



롯데 고객 행동분석을 통한 개인화 상품 추천 및 서비스 제안

신지은, 배수연

목차

INDEX

STEP1
데이터 소개 및
탐색

분석 목적
데이터 소개
데이터 탐색

STEP2
군집분석

온라인 데이터
오프라인 데이터

STEP3
추천 시스템
모델링

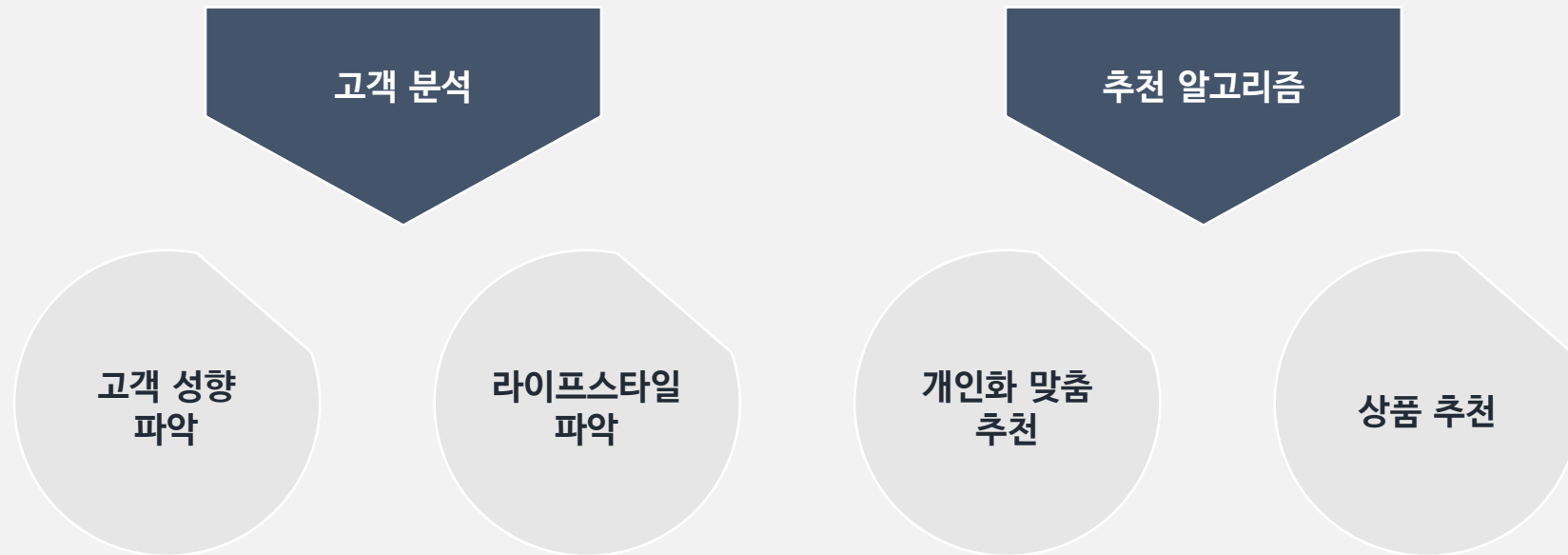
오프라인 : 연관성분석
온라인 : 모델기반 추천시스템

STEP4
서비스 제안

온라인 서비스 제안
오프라인 서비스 제안

데이터 소개 및 탐색

분석 목적

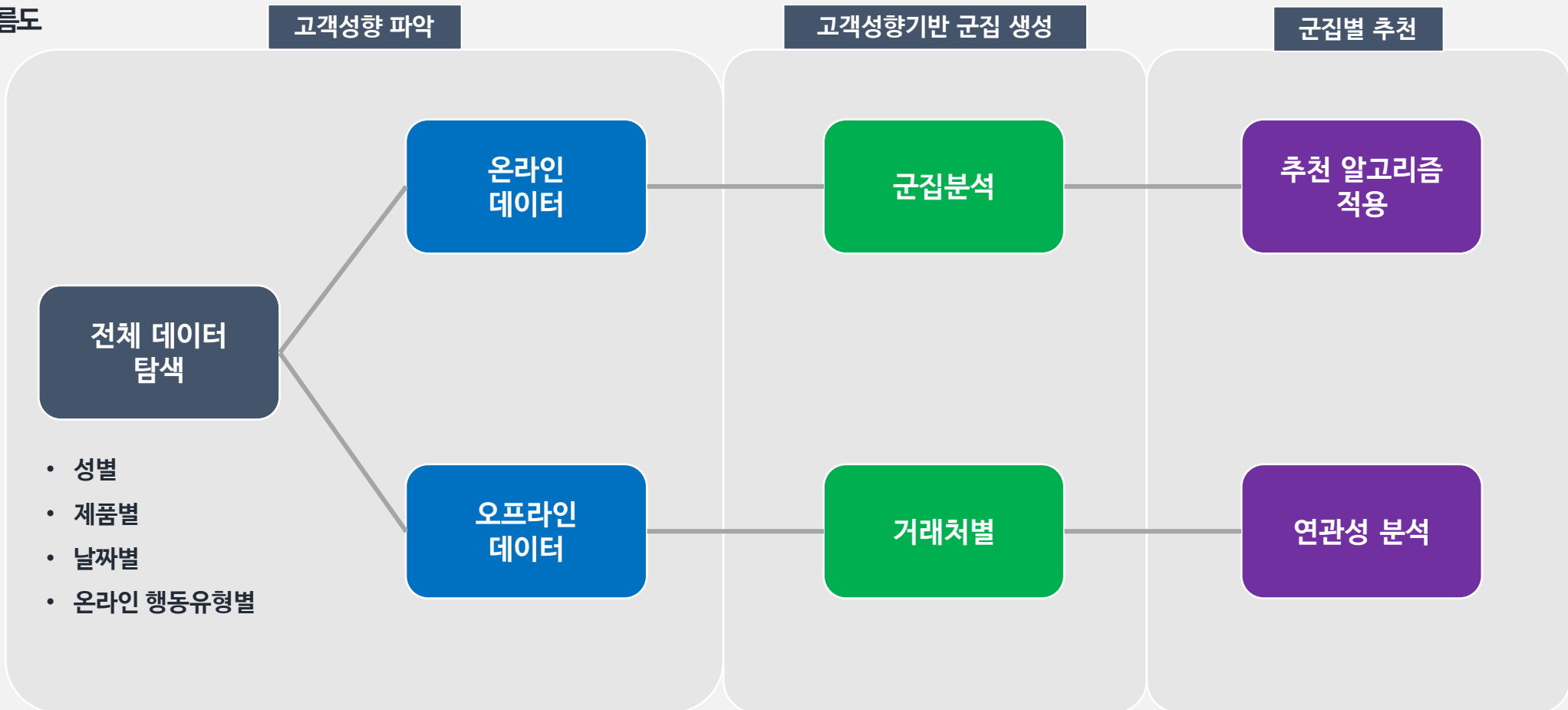


고객의 구매성향을 통해 라이프스타일을 파악하여
각 고객에게 맞춤상품을 추천하는 것이 목표

데이터 소개 및 탐색

분석 목적

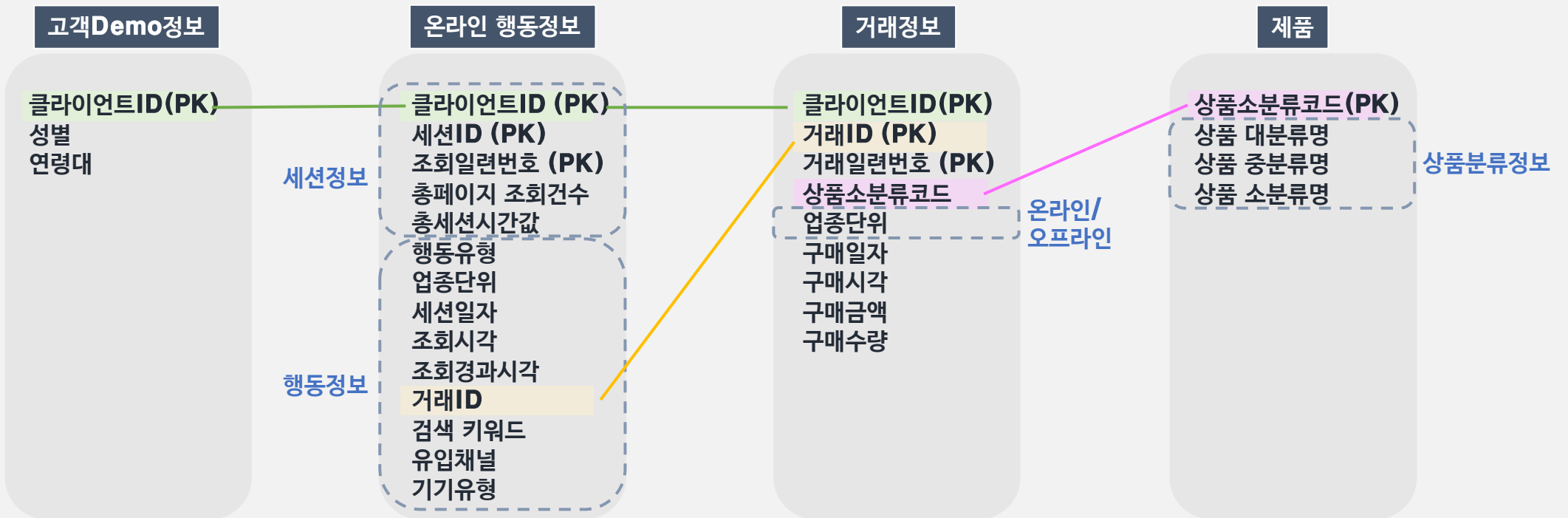
| 분석 흐름도



데이터 소개 및 탐색

데이터 소개

| 데이터 흐름도



2019년 7월~9월(3개월) 간 수집되었음

데이터 소개 및 탐색

데이터 탐색

| 전처리_온라인행동정보

clnt_id	sess_id	hit_seq	action_type	biz_unit	sess_dt	hit_tm	hit_pss_tm	trans_id	sech_kwd	tot_pag_view_ct	tot_sess_hr_v	trfc_src	dvc_ctg_nm	ox
7809	1	8	5	A03	20190913	01:16	2571103	NaN	NaN	34.0	2663.0	DIRECT	NaN	0.0
7809	1	4	2	A03	20190913	01:14	2485909	NaN	NaN	34.0	2663.0	DIRECT	NaN	0.0
7809	1	11	5	A03	20190913	01:17	2646597	NaN	NaN	34.0	2663.0	DIRECT	NaN	0.0
7809	1	1	2	A03	20190913	00:46	788304	NaN	NaN	34.0	2663.0	DIRECT	NaN	0.0
7809	1	9	5	A03	20190913	01:17	2617609	NaN	NaN	34.0	2663.0	DIRECT	NaN	0.0
...
31040	1	1	0	A02	20190815	03:47	13525	NaN	다이슨 에어랩	4.0	62.0	unknown	mobile_web	0.0
4129	1	2	0	A02	20190721	01:22	250099	NaN	여성메탈 시계	3.0	250.0	unknown	mobile_web	0.0
4129	1	1	0	A02	20190721	01:18	1525	NaN	여성메탈 시계	3.0	250.0	unknown	mobile_web	0.0
54403	3	1	0	A02	20190731	20:14	16905	NaN	비비고왕 교자	3.0	68.0	unknown	mobile_app	1.0
54403	4	1	0	A02	20190818	16:35	82700	NaN	에포테쉬 볼룸팝브 러쉬	5.0	101.0	unknown	mobile_app	1.0

✓ 'ox' 변수 생성
:구매한 고객은 1, 구매하지 않은 고객은 0의 ox 변수 생성

데이터 소개 및 탐색

데이터 탐색

| 전처리_거래정보

✓ 상품코드가 “unknown”인 행 제거

clnt_id	trans_id	trans_seq	biz_unit	pd_c	de_dt	de_tm	buy_am	buy_ct
1767	42911	107571	11	A03	1183	20190923 15:24	0	0
1830	42911	107571	12	A03	1213	20190923 15:24	0	0
1841	42911	107571	9	A03	1215	20190923 15:24	0	0
1845	42911	107571	2	A03	1213	20190923 15:24	0	0
1851	42911	107571	10	A03	1189	20190923 15:24	0	0
...
599845	33368	1989	2	B03	unknown	20190807 17:01	0	1
599865	33296	2369	2	B03	unknown	20190901 16:30	0	1
599866	21682	2136	4	B03	unknown	20190901 15:54	0	1
599934	65411	3740	2	B03	unknown	20190927 18:00	0	1
599942	16091	5636	3	B03	unknown	20190927 21:08	0	1

clnt_id	trans_id	trans_seq	biz_unit	pd_c	de_dt	de_tm	buy_am	buy_ct
0	21922	104999	1	A03	unknown	20190920 12:41	5990	1
1	21279	104907	4	A03	unknown	20190920 10:27	10900	1
2	39423	105124	11	A03	unknown	20190920 17:26	12900	1
3	18362	104010	1	A03	unknown	20190920 09:57	9900	1
9	58071	104894	2	A03	unknown	20190920 10:12	4480	1
...
599923	67063	14000	7	B03	unknown	20190901 13:35	50	1
599933	42297	12534	3	B03	unknown	20190927 18:53	100	1
599934	65411	3740	2	B03	unknown	20190927 18:00	0	1
599942	16091	5636	3	B03	unknown	20190927 21:08	0	1
599943	20325	37494	2	B03	unknown	20190927 13:50	2500	1

✓ 상품 구매금액과 구매량이 0인 행 제거

데이터 소개 및 탐색

데이터 탐색

| 전처리_거래정보

	clnt_id	trans_id	trans_seq	biz_unit	pd_c	de_dt	de_tm	buy_am	buy_ct
19237	50600	102331	5	A03	0382	20190917	14:19	100000016899	1
163188	50841	85664	3	A03	unknown	20190827	15:04	100000007199	4
285349	43143	127882	1	B01	0237	20190801	15:16	11800000	500
501157	5492	142744	1	B01	0158	20190827	13:44	11578450	235
83058	43143	74878	1	A03	0157	20190813	16:03	10500000	500
151228	43143	84804	1	A03	0157	20190826	17:09	10500000	500
154811	43143	84815	1	A03	0157	20190826	17:16	10500000	500
41677	63500	65340	1	A03	0237	20190808	11:33	10121800	520
52407	63500	71894	1	A03	0237	20190809	13:20	10121800	520
39779	63500	70312	1	A03	0237	20190807	15:29	10121800	520
51166	63500	68266	1	A03	0237	20190809	11:31	10121800	520
40081	63500	65329	1	A03	0237	20190807	11:30	10121800	520
48274	63500	68259	1	A03	0237	20190810	11:30	10121800	520
347184	70438	126347	1	B01	0158	20190902	17:11	5680000	100
510303	43143	127955	1	B01	0241	20190905	14:07	4935000	470

✓ 1건의 구매에서 구매액이 매우 큰 값 제거 (1천만원 초과)

데이터 소개 및 탐색

데이터 탐색

| 전처리_거래정보

	clnt_id	trans_id	trans_seq	biz_unit	pd_c	de_dt	de_tm	buy_am	buy_ct
541716	43143	6453	1	B02	0241	20190904	17:44	4500	3600
407912	43143	6454	1	B02	1056	20190904	17:46	2550	1600
319230	21768	121679	1	B01	0304	20190727	16:45	98400	984
219178	64311	118788	1	B01	0721	20190814	20:11	93100	931
221748	5370	116890	1	B01	0320	20190829	20:53	93000	930
...
349423	7740	123215	1	B01	0546	20190928	17:32	5051	0
288164	72267	124955	15	B01	1604	20190901	17:32	899	0
277857	29530	140873	2	B01	0970	20190720	21:32	7637	0
267753	18947	115960	2	B01	1142	20190809	21:11	11260	0
332202	48481	137880	9	B01	1604	20190818	18:13	1603	0

✓ 1건의 구매에서 구매량이 매우 큰 값 제거 (1,600개 초과)

데이터 소개 및 탐색

데이터 탐색

| 전처리_거래정보

	clnt_id	trans_id	trans_seq	biz_unit	pd_c	de_dt	de_tm	buy_am	buy_ct
0	21922	104999	1	A03	unknown	20190920	12:41	5990	1
1	21279	104907	4	A03	unknown	20190920	10:27	10900	1
2	39423	105124	11	A03	unknown	20190920	17:26	12900	1
3	18362	104010	1	A03	unknown	20190920	09:57	9900	1
4	39423	105124	13	A03	0565	20190920	17:26	2990	1
...
599956	35311	2839	3	B03	0339	20190927	14:09	3500	1
599957	35311	2839	4	B03	0339	20190927	14:09	3600	2
599958	35311	2839	5	B03	0339	20190927	14:09	3500	1
599959	35311	2839	2	B03	0339	20190927	14:09	4500	1
599960	35311	2839	6	B03	0324	20190927	14:09	6800	1

- ✓ **pd_c : str → int**로 변환
- ✓ **de_dt : int → date type**으로 변환

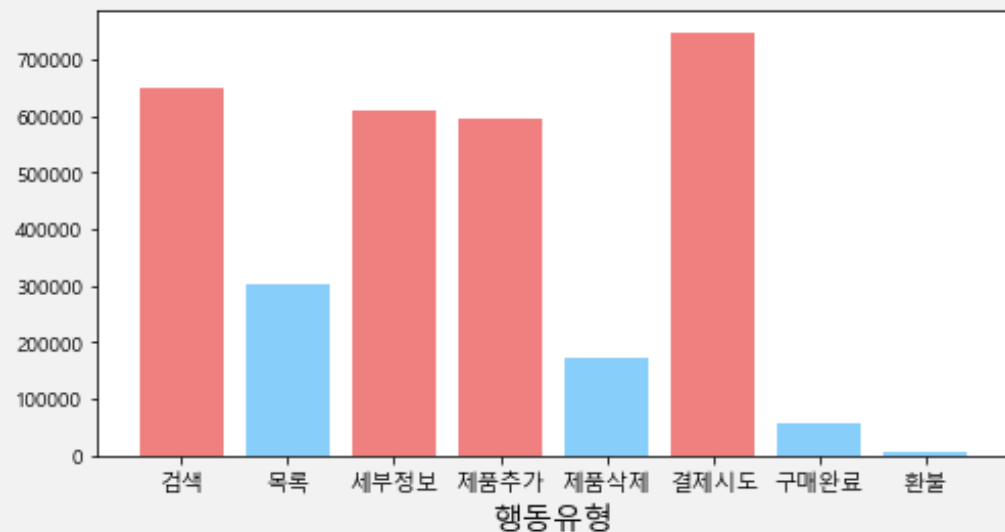
데이터 소개 및 탐색

데이터 탐색

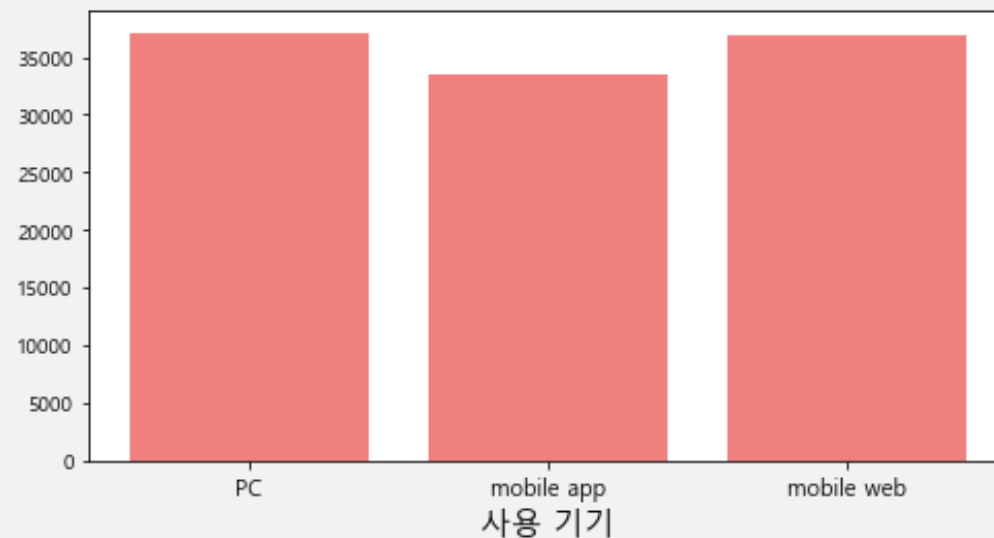
| EDA 온라인 행동유형별

롯데그룹 온라인 계열사 3곳 기준
이용 고객 67,185명 대상

행동유형 별 빈도



사용 기기별 빈도



- ✓ 결제 시도에 비해 구매완료의 빈도가 매우 낮음.
- ✓ 사용 기기는 PC, 모바일 웹, 모바일 앱 순으로 많다.

