212179

● 인사이트:

Headgears→ Bags : 신뢰도 0.56, 향상도 1이상

Notebooks & Journals → Bags : 신뢰도 0.6, 향상도 1이상

저인기 품목중, 헤드기어와 노트북&저널을 구매한 고객에게는 가방을 추천해준다.

또한, 이러한 순서대로 구매한 고객에게 가방 할인쿠폰을 제공한다.

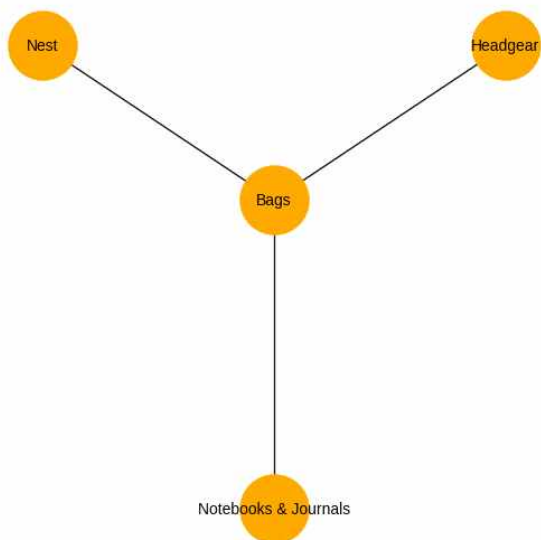
반대로, Bags → Headgears, Bags → Notebooks & Journals 향상도는 1이상이나, 신뢰도가 0.3 이하
Bags를 이미 산사람이 Headgears, Notebooks & Journals을 사는 것보다

Headgears, Notebooks & Journals를 산사람이 Bags를 사는게 더 자주, 필연적으로 발생.

Bags를 구매할 때에는 Headgears, Notebooks & Journals에 비해 비교적 쇼핑 탐색시간, 가격 비교 시간 등 소요시간이 길어질 것이라 생각.

Bags를 구매후 다른 작은 것들을 굳이 같이 이어서 사기 보다는 Headgears, Notebooks & Journals를 산사람이 Bags를 사는게 더 일반적.

→ 가능한 이유 : 이미 가방을 사기로 마음을 먹고 들어온 고객이 소요시간이 비교적 짧은것부터 사기 때문 or 가방을 살생각이 없다가 Headgears, Notebooks & Journals 사는 케이스



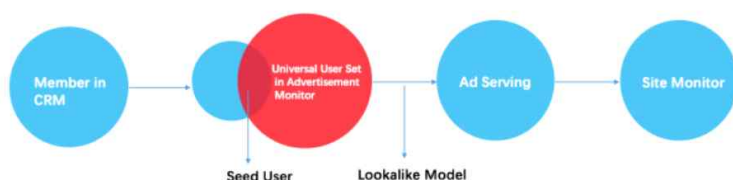
4. Cluster 3 - 저성과 고객

: L↓ R↑ F↓ M↓ V↓

저성과 고객은 가입 기간이 짧으며 구매한 이력이 없는 고객으로, 이익을 적게 벌어주는 고객으로 정의하였다.

[비즈니스 솔루션 아이디어]

- Lookalike 모델 기법 활용 : 기계 학습 알고리즘을 기반으로 시드 세트 간의 행동 패턴을 기반으로 모델 개선을 가능하게 함으로써 유사한 모델링 지원



● 고객 정보가 없는 Cluster 3의 고객을 유치하기 위해 Cluster 0,1,2의 일부 오디언스 풀을 제공하여, 해당 오디언스 풀에 가장 가까운 고객들을 원하는 수만큼 검색

○ Lookalike 모델 적용 방안 : Cluster 3 고객의 취향을 기반으로 유사한 취향을 가졌지만, 우수한 성과를 내는 로열 고객군과 네트워크로 관계를 연결하여, 이커머스에 호감도를 높이도록 유도

● 현재까지는 신규고객으로, 생애주기(L)이 길어질 때 RFMV를 모두 반전