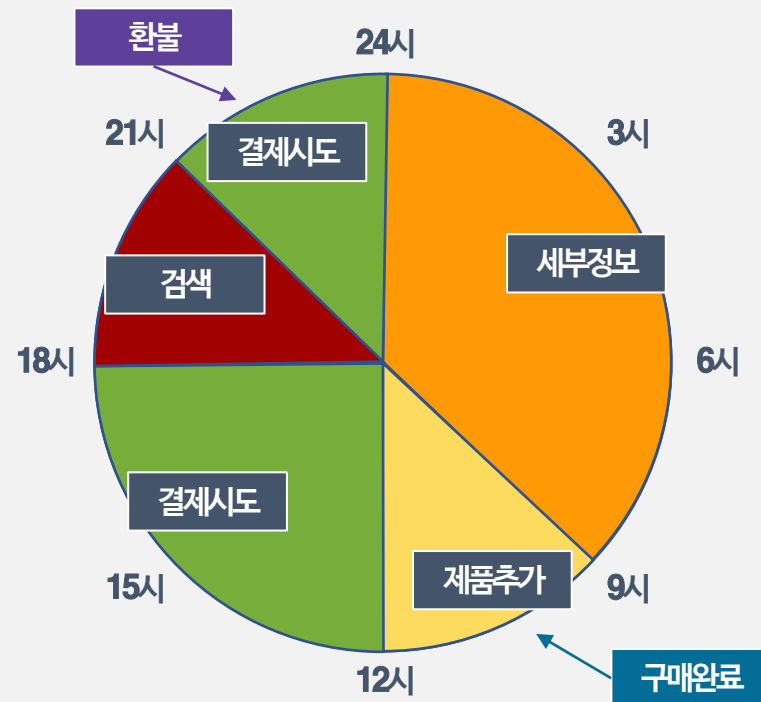
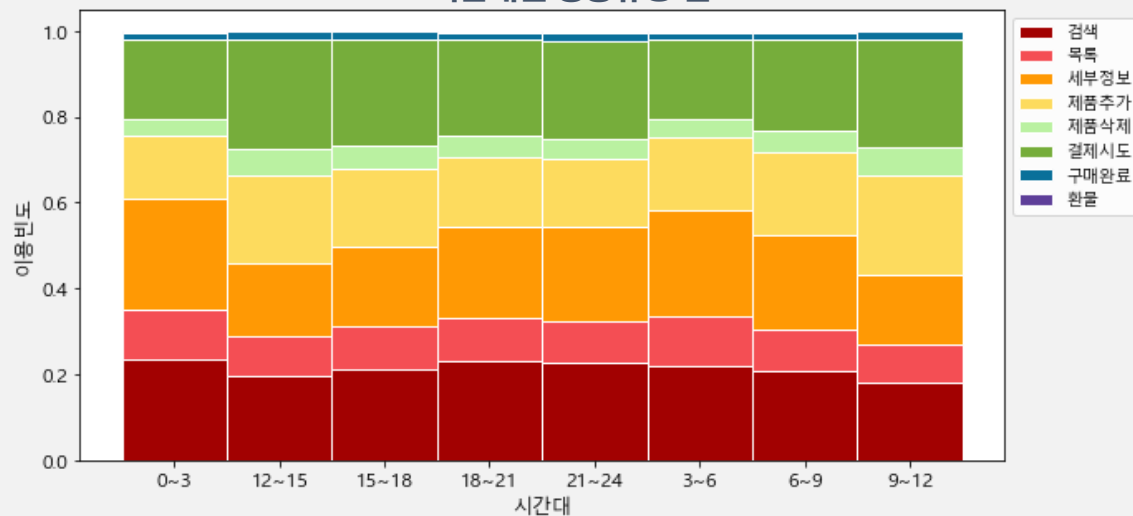
데이터 소개 및 탐색

데이터 탐색

EDA 온라인 행동유형별

롯데그룹 온라인 계열사 3곳 기준
이용 고객 67,185명 대상

시간대별 행동유형 빈도



- ✓ 새벽 시간대에는 주로 제품의 세부정보를 본다.
- ✓ 오전 9~12시에 제품 추가, 구매 완료하는 빈도가 가장 높다.
- ✓ 오후 시간대에는 결제를 시도하는 비중이 높다.
- ✓ 특히 하루 일과가 끝나는 시간대인 21~24시에 환불 빈도가 가장 높다.

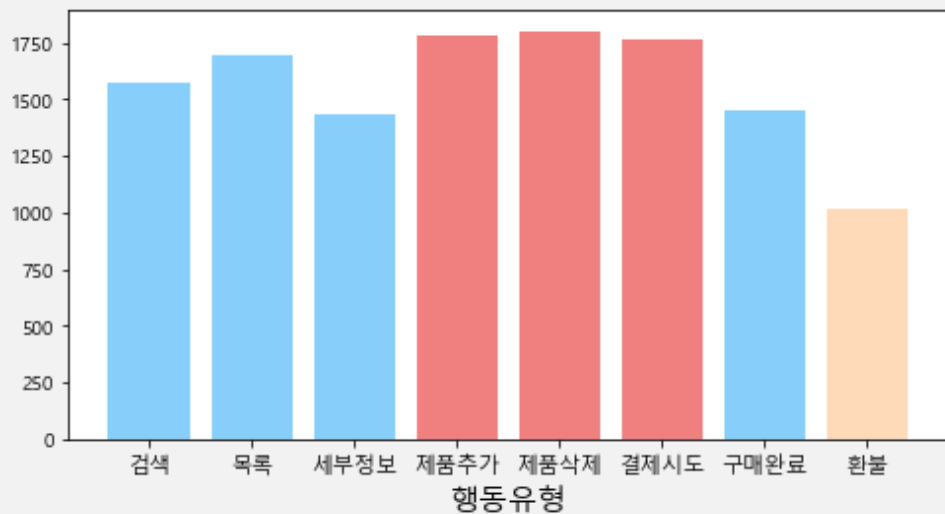
데이터 소개 및 탐색

데이터 탐색

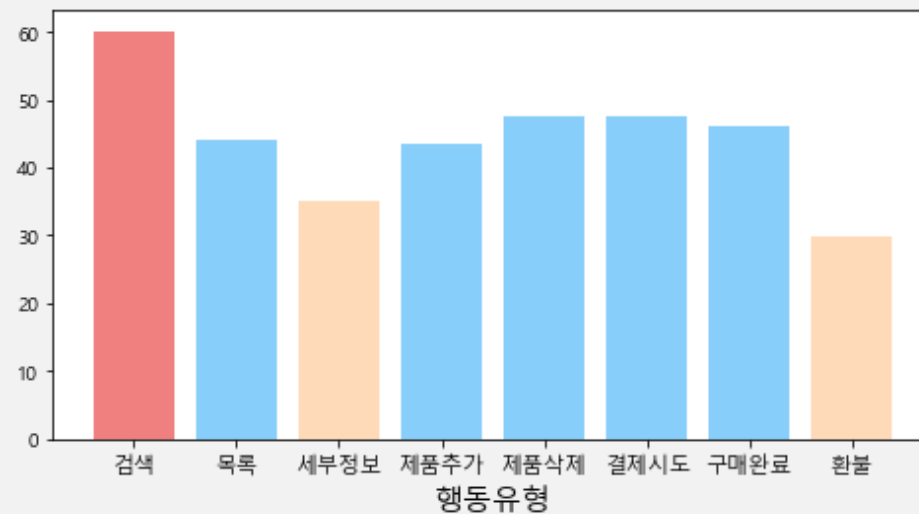
| EDA 온라인 행동유형별

롯데그룹 온라인 계열사 3곳 기준
이용 고객 67,185명 대상

행동유형별 세션시간 평균



행동유형별 페이지조회수 평균



- ✓ 제품 정보를 살피는 것 보다 결제행동을 할 때 페이지에 머무르는 시간이 김.
- ✓ 반면 환불할 때는 페이지에 머무르는 시간이 가장 짧음.
- ✓ 페이지 조회 수는 검색을 할 때 가장 많고, 세부정보를 보거나 환불할 때 가장 낮다.

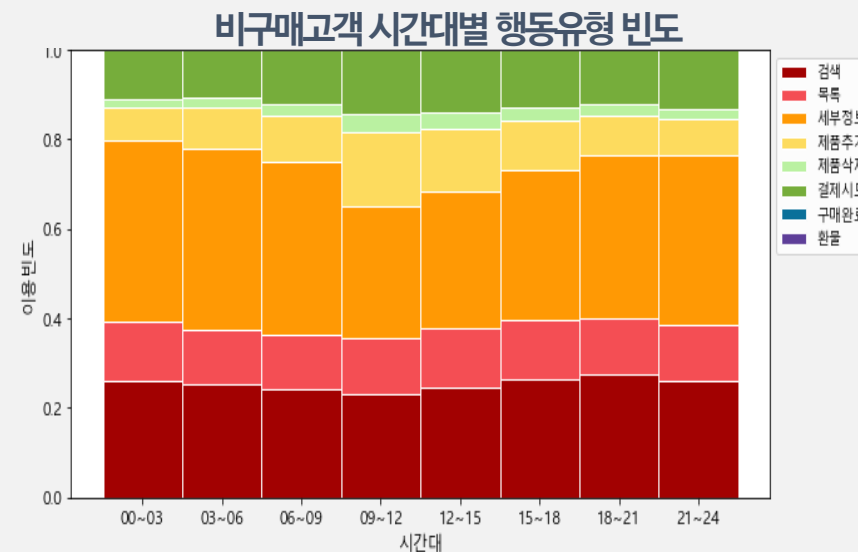
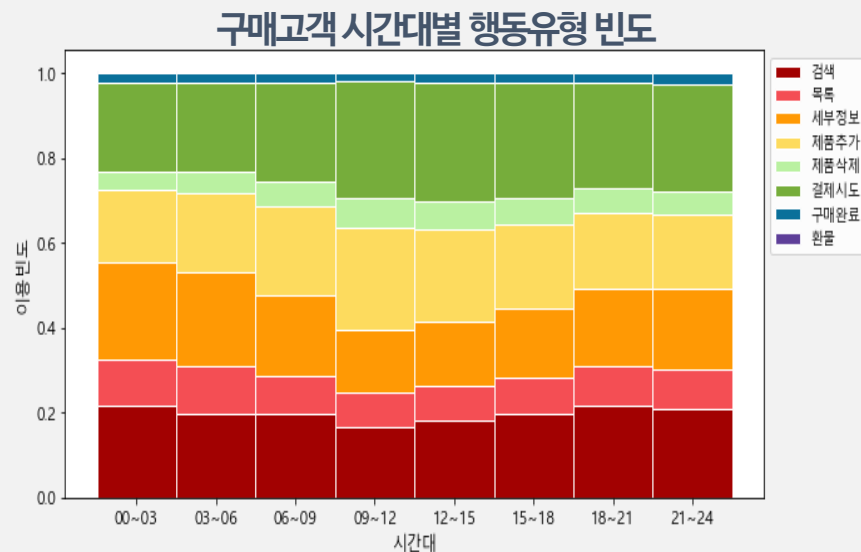
데이터 소개 및 탐색

데이터 탐색

EDA_구매고객/비구매고객 별 시간대별 행동유형 빈도

구매고객 Client Id : 20833개

비구매고객 Client Id : 51566개



- ✓ 비구매고객은 구매고객에 비해 **검색, 세부 정보 확인, 제품 목록**의 비율이 큼
→ 구매와 직접적으로 관련 없는 행동을 통해 구매에 목적이 있어 보이지 않음
- ✓ 비구매고객이 검색의 비율이 큰 점에서 **원하는 상품이 없거나, 구매 목록을 정하지 않고 즉흥적으로 들어온 것으로 보임**
→ 전자의 경우 **검색 체계 보완 필요**

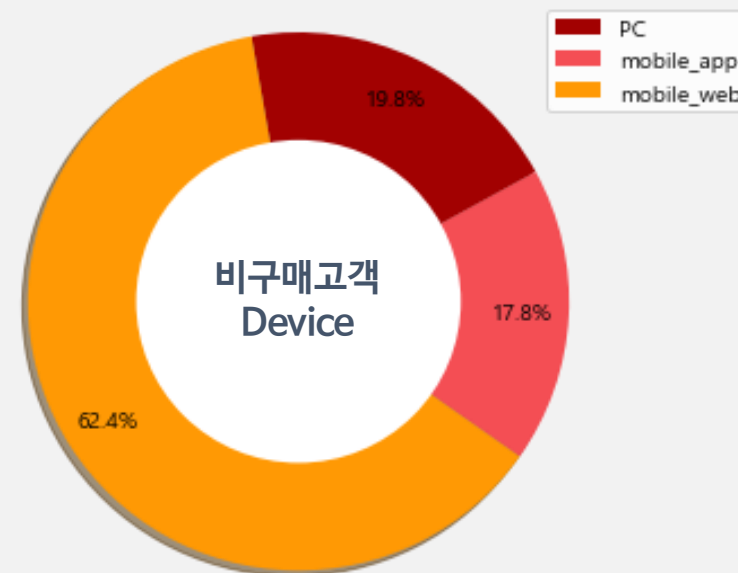
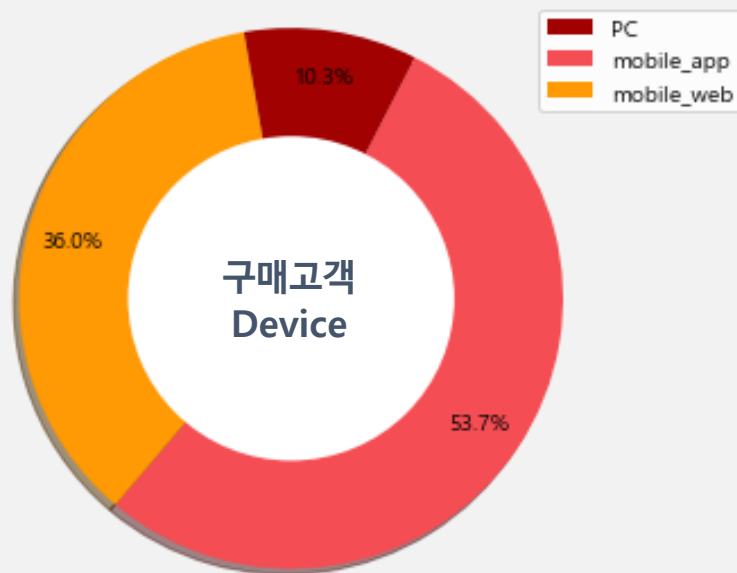
데이터 소개 및 탐색

데이터 탐색

| EDA_구매고객/비구매고객 별 Device

구매고객 Client Id : 20833개

비구매고객 Client Id : 51566개



- ✓ 구매고객일수록 APP의 비중이 크며 PC와 mobile web의 비중이 작음
- ✓ PC와 web의 쇼핑 환경이 불편함을 의미 → 개선 필요
- ✓ Mobile web으로 들어온 고객에게 APP 설치 유도

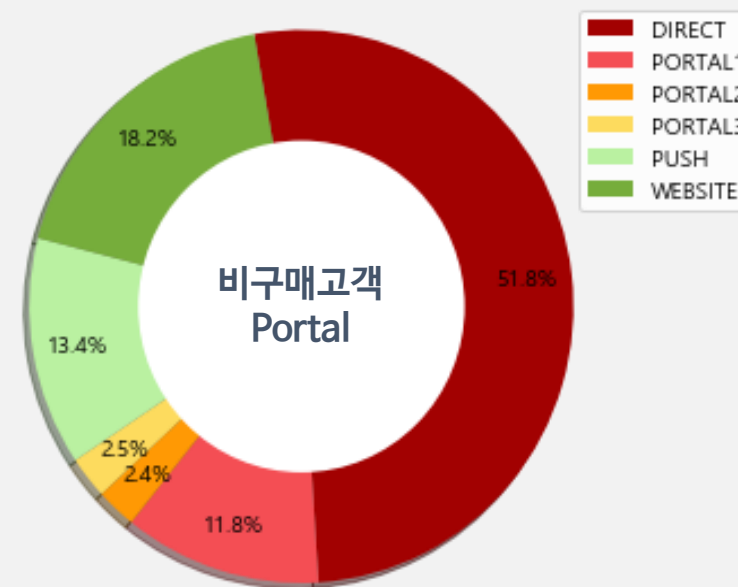
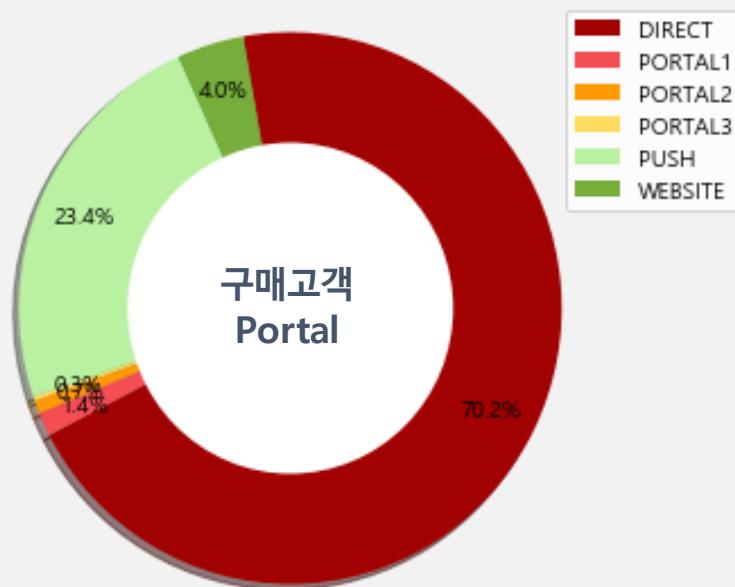
데이터 소개 및 탐색

데이터 탐색

EDA 온라인 행동유형별

구매고객 Client Id : 20833개

비구매고객 Client Id : 51566개



- ✓ 구매고객일수록 **DIRECT**, **PUSH**의 비율이 높음
- ✓ 고객의 성향을 파악해 주기적으로 **고객맞춤 PUSH** 알람을 넣도록 함
- ✓ 구매고객 중 다른 PORTAL로 부터의 유입 비중이 적음
→ PORTAL 사이트의 광고 비중을 줄일 필요가 있음

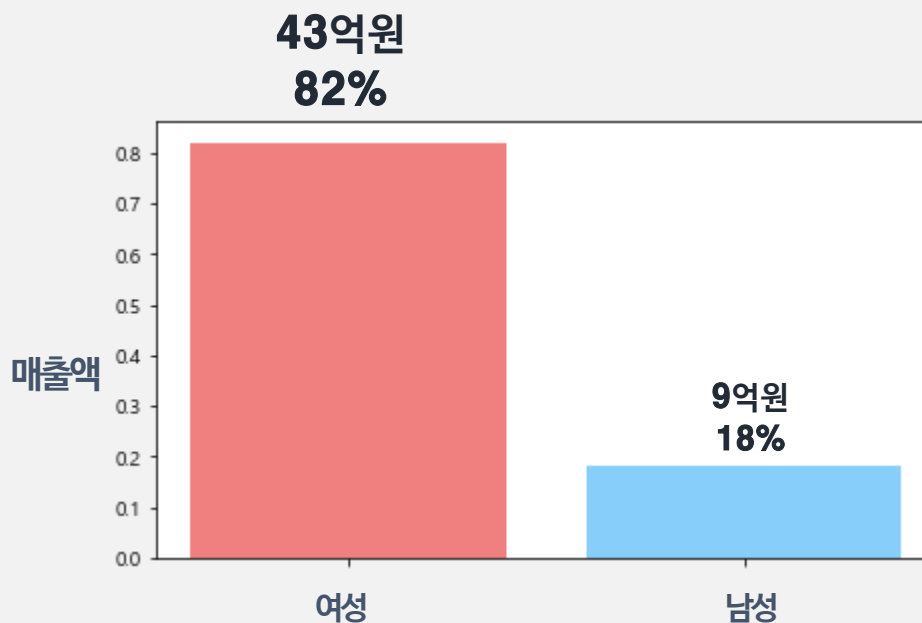
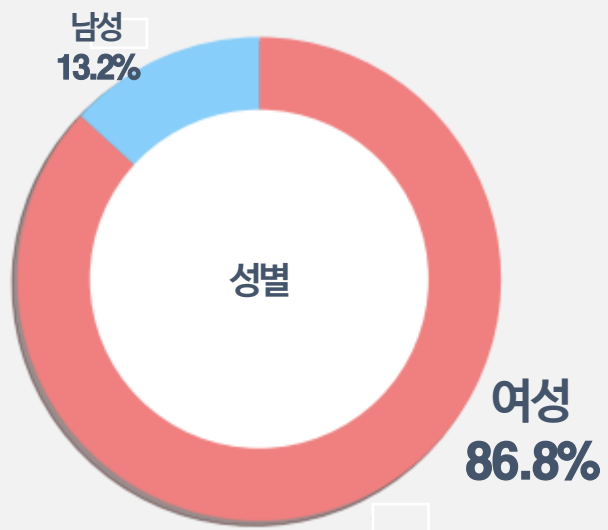
데이터 소개 및 탐색

데이터 탐색

| EDA_성별

여성 : 499,505명

남성 : 76,009명



✓ 전체 고객 중 여성이 약 87%, 남성이 13% 차지하며 매출액도 여성의 비중이 훨씬 많다.

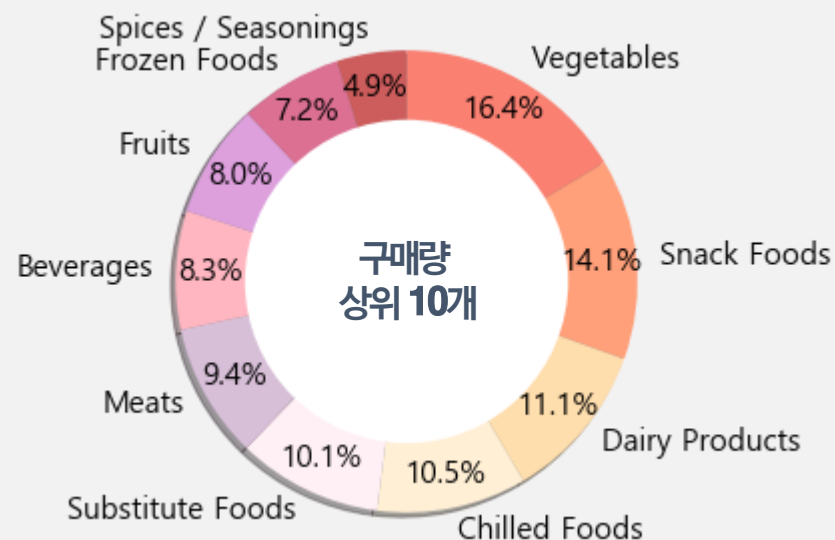
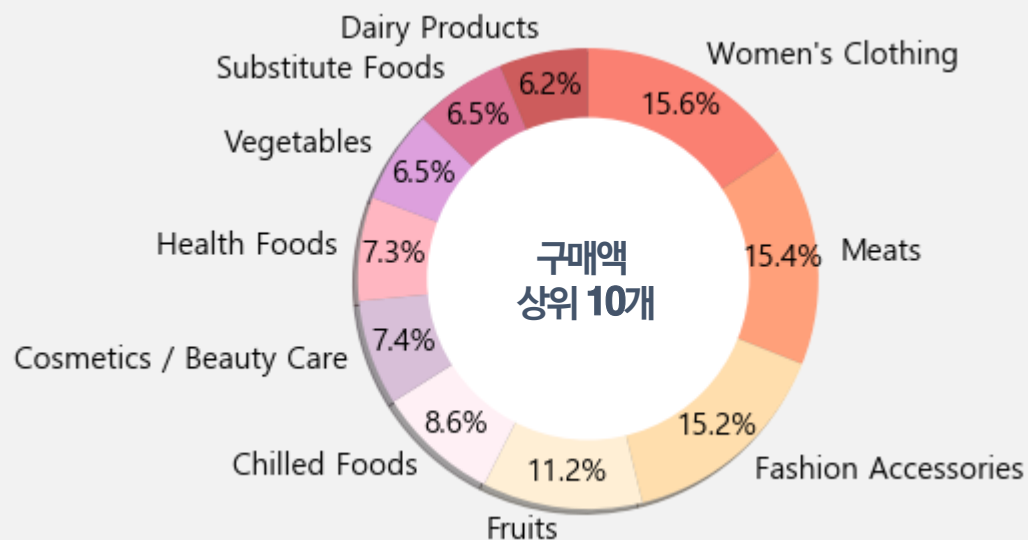
데이터 소개 및 탐색

데이터 탐색

EDA_제품별

총 구매액 : 53억 원

총 구매량 : 57만 건



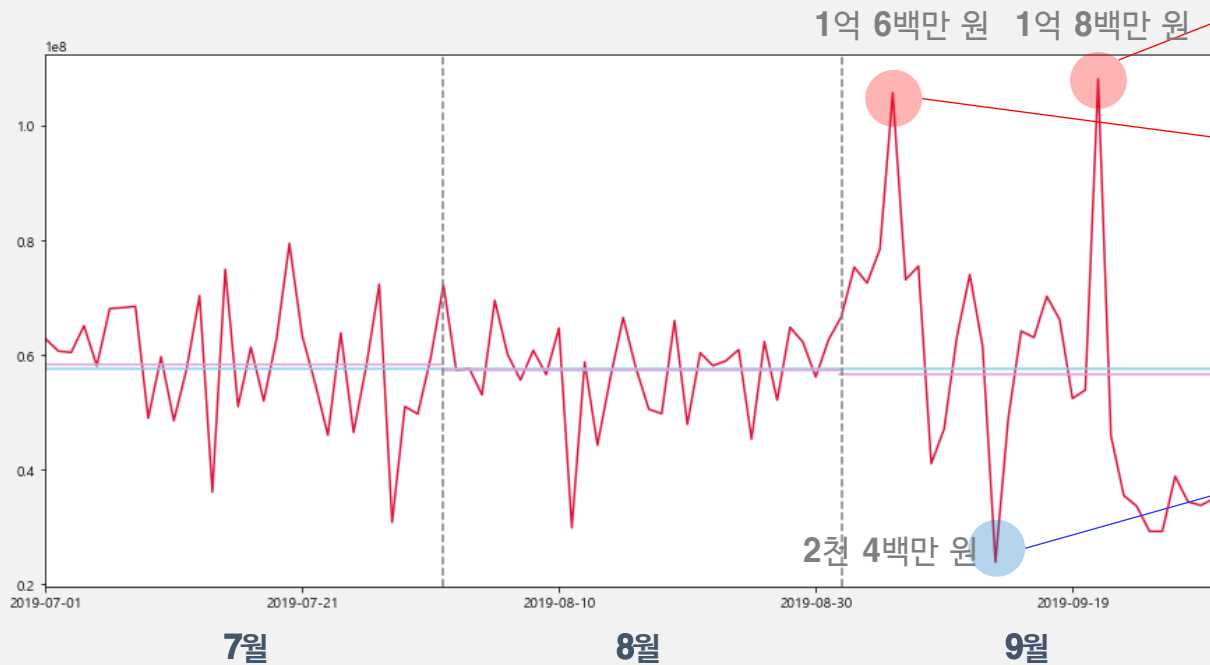
- ✓ 구매액은 여성의류, 먹거리(고기, 과일, 채소), 패션 악세서리, 화장품의 비중이 골고루 있음.
- ✓ 구매량은 음식의 비중이 대부분을 차지함.

데이터 소개 및 탐색

데이터 탐색

EDA 날짜별

7월 평균 구매액 : 5천 8백만 원
8월 평균 구매액 : 5천 7백만 원
9월 평균 구매액 : 5천 6백만 원



✓ 9월 21일
여성의류 최다구매일 (4천 3백만 원)

✓ 9월 5일
커피/티 최다구매일 (2천만 원)
추석 전 선물세트 영향 (7백만 원)

✓ 9월 13일
전체 매출액을 책임지는 여성의류, meat의
판매가 저조함.

데이터 소개 및 탐색

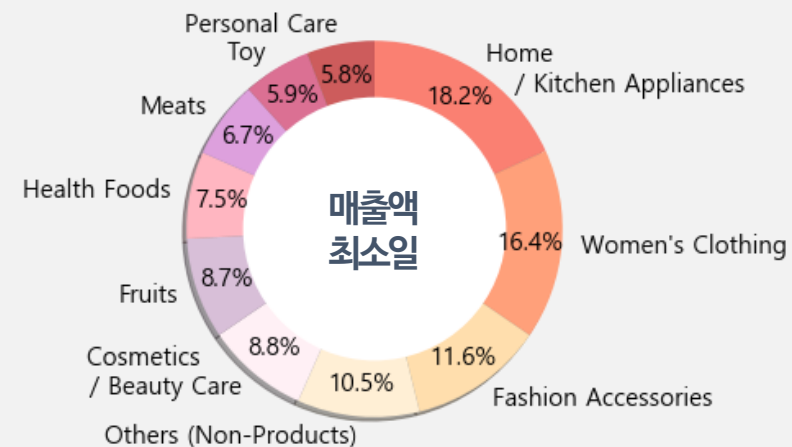
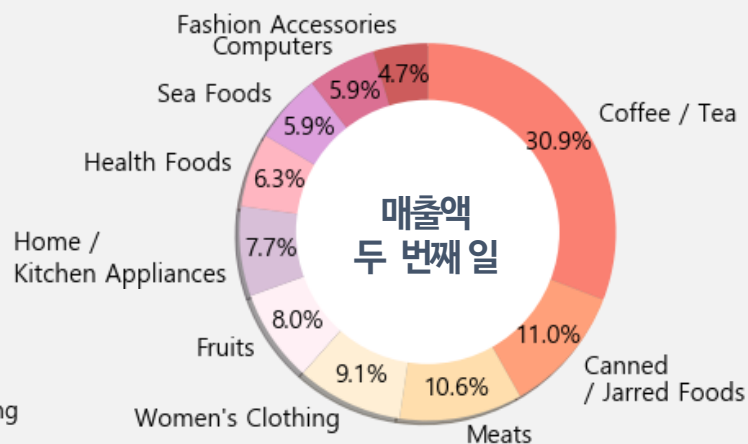
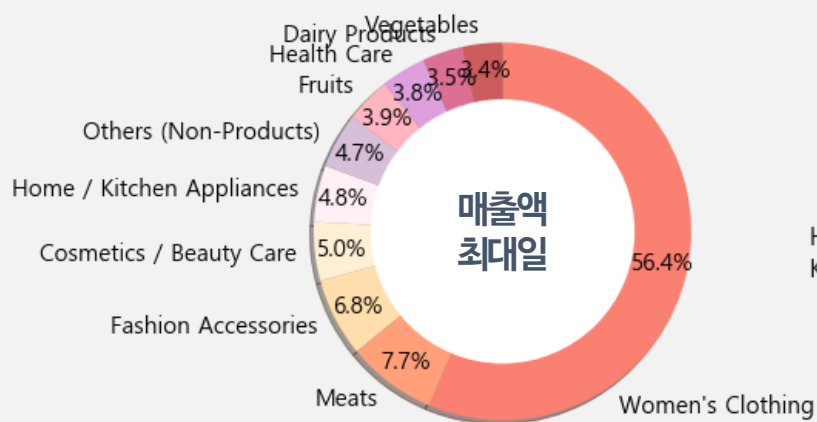
데이터 탐색

EDA 날짜별

7월 평균 구매액 : 5천 8백만 원

8월 평균 구매액 : 5천 7백만 원

9월 평균 구매액 : 5천 6백만 원



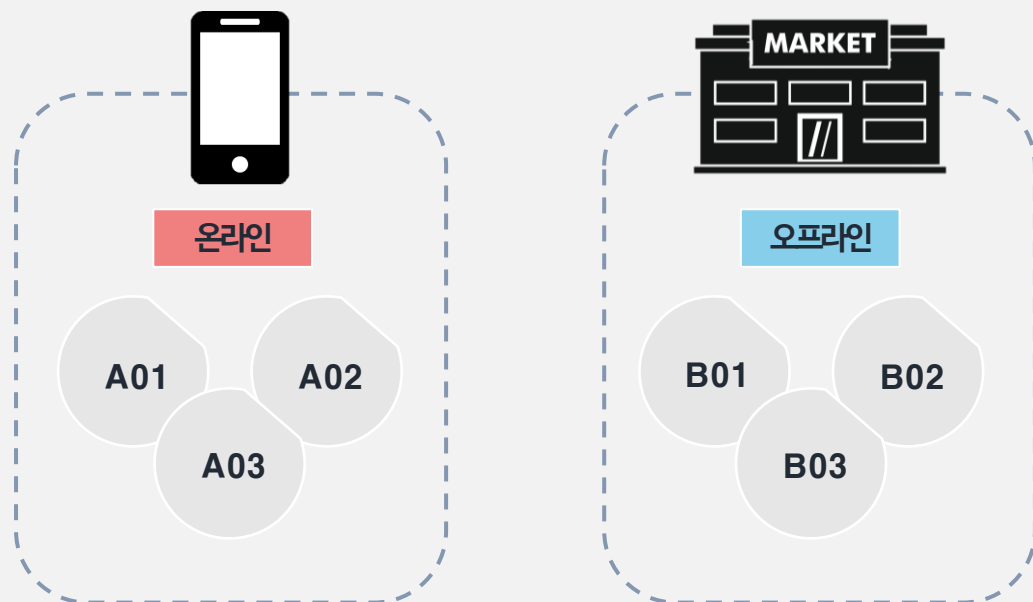
- ✓ 매출액 최대 일에는 **여성의류**가 많이 팔림
- ✓ 매출액 두 번째 일에는 **선물세트(canned/jarred foods)**가 많이 팔렸는데, 이는 추석의 영향이라 볼 수 있음.
- ✓ 매출액 최소일에는 전체 매출액을 책임지는 여성의류와 **meat**의 판매가 저조함.

데이터 소개 및 탐색

데이터 탐색

EDA 온라인/오프라인 이용정보 비교

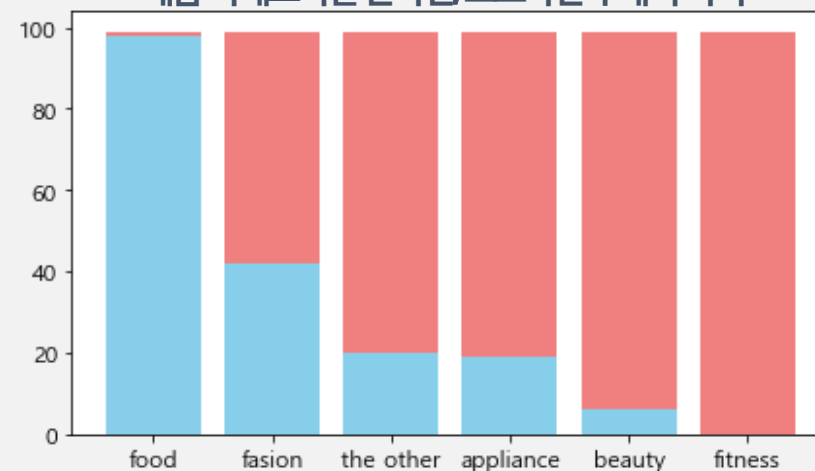
온라인 매장 이용고객 10,026명,
오프라인 매장 이용고객 7,394명을 대상으로



총 구매액 : 33억
총 구매량 : 218,571건

총 구매액 : 20억
총 구매량 : 356,943건

제품 카테고리별 온라인/오프라인 구매액 차이



- ✓ 오프라인에서는 주로 식료품을 많이 구매함.
- ✓ 온라인에서는 식료품을 제외한 옷, 가전/전자제품, 화장품 등을 많이 구매함.
- ✓ 온라인, 오프라인의 제품구매 차이가 뚜렷하게 드러남.
- ✓ 총 구매액은 온라인이 많고, 총 구매량은 오프라인이 많음.

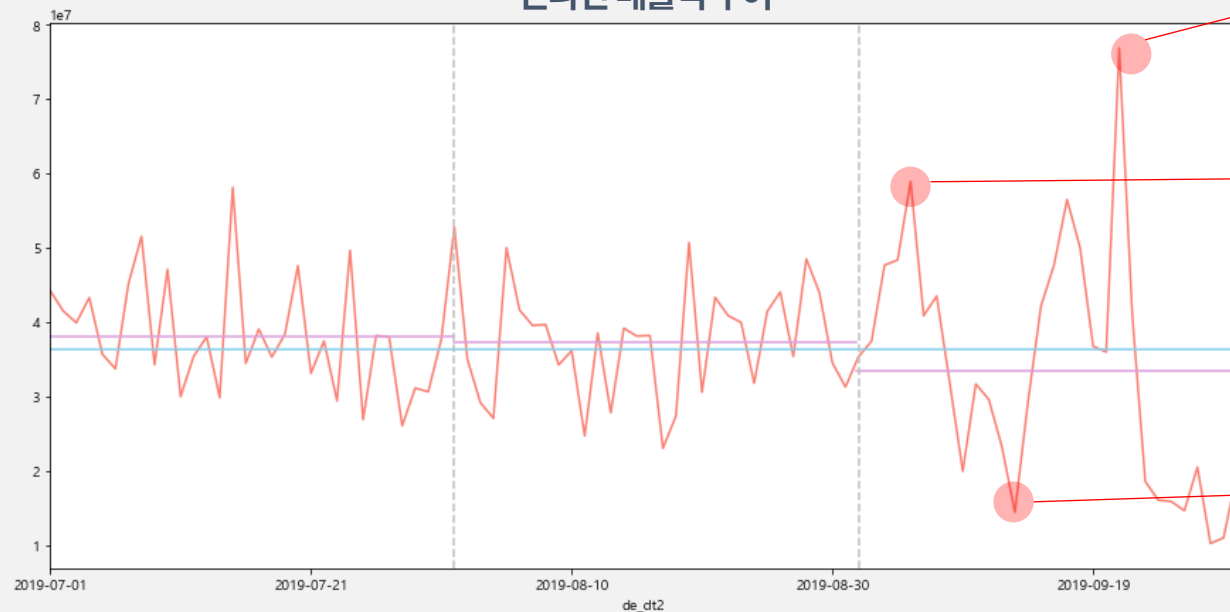
데이터 소개 및 탐색

데이터 탐색

| EDA_온라인 3개월 매출액 추이

온라인 총 매출액 : 33억

온라인 매출액 추이



9월 21일
여성의류 판매액 최대(약 4천만원)

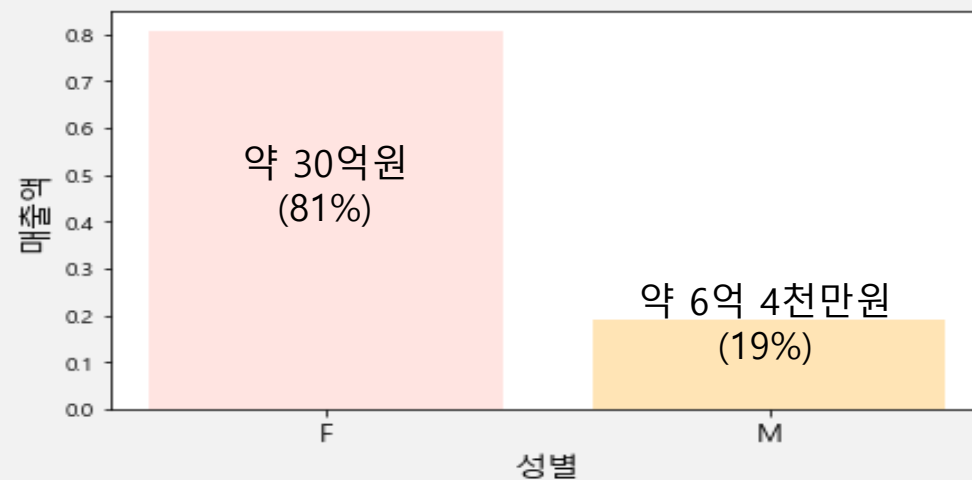
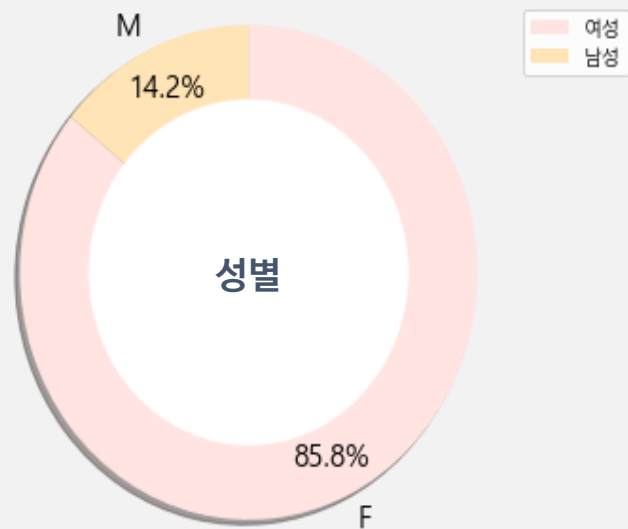
9월 5일
고기, 과일, 해산물 등 판매액 최대
(추석 영향)

9월 13일
추석 연휴 (매출 최하위)

데이터 소개 및 탐색

데이터 탐색

| EDA 온라인 성별 비율 및 매출액



데이터 소개 및 탐색

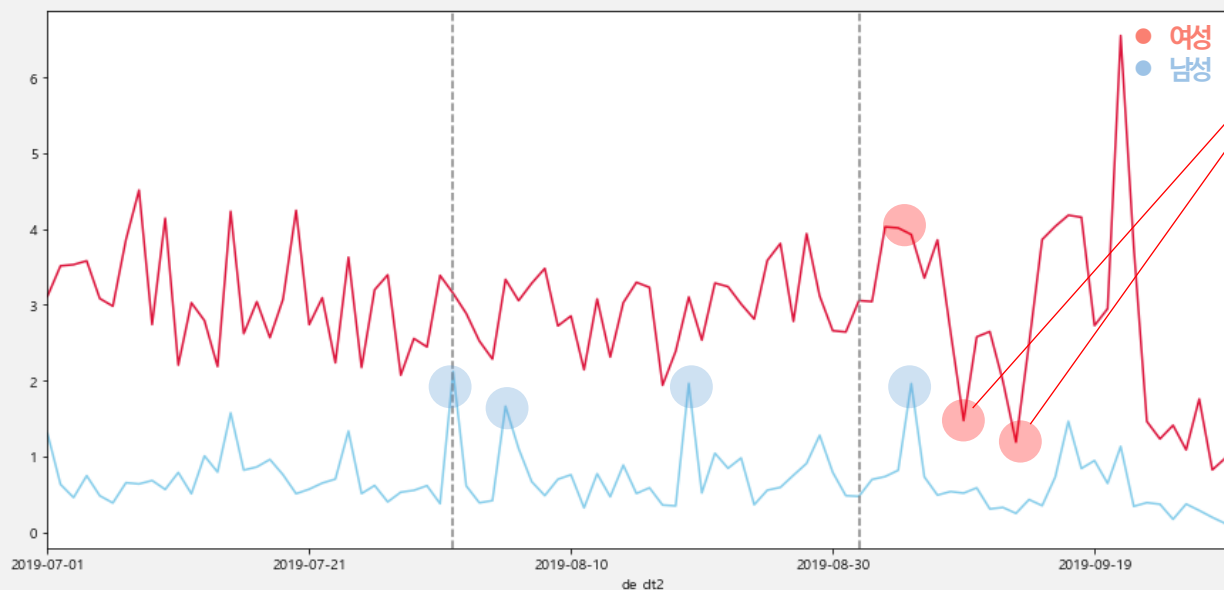
데이터 탐색

| EDA 온라인 성별 3개월 구매액 추이

여성 1인당 평균 구매액 : 32 만원

남성 1인당 평균 구매액 : 45 만원

성별 구매액 추이



✓ 9월 13일(추석)
여성 매출액 최하위

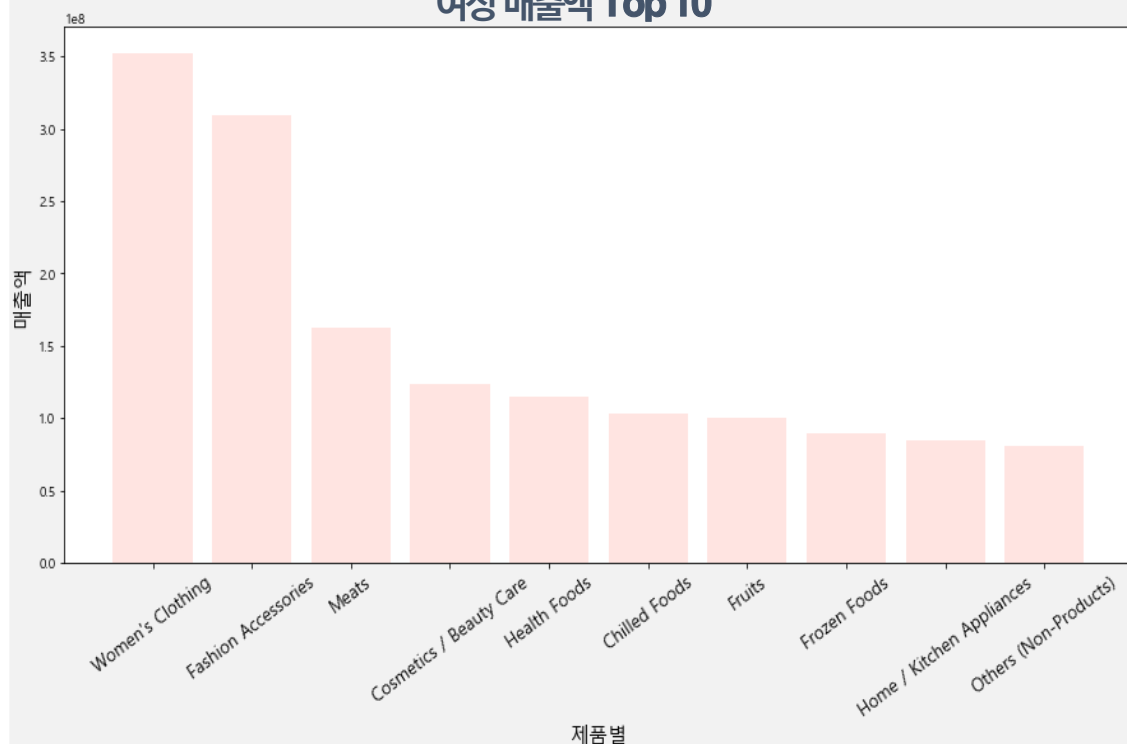
- ✓ 9월 여성 매출액 변동성이 높음
- ✓ 여성은 연휴에 영향을 많이 받는 반면 남성은 연휴에 영향을 덜 받음
- ✓ 남성이 매출액이 높을 때 주로 가전제품을 많이 삼
- ✓ 여성 총 매출액이 남성보다 압도적이지만 1인당 평균 구매액은 남성이 여성보다 높음

데이터 소개 및 탐색

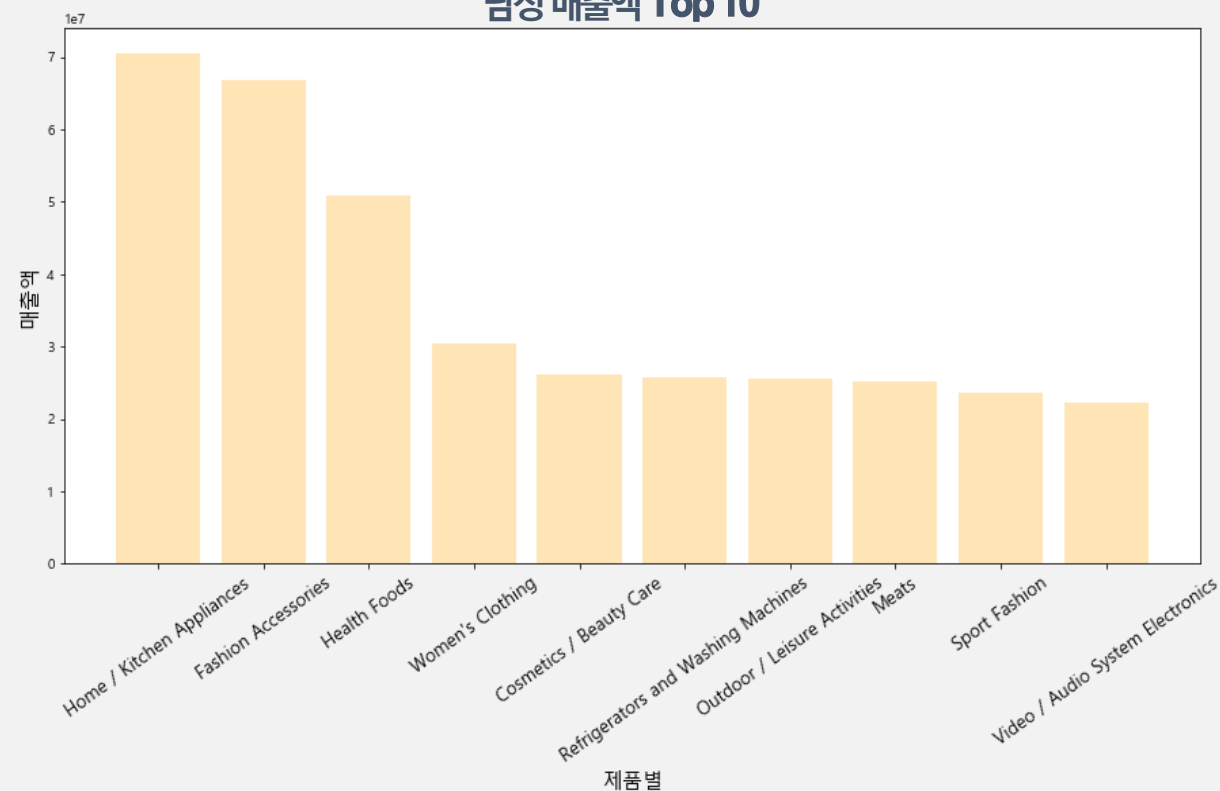
데이터 탐색

| EDA 온라인 성별 매출액 Top10

여성 매출액 Top 10



남성 매출액 Top 10



- ✓ 여성의 경우 의류와 악세서리 매출액이 압도적이며 **의류/뷰티/식품** 쪽 소비가 큰 것으로 보임
- ✓ 남성의 경우 식품의 비중은 적으며 **가전/가구**의 매출액이 상당히 높음

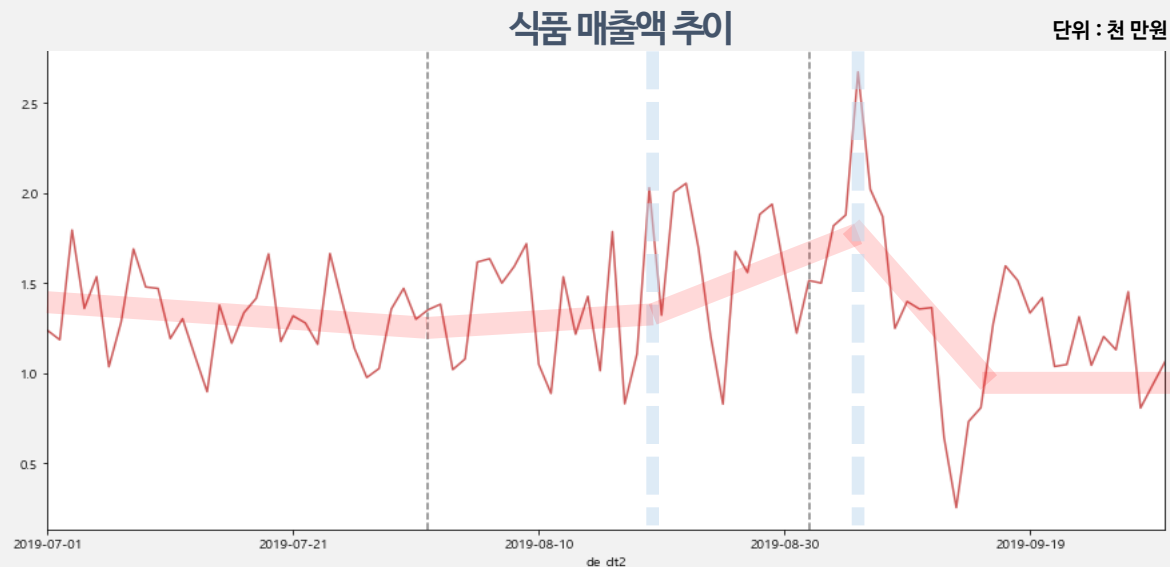
데이터 소개 및 탐색

데이터 탐색

| EDA_온라인 제품별 3개월 매출액 추이

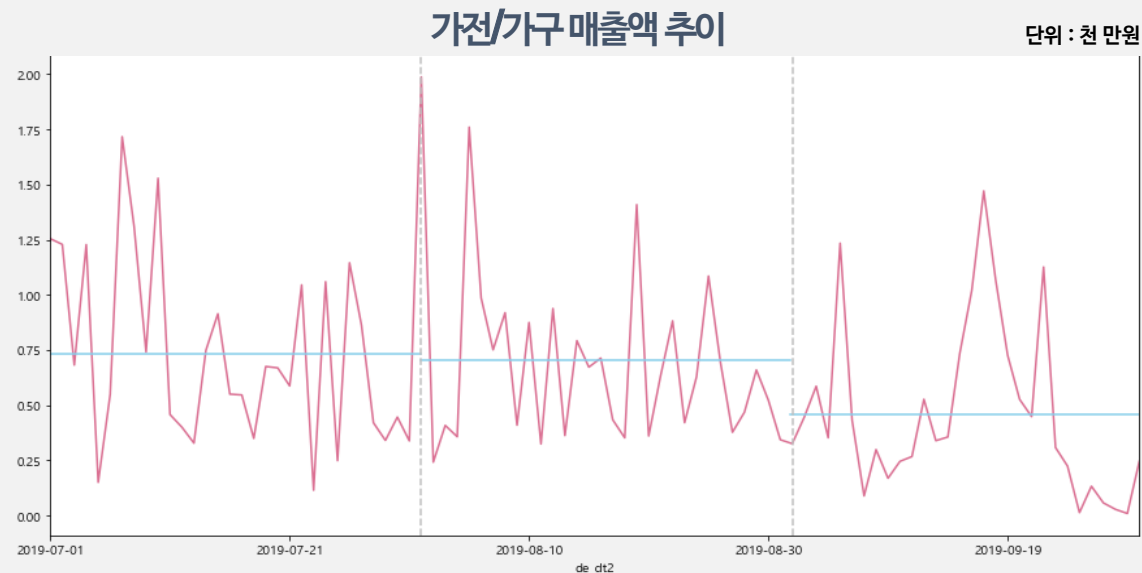
식품 일 평균 매출액 : 1천 4백 만원

가전/가구 일 평균 매출액 : 6백 30 만원



✓ 8월19일부터 9월 5일까지의 식료품 매출액 증가

✓ 추석의 영향으로 보임



✓ 7, 8월 가전/가구 매출액이 높음

✓ 남성의 영향을 많이 받는 것으로 보임

데이터 소개 및 탐색

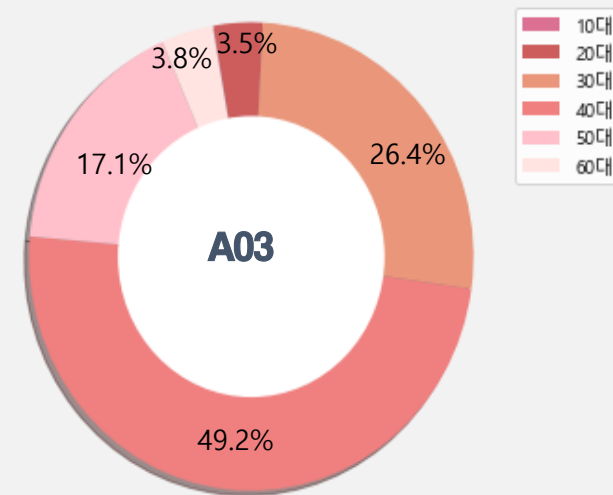
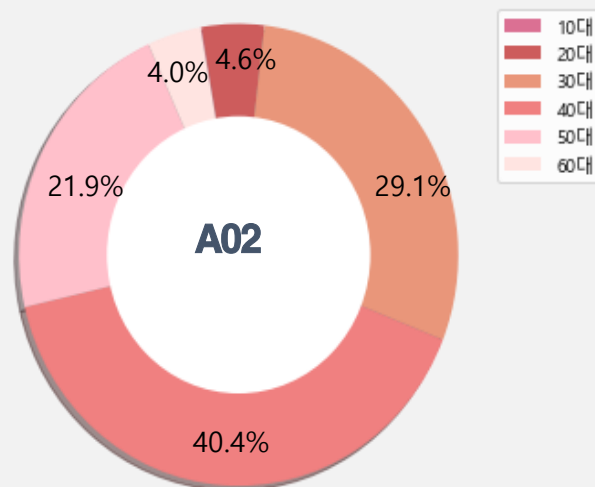
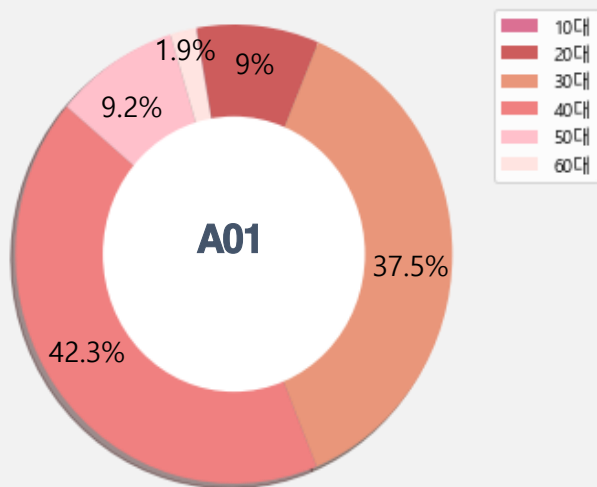
데이터 탐색

EDA 온라인 거래처 별 나이대 매출액

A01 일 평균 매출액 : 3백 만원

A02 일 평균 매출액 : 2천 4백 만원

A03 일 평균 매출액 : 1천 2백 만원

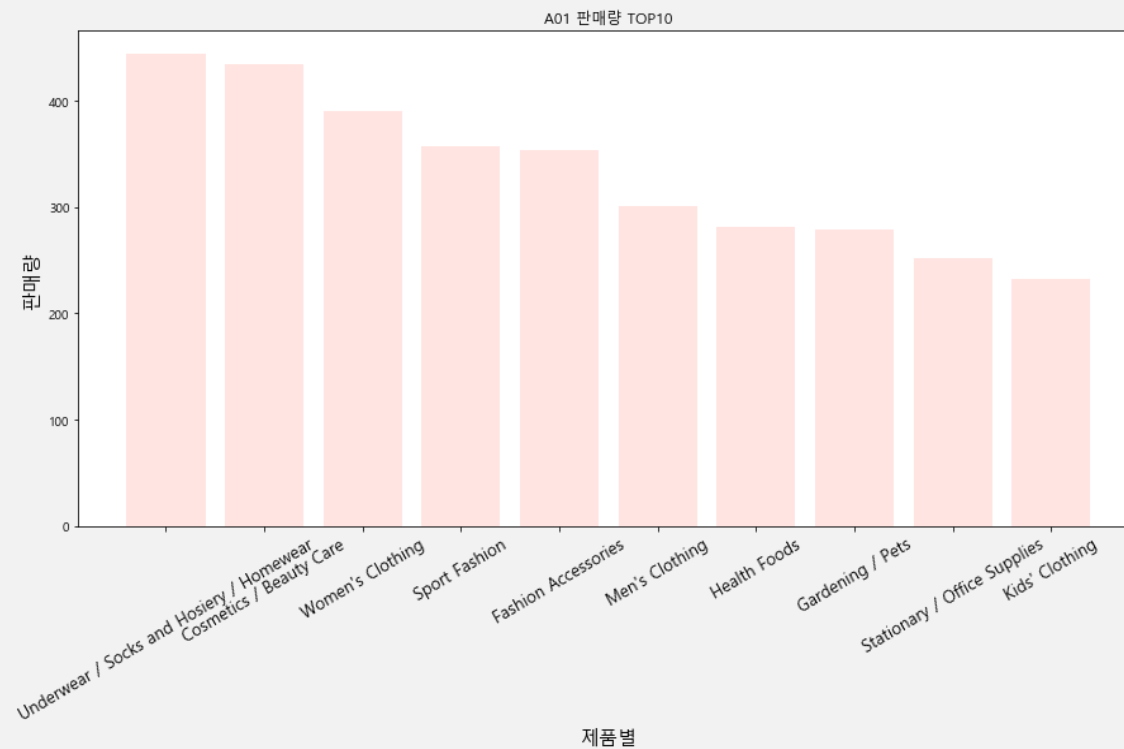
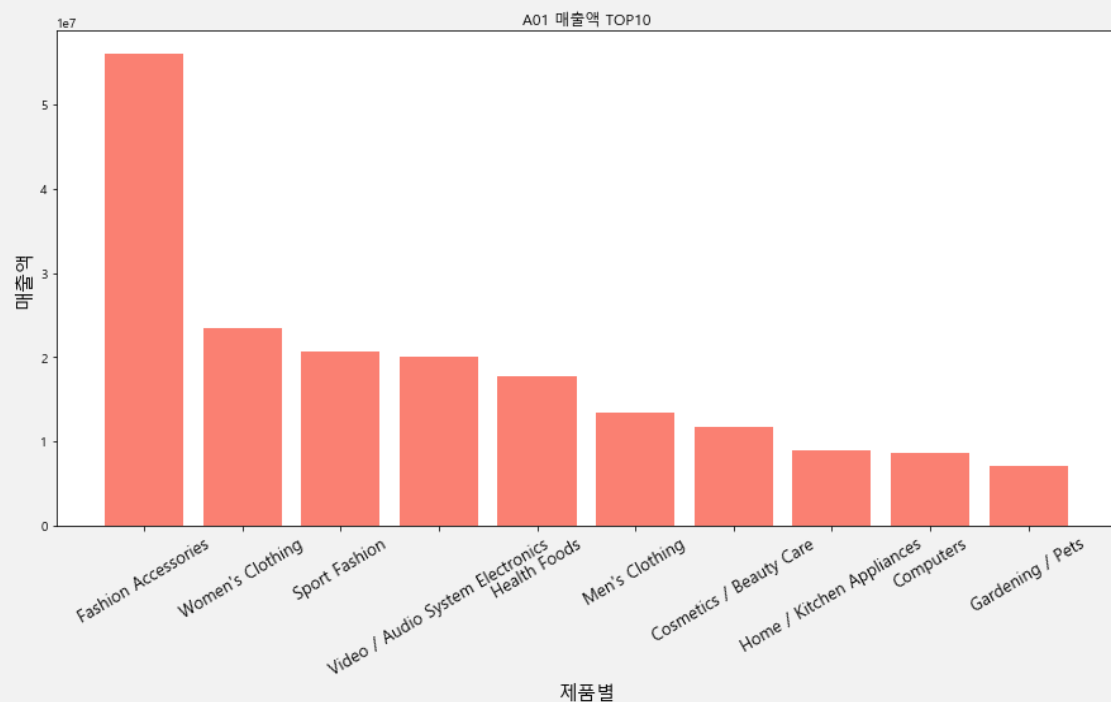


- ✓ 10대는 모두 0.1% 미만
- ✓ 50대에서 A02 매출액이 상당히 높음
- ✓ 20대, 30대에서 A01 매출액이 상당히 높음
- ✓ 40대에서 A03 매출액이 상당히 높음

데이터 소개 및 탐색

데이터 탐색

| EDA_온라인 A01 제품별 매출액 Top 10

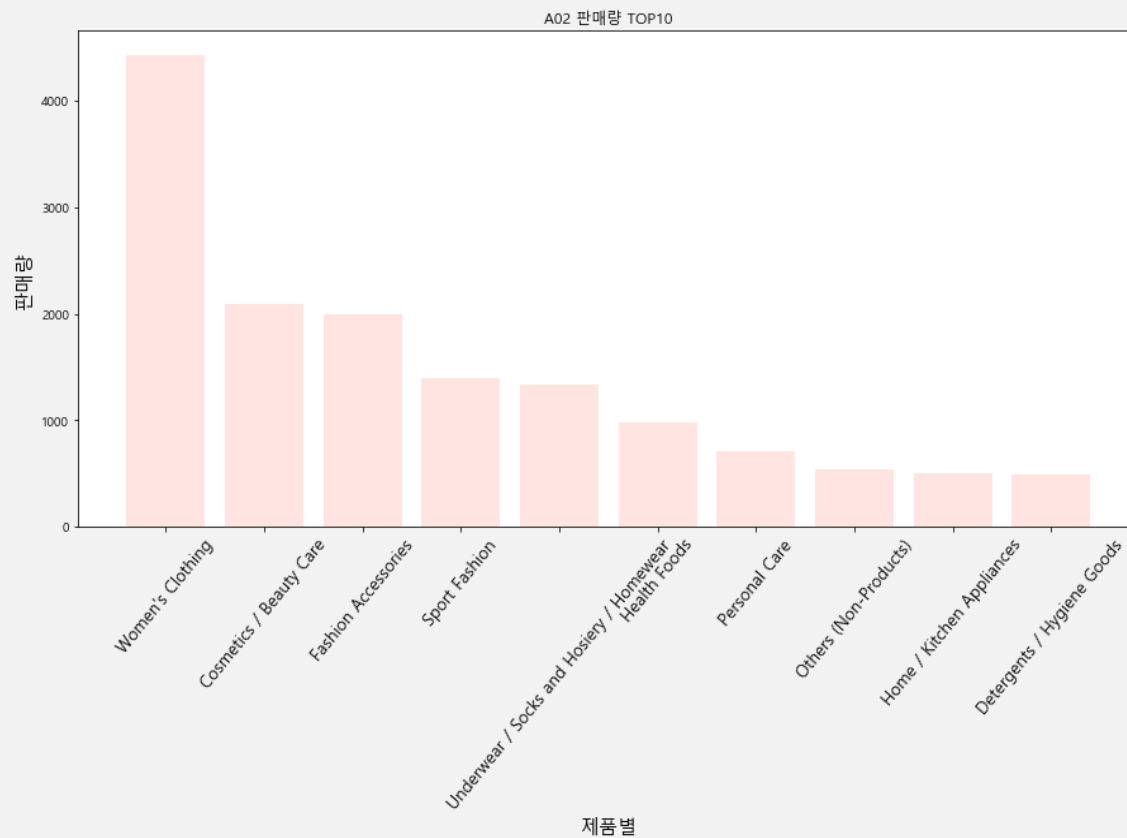
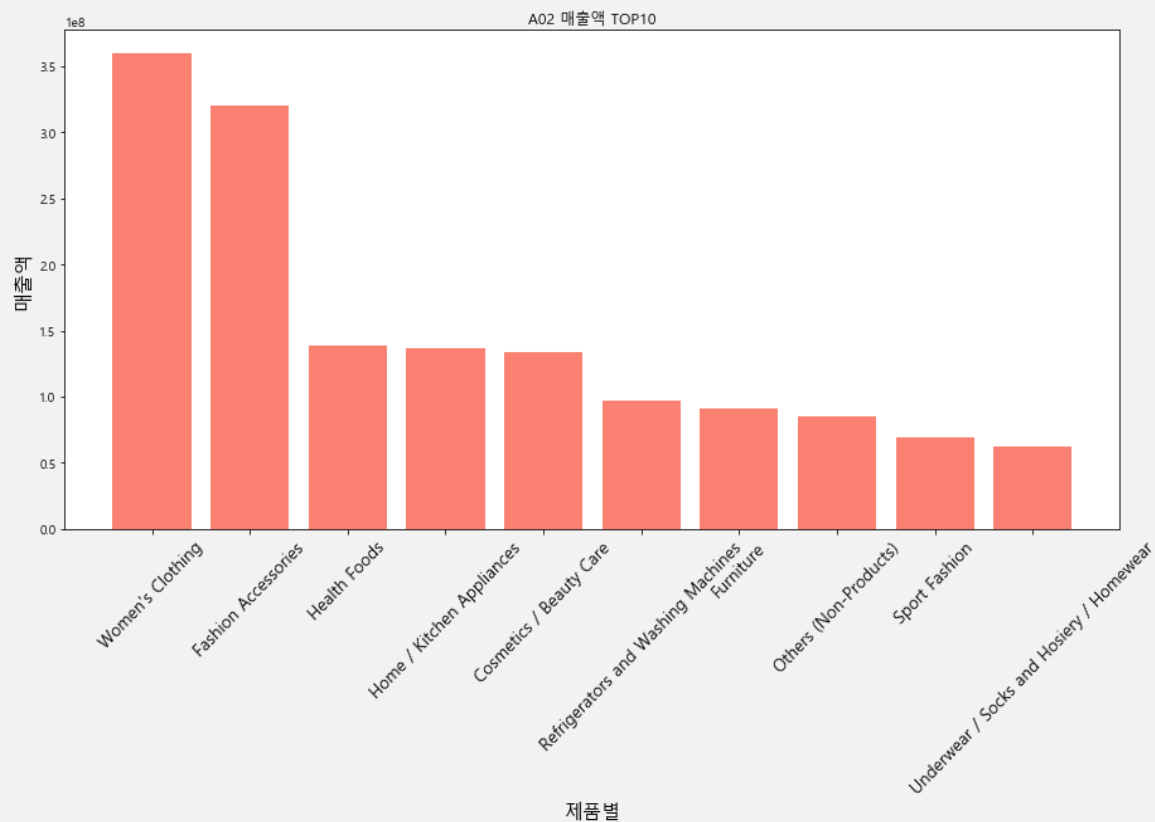


- ✓ 패션 악세서리의 매출액이 압도적으로 높음
- ✓ 뷰티/패션과의 연관성이 높아 보임

데이터 소개 및 탐색

데이터 탐색

| EDA_온라인 A02 제품별 매출액 Top 10

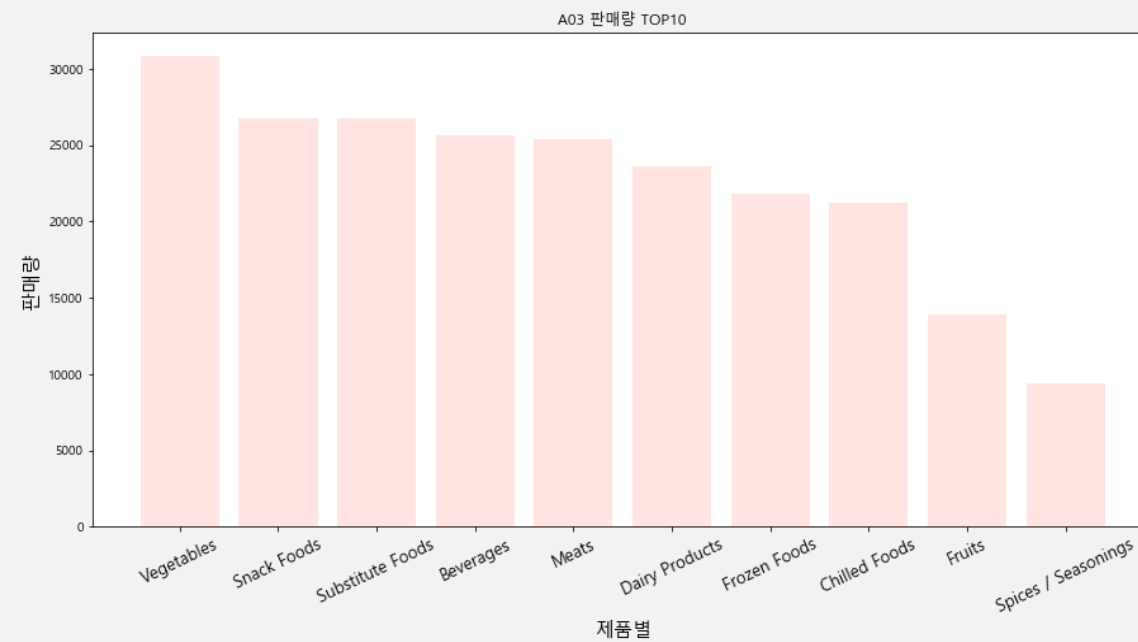
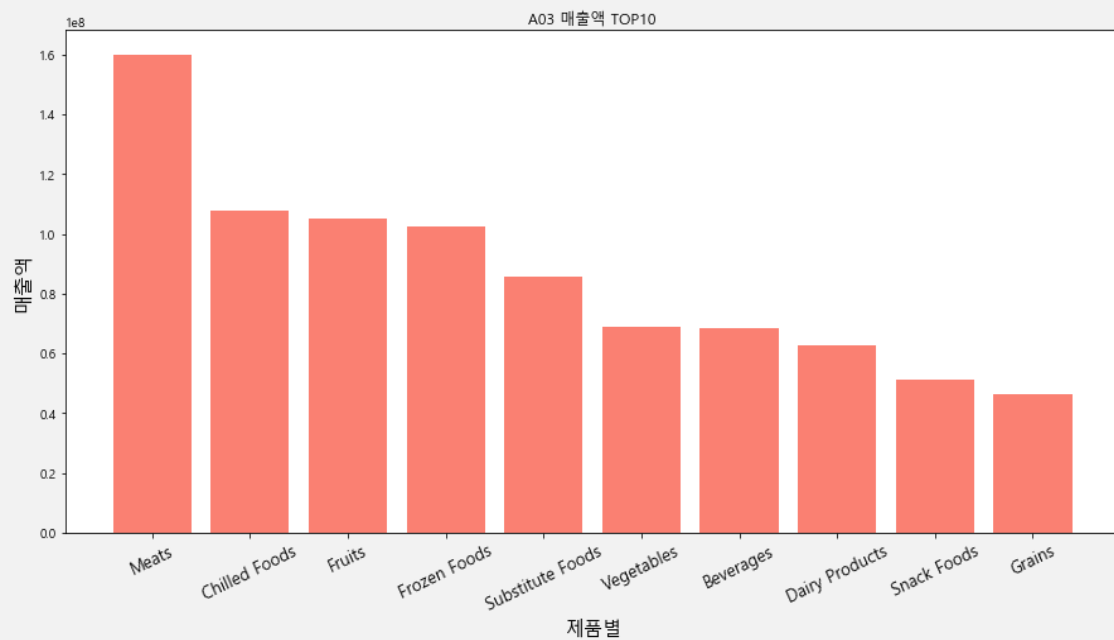


- ✓ 패션 악세서리와 여성 의류의 매출액이 압도적으로 높음
- ✓ 여성의류가 대부분의 판매량을 차지하고 있음

데이터 소개 및 탐색

데이터 탐색

| EDA_온라인 A03 제품별 매출액 Top 10



✓ 대부분 음식에서 판매량과 매출액이 높다.

데이터 소개 및 탐색

데이터 탐색

| EDA_온라인 거래처 별 특징

A01



주로 패션/뷰티 쪽에 매출액이 높음
→ 20대, 30대

A02



여성의류가 매출액/판매량 모두 압도적
→ 50대

A03



대부분 음식에서 매출액이 압도적
→ 40대

데이터 소개 및 탐색

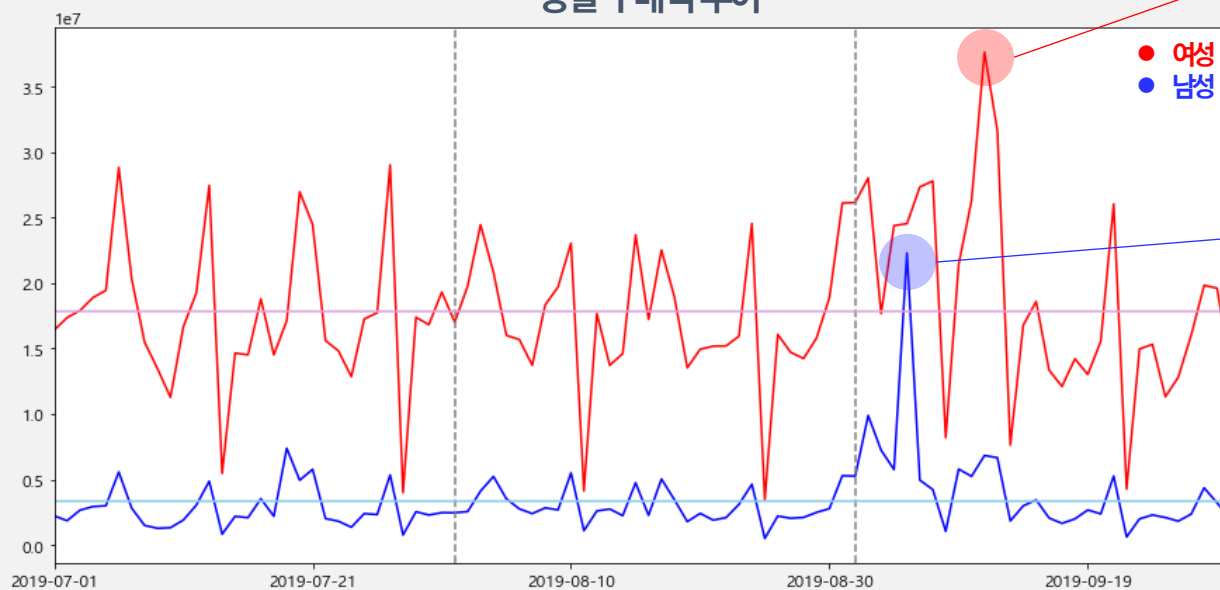
데이터 탐색

| EDA_오프라인 성별

여성 평균 구매액 : 1천 8백만 원

남성 평균 구매액 : 3백 36만 원

성별 구매액 추이



✓ 9월 11일
여성 최대구매액 (3천 7백만 원)
각종 선물세트 매출이 큼.

✓ 9월 5일
남성 최대구매액 (2천만 원)
선물세트와 여성의류 매출이 큼.

✓ 여성이 남성보다 평균적으로 더 많이 구매함.

✓ 매출액 최대 일에는 여성의류와 선물세트를 많이 구매함. 이는 추석의 영향이라고 볼 수 있음.

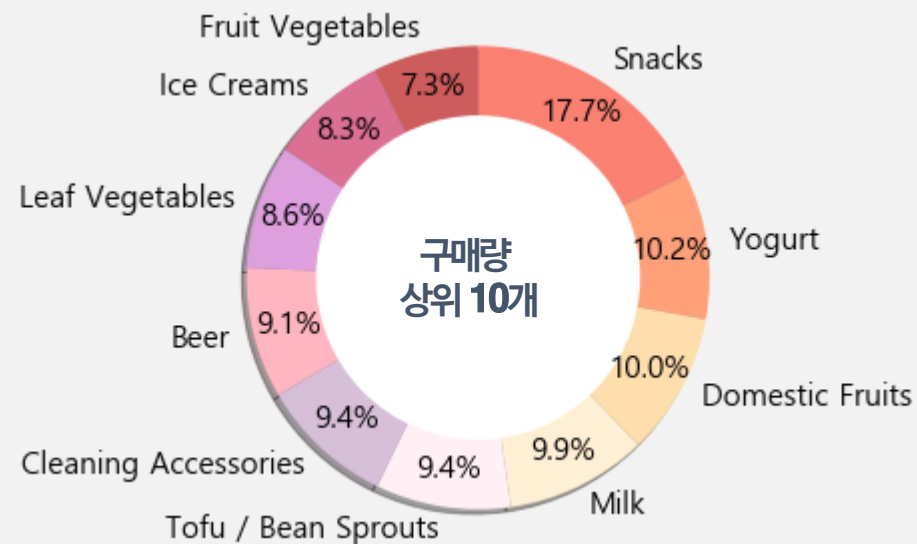
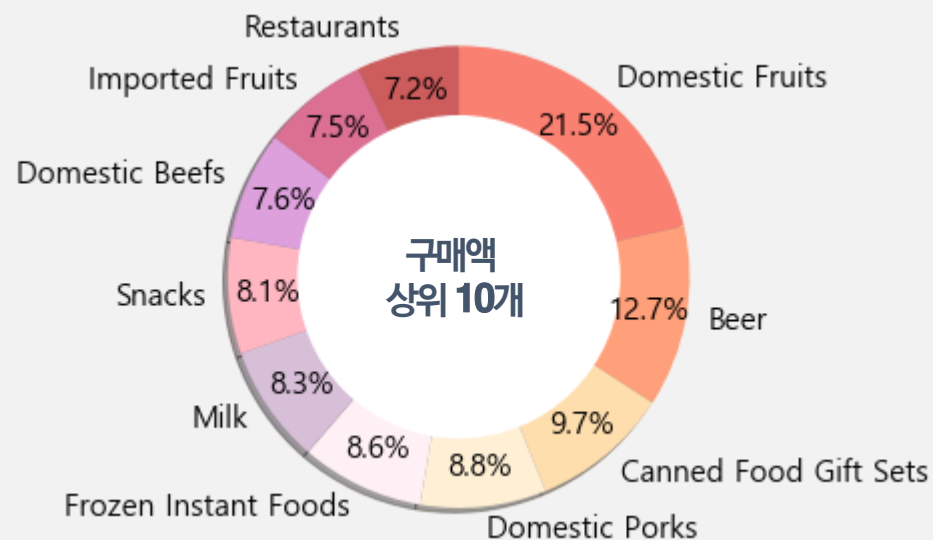
데이터 소개 및 탐색

데이터 탐색

EDA 오프라인 제품별

총 구매액 : 20억

총 구매량 : 356,943건



- ✓ 오프라인에서는 온라인 구매특성과는 달리 주로 **가정에서 먹는 식품**을 많이 구매함을 알 수 있다.
- ✓ 특히 **7~9월**의 시기에 수집된 데이터 특성에 따라 '선물세트'의 매출액 비중이 크게 자리잡고 있다.

데이터 소개 및 탐색

데이터 탐색

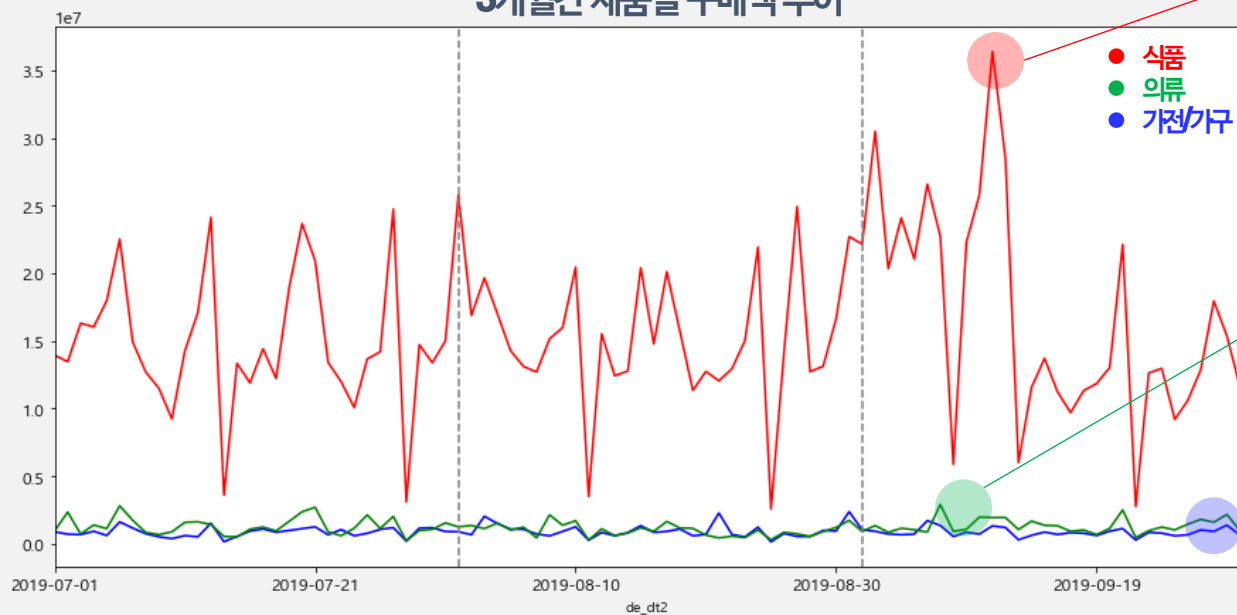
EDA_오프라인 날짜별

식품 평균 구매액 : 1천 5백만 원

의류 평균 구매액 : 1백 20만 원

가전/가구 평균 구매액 : 90만 원

3개월간 제품별 구매액 추이



✓ 식품
오프라인 제품 카테고리 중 가장 매출액이 높음.
9월에 매출액 크게 증가.
추석 시즌에 매출액이 증가한 것으로 보임.

✓ 의류
7월에 남성의류 매출액이 특히 많았음.
8월 매출이 전체적으로 저조함.

✓ 가전/가구
8월 매출액이 가장 높았음.
7, 8월 여름철 에어컨 구매가 증가함.

데이터 소개 및 탐색

데이터 탐색

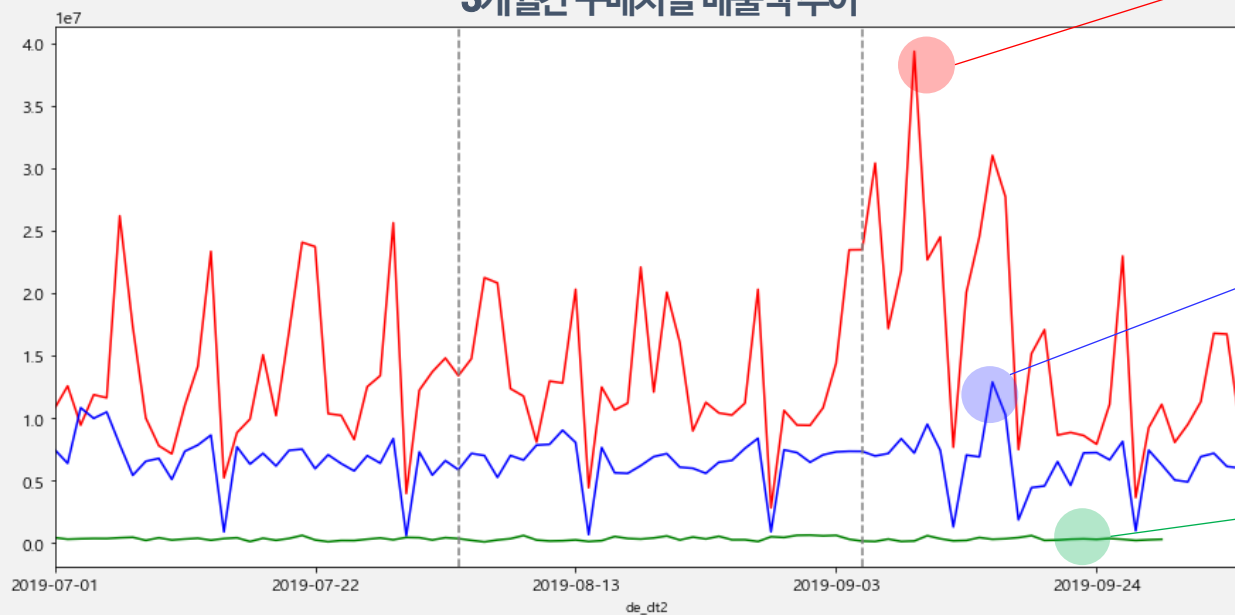
| EDA_오프라인구매처별

B01 일 평균 매출액 : 1천 4백 60만 원

B02 일 평균 매출액 : 6백 58만 원

B03 일 평균 매출액 : 32만 원

3개월간 구매처별 매출액 추이



✓ B01
식품 판매 매장
구매처 중 가장 많은 매출액을 가짐
9월에 매출 크게 증가(추석 선물세트)

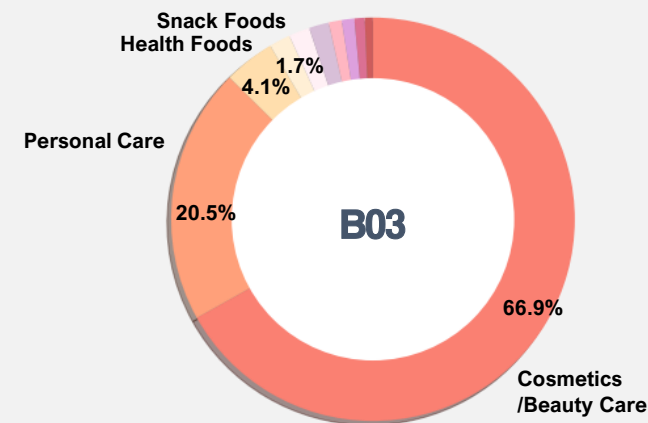
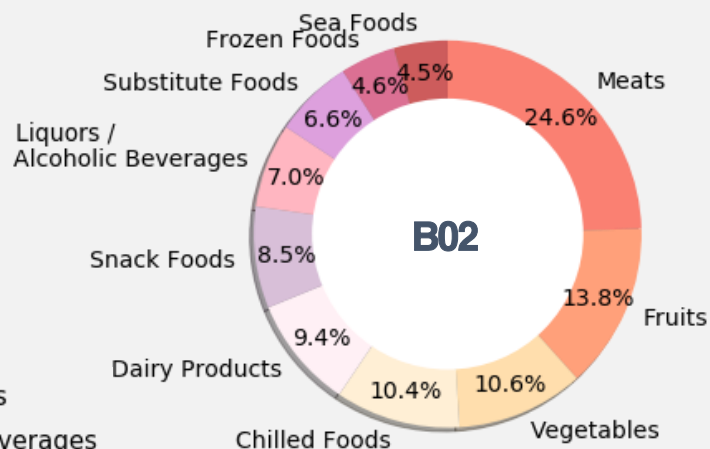
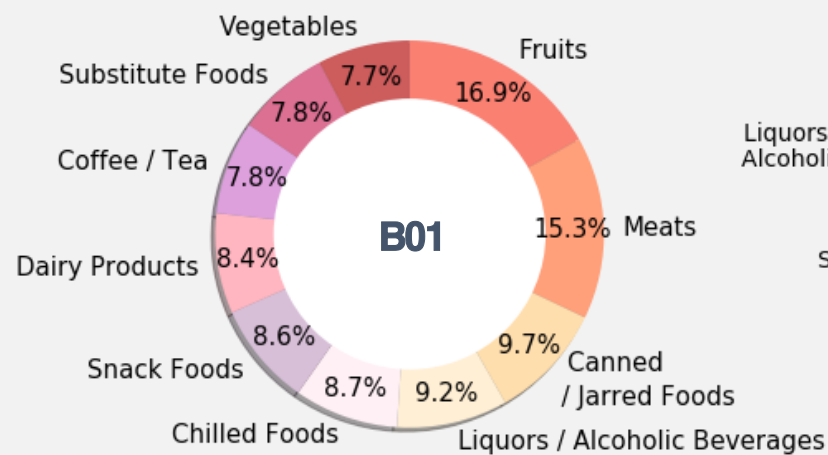
✓ B02
식품 판매 매장
월간 일정한 매출액 유지

✓ B03
화장품 매출액이 많은 매장
구매처 중 가장 낮은 매출액을 가짐

데이터 소개 및 탐색

데이터 탐색

| EDA_오프라인구매처별



- ✓ B01와 B02는 주로 식료품을 판매하는 매장이다.
- ✓ B01은 월간 매출 비슷하고 선물세트가 많이 팔리며 9월 매출이 크게 증가한 반면, B02는 선물세트 판매에 영향이 없고 매 월 일정한 매출액을 유지하고 있다.
- ✓ B03은 화장품 판매액이 많은 매장이다. 구매처 중 가장 낮은 매출액을 가진다.

군집분석

| 분석 흐름도

RFM변수 생성

K-Means Clustering

군집 별 특징 확인

- 군집분석을 위한 변수로 RFM을 정의함
- **Recency** = 2019년 10월 1일 기준, 가장 최근 구매일과의 차이(일)
- **Frequency** = 3개월 간 총 방문횟수(회)
- **Monetary** = 3개월 간 총 구매액(원)

RFM

표준화한 RFM

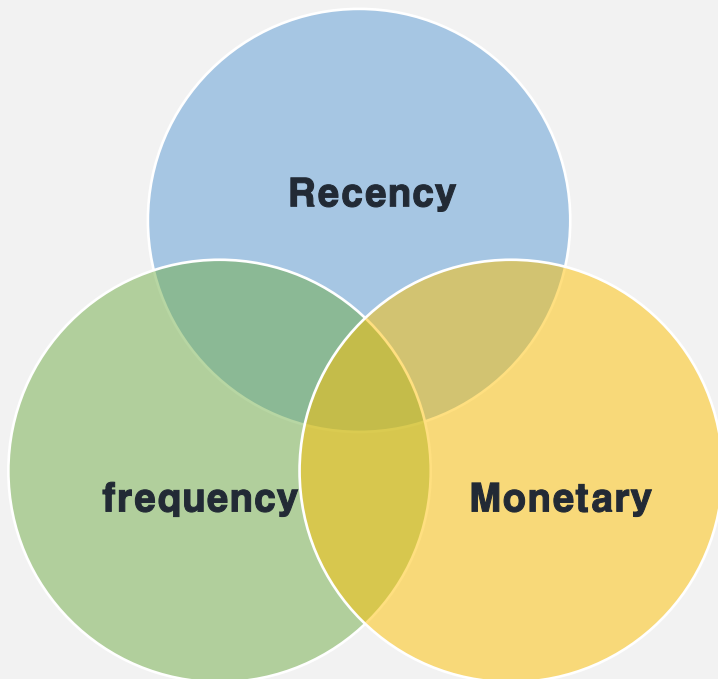
clnt_id	manetary	frequency	recency3	z_frequency	z_manetary	z_recency
9	239870	5	17	-0.301769	-0.050582	-0.256380
12	29900	1	38	-0.739296	-0.486327	0.628367
20	18900	1	89	-0.739296	-0.509155	2.777038
23	175964	8	3	0.026376	-0.183204	-0.846212
24	2300	1	60	-0.739296	-0.543605	1.555245

- 사전에 지정한 군집의 수를 사용하여 고객을 각각의 군집에 할당함
- R/F/M변수를 각각 표준화한 후 K-means Cluster 생성
- 최적의 군집 수 찾기

- 온라인, 오프라인 고객별 각각의 군집을 생성하여 특징 파악

군집분석

| 변수



✓ 구매 최근성 **Recency**

고객의 마지막 구매 시점이 언제인지를 나타내는 변수로써 산업에 따라 다소 차이가 있지만 일반적으로 최근에 구매한 고객일수록 현재의 관계가 유의하다고 판단할 수 있음.



✓ 구매 빈도 **Frequency**

정해진 기간 동안 얼마나 자주 구매했는지를 나타내는 변수로써 동일한 기간 동안 구매횟수가 많을수록 높은 점수를 부과하며, 고객의 구매/이용활동성 판단이 가능



✓ 구매 금액 **Monetary**

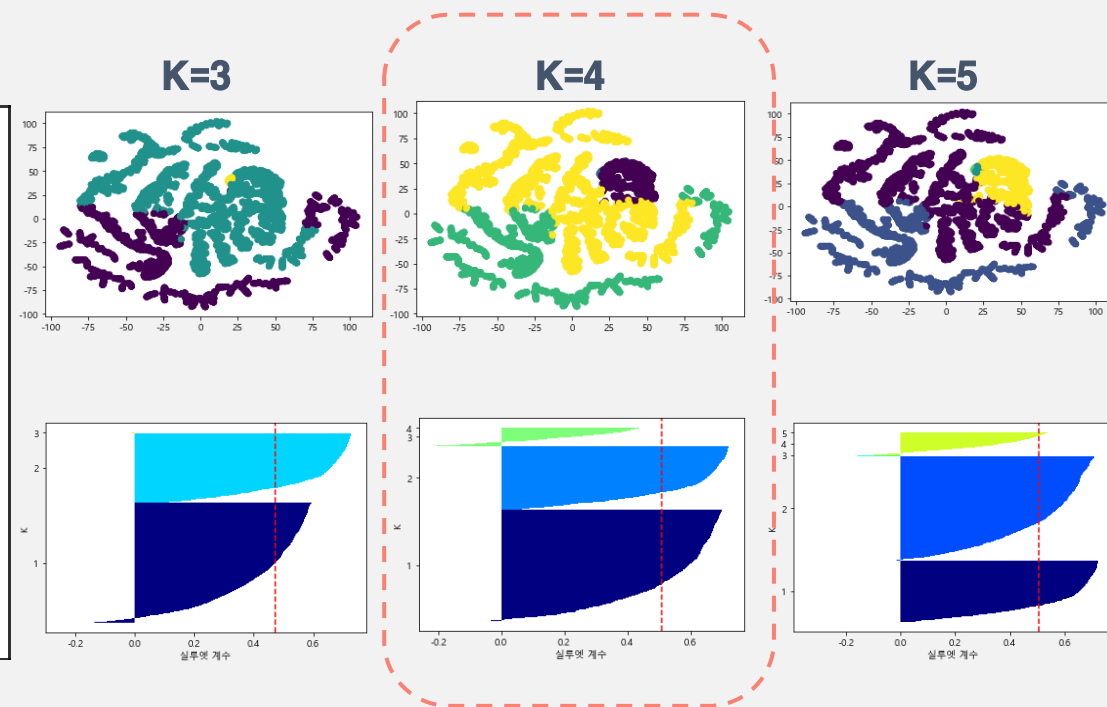
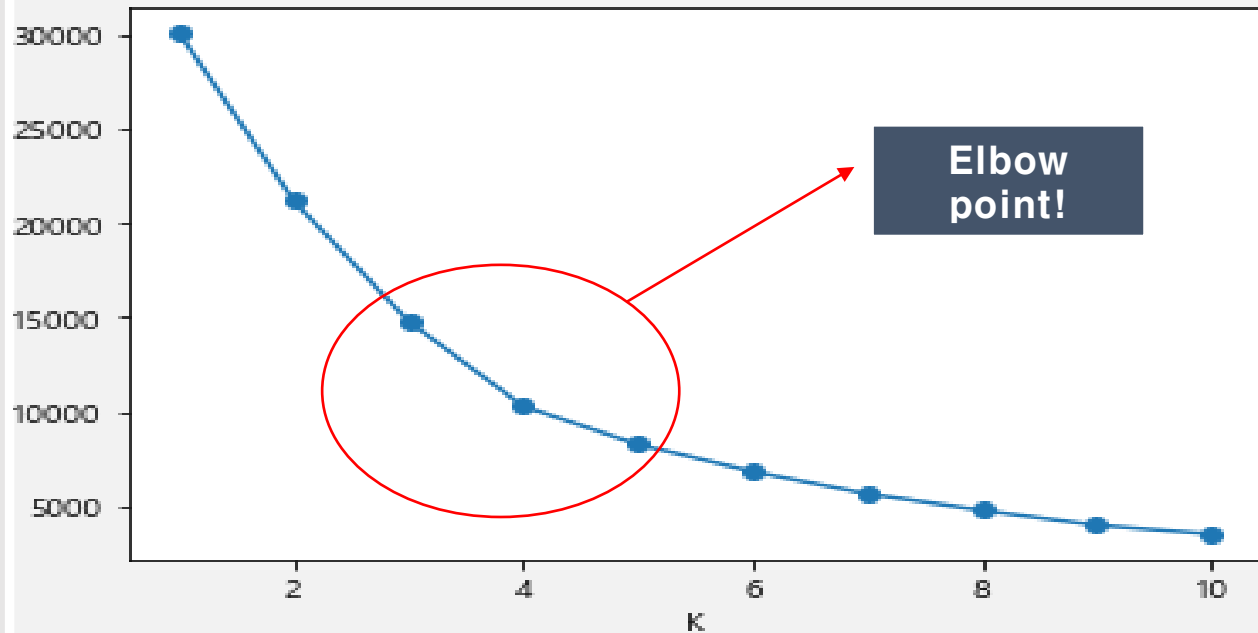
일정 기간 동안에 고객의 총 구매금액을 나타내는 변수로써, 구매액이 높을수록 높은 점수를 획득할 수 있다. 지나치게 높은 구매액이 존재할 경우 상한선을 두는 것이 RFM지수 왜곡을 방지함.

군집분석

온라인 데이터

| 최적의 군집 수 찾기

군집 수에 따른 오차제곱합



- ✓ Elbow method를 통해 K가 3~5사이에 최적 군집을 형성함을 알 수 있다.
- ✓ t-SNE와 실루엣 분석을 통해 K가 4일 때 군집이 보다 잘 형성됨을 볼 수 있다.

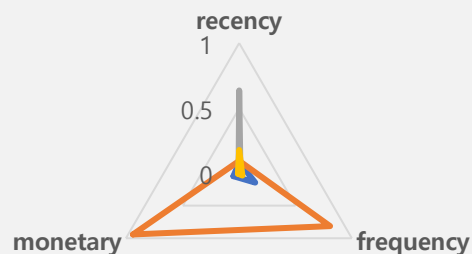
군집분석

온라인 데이터

I 군집 별 특징

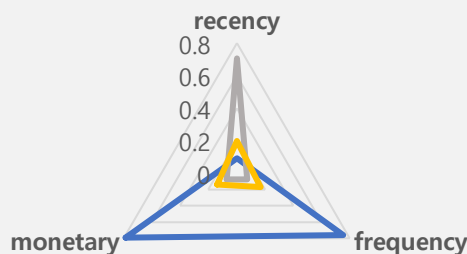
군집 1 : 924명
 군집 2 : 6명
 군집 3 : 3329명
 군집 4 : 5767명

— 군집1 — 군집2 — 군집3 — 군집4



이상 군집 (군집2)
제거 후

— 군집1 — 군집3 — 군집4



✓ 군집3, 군집 4에 마케팅 전략 필요

군집1

- ✓ 방문 시점이 최근, 방문 빈도 높음. 구매 액수가 큰 군집
- ✓ VIP 고객들

군집2

- ✓ 방문 시점이 최근, 방문 빈도는 매우 높음, 구매액수 매우 큼
- ✓ 이상 군집

군집3

- ✓ 평균적으로 2달 전 방문, 구매빈도, 구매 액수 모두 매우 낮음
- ✓ 이탈 고객으로써 마케팅 전략 필요

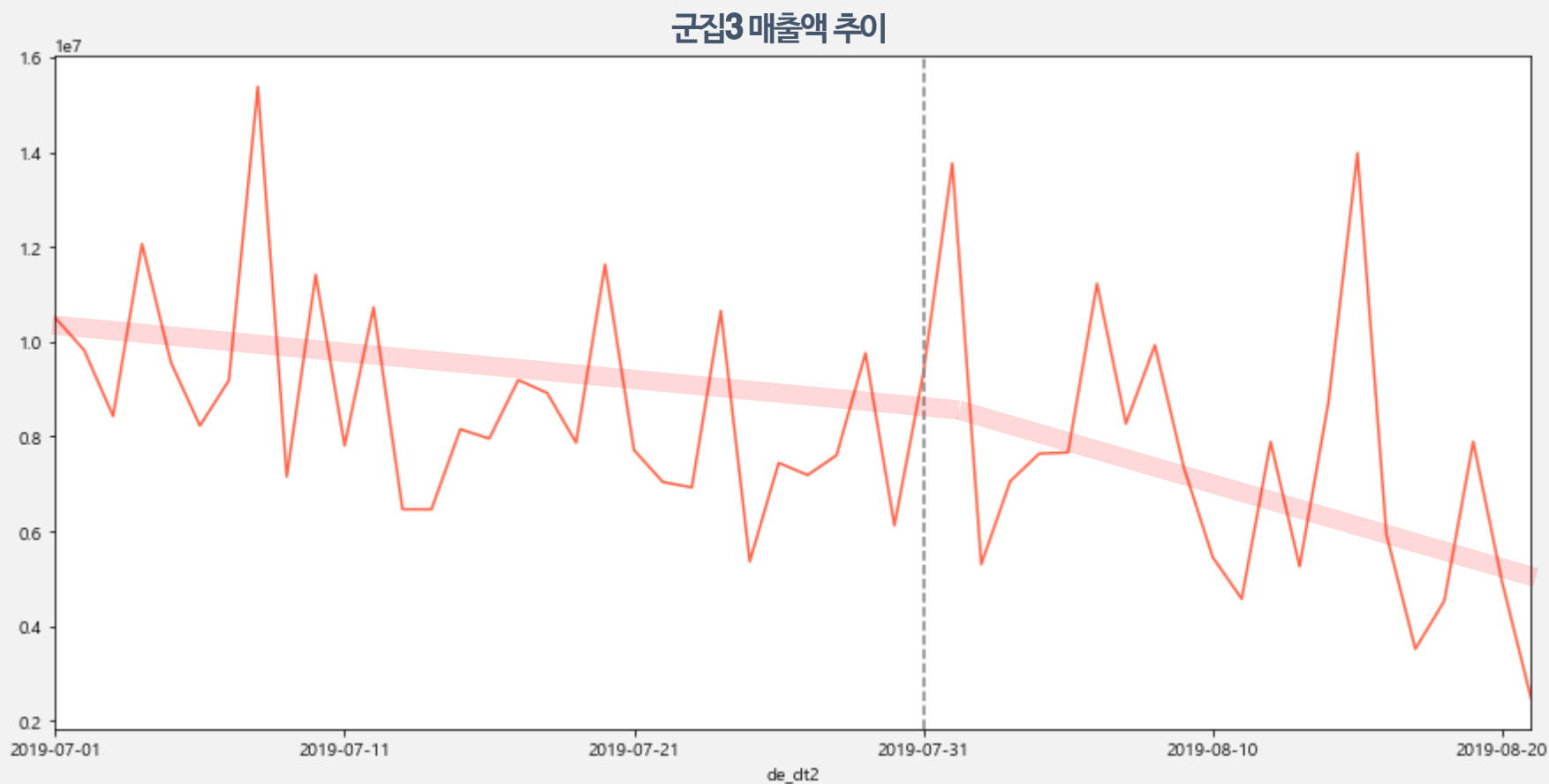
군집4

- ✓ 방문 시점, 구매빈도, 구매액수 모두 중간
- ✓ 평범한 고객
- 마케팅 전략을 통해 단골고객으로 만들 필요가 있음

군집분석

온라인 데이터

| 군집3 매출액 추이



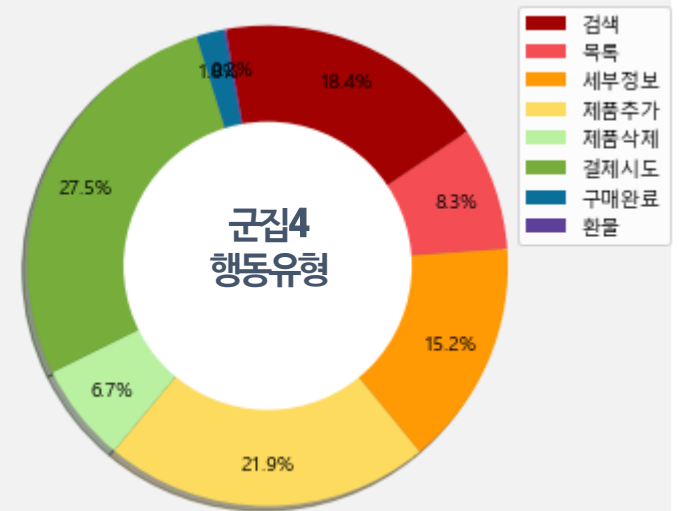
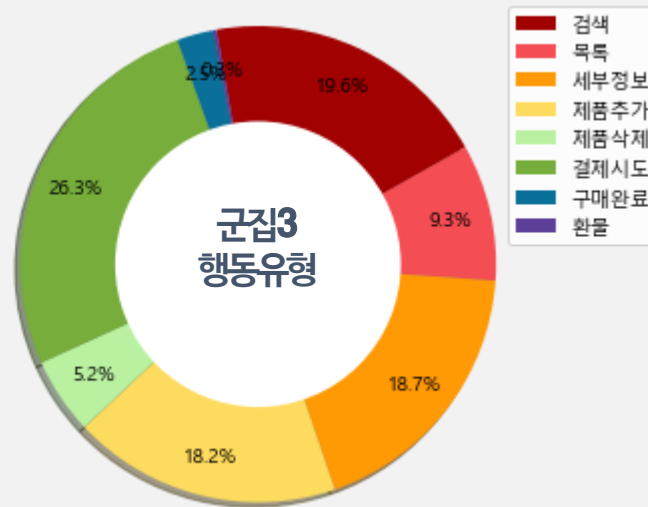
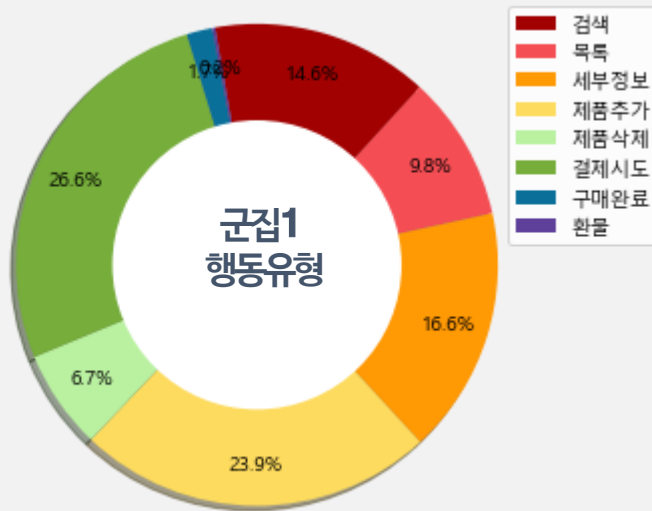
- ✓ 8월 22일 이후 구매 내역이 없음
- ✓ 매출액이 점점 감소하는 경향을 보임

군집분석

온라인 데이터

| 군집 별 행동유형

군집 1 : 924명
 군집 3 : 3329명
 군집 4 : 5767명



- ✓ 크게 차이나는 부분은 검색/제품추가
- ✓ 검색의 경우 군집3>군집4>군집1 → 단골 고객일수록 검색보단 살 물건을 정하고 들어옴
- ✓ 제품추가 경우 군집1>군집4>군집3 → 단골 고객일수록 제품추가를 많이 함

군집분석

온라인 데이터

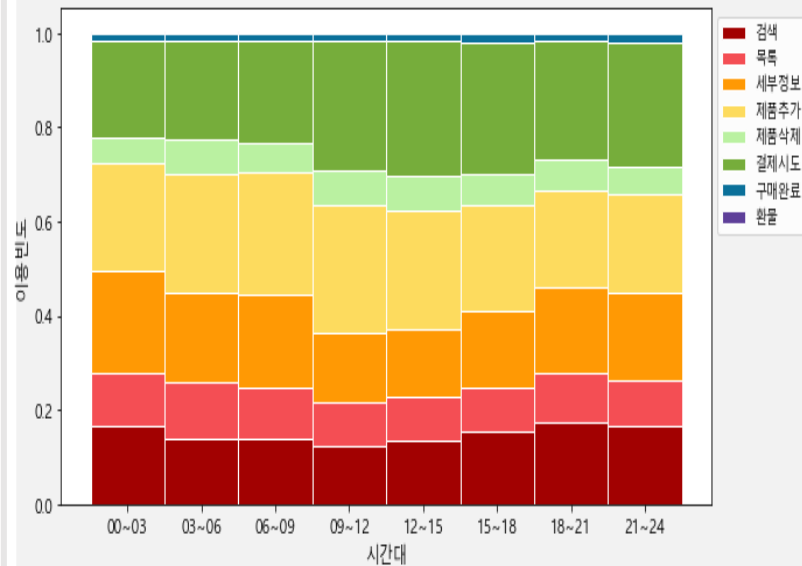
| 군집 별 시간별 행동 유형

군집 1 : 924명

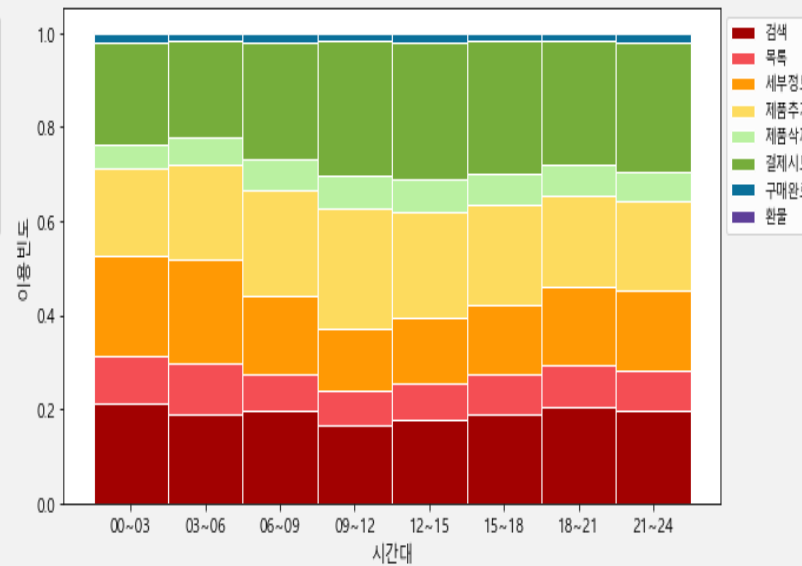
군집 3 : 3329명

군집 4 : 5767명

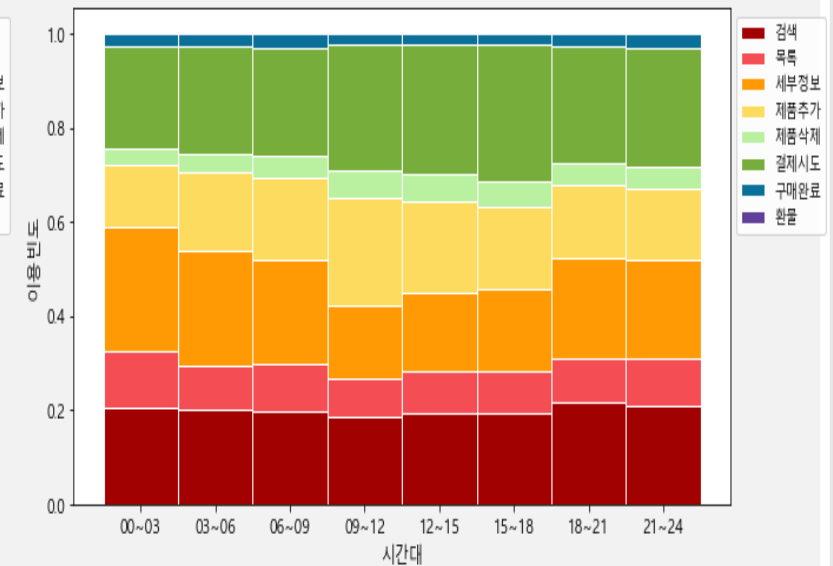
군집1 시간별 행동 유형



군집3 시간별 행동 유형



군집4 시간별 행동 유형



✓ 군집1이 군집3, 군집4에 비해서 검색 비율이 적은 걸 볼 수 있다.

군집분석

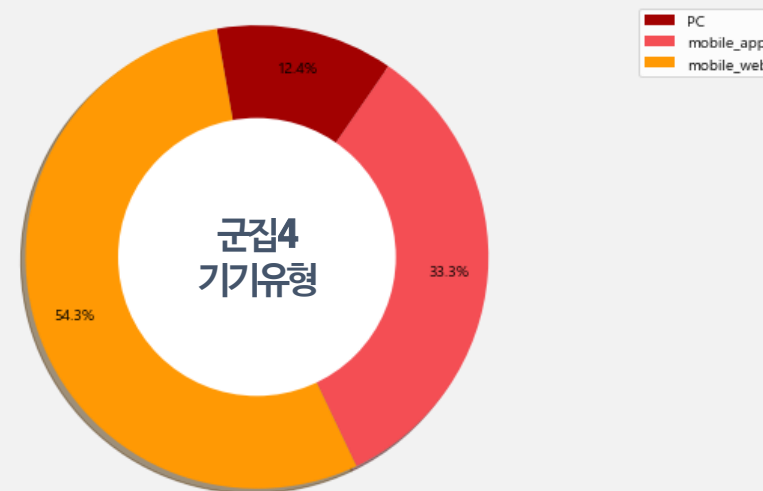
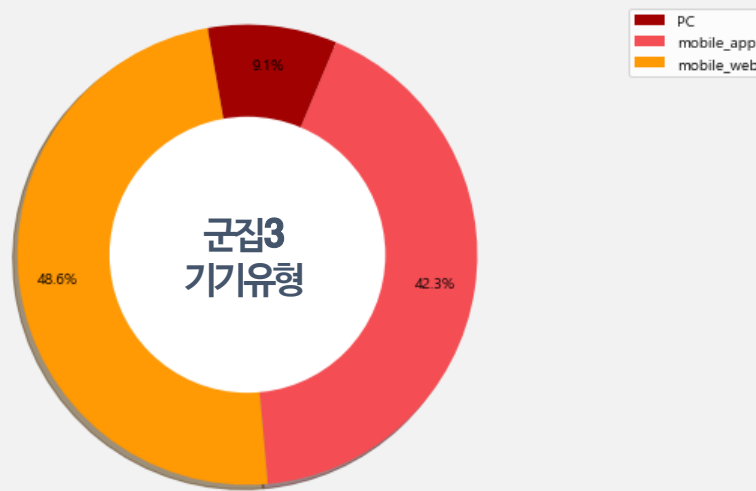
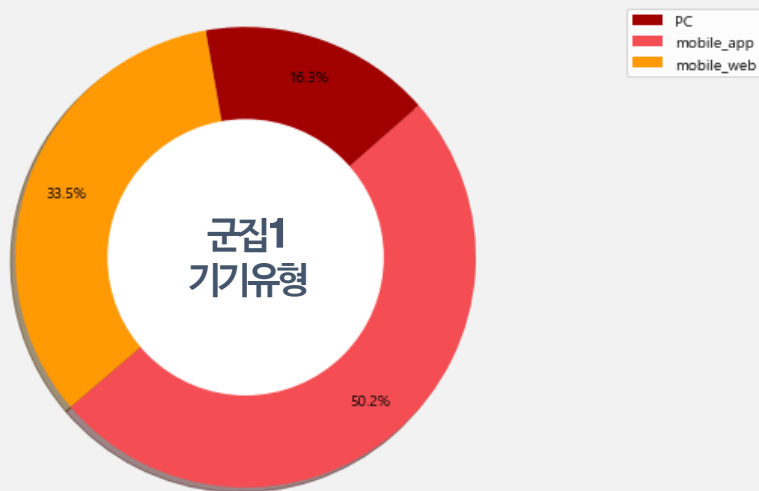
온라인 데이터

| 군집 별 Device 사용유형

군집 1 : 924명

군집 3 : 3329명

군집 4 : 5767명



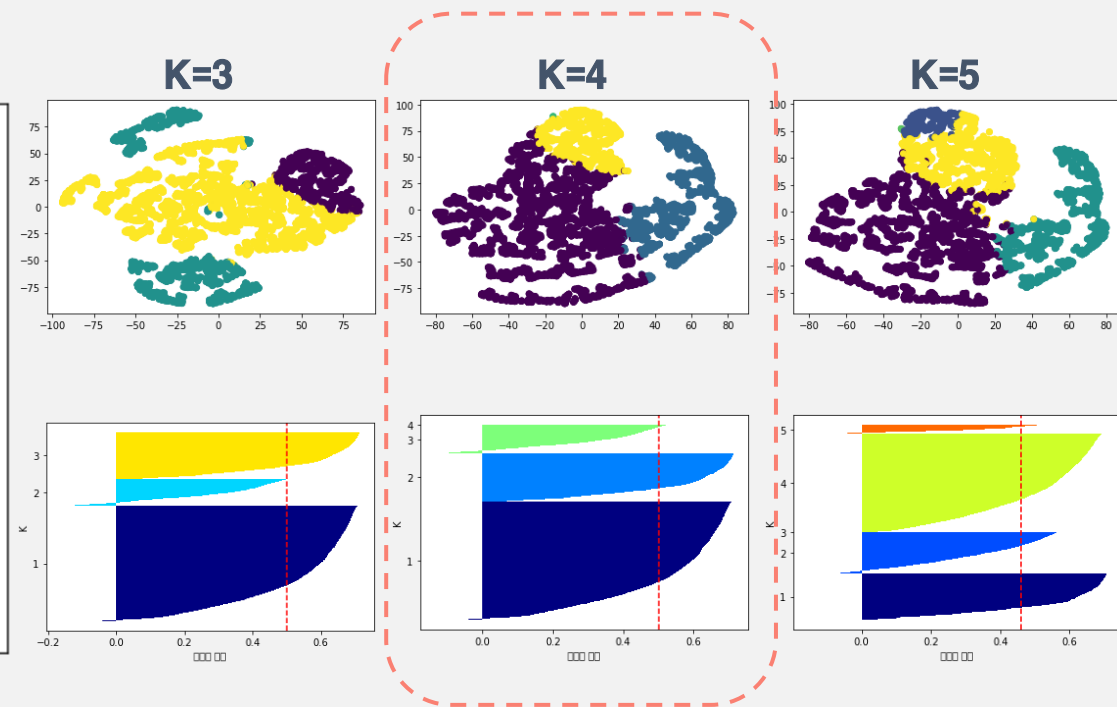
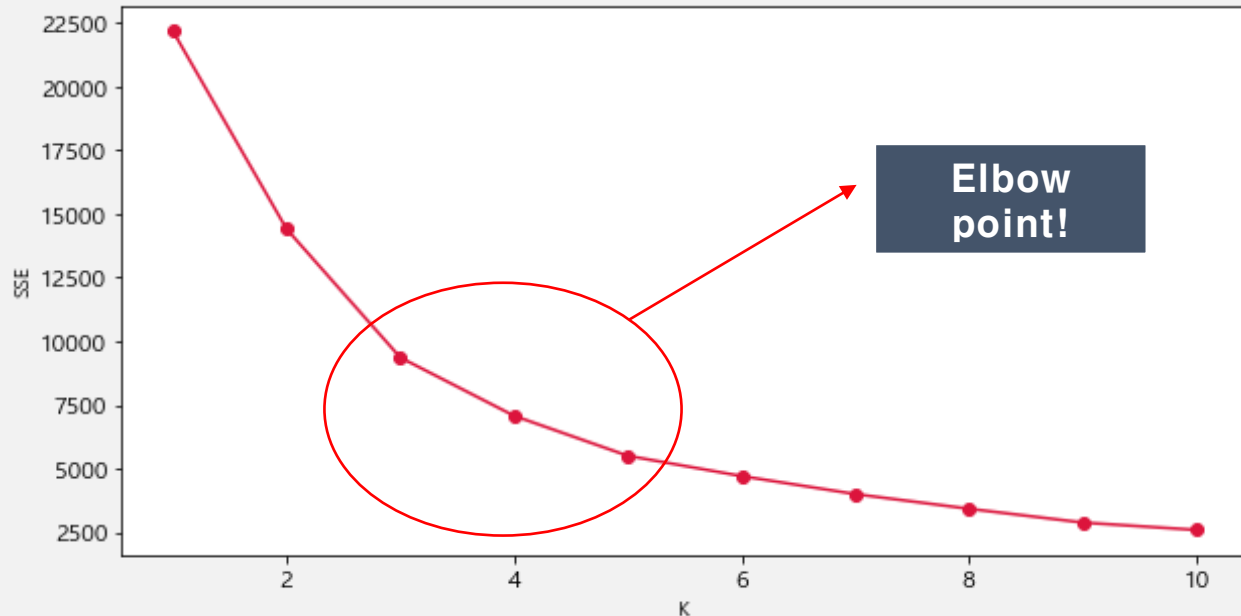
- ✓ 단골 고객일 수록 PC와 mobile app의 비율이 크며 mobile web의 비율이 낮다.
- ✓ 이는 **mobile web**이 PC와 mobile app에 비해 **쇼핑에 불편함**을 의미
- ✓ **Mobile web의 개선**을 통해 **편리한 쇼핑환경** 제공
- ✓ 군집3과 군집4 고객들에게 **app설치**를 권장하도록 해야함

군집분석

오프라인 데이터

| 최적의 군집 수 찾기

군집 수에 따른 오차제곱합

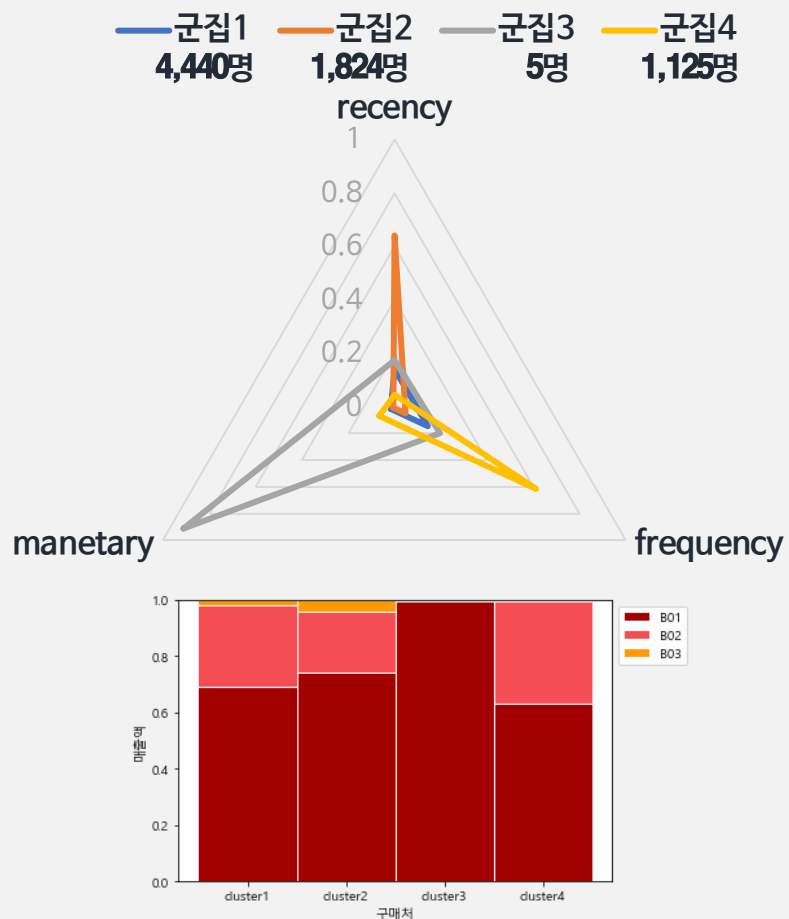


- ✓ 군집 1~10개의 오차제곱합을 계산하여 **elbow point**로 최적의 군집 수를 추적한다.
- ✓ K=3, 4, 5의 산점도와 실루엣그림을 각각 비교하여 **군집의 수를 4개가 적절하다고 판단했다.**

군집분석

오프라인 데이터

군집별 특징



군집1

- ✓ 방문 시점이 최근, 방문 빈도는 적음. 구매 액수가 작은 군집
→ 매출액 9월에 증가시킴
- ✓ 최근 방문 고객들

군집2

- ✓ 방문 시점이 오래됨, 방문 빈도는 적음, 돈은 적게 씀, 구매 액수가 작은 군집
- ✓ 20대 비중이 높고 B03방문이 가장 많은 고객들

군집3

- ✓ 최근에 온 소수인원들, 구매 액수가 큰 군집
- ✓ 선물세트 구매한 고객들

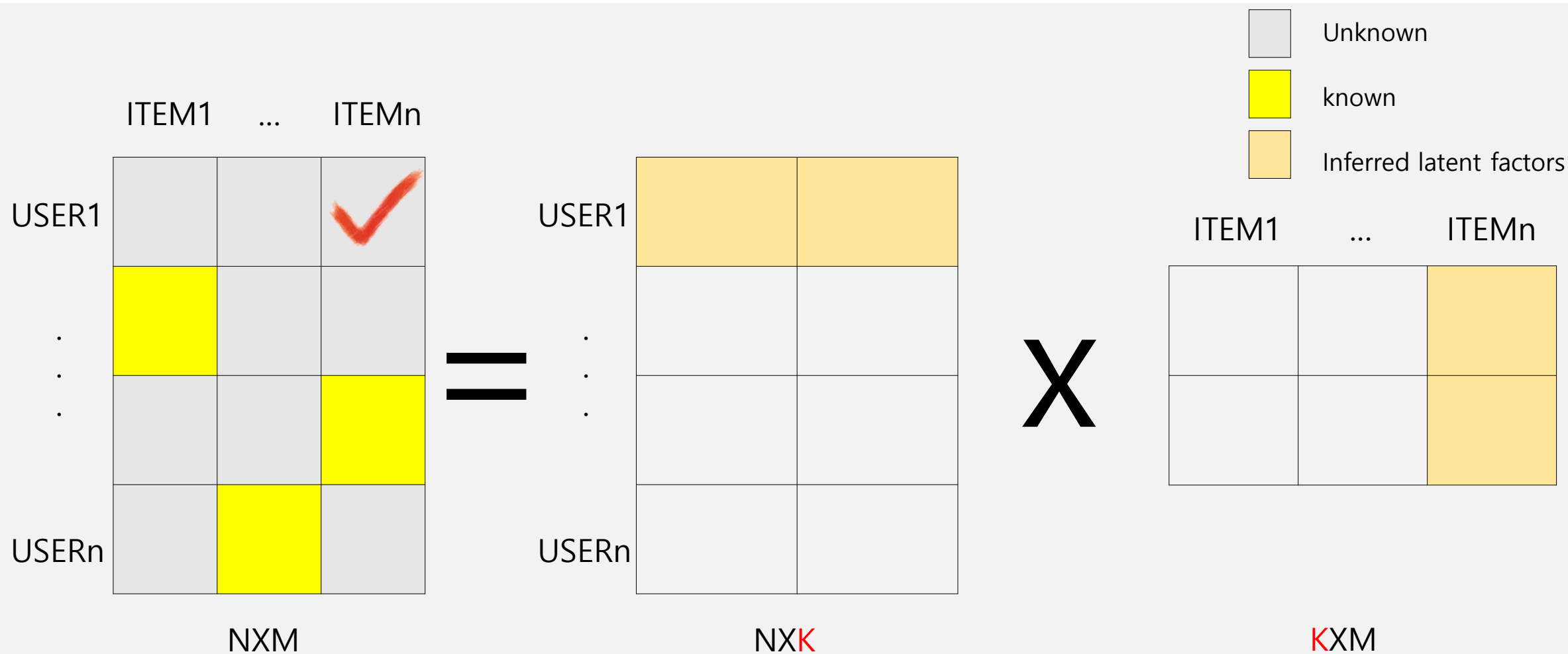
군집4

- ✓ 방문 빈도가 높고. 구매 액수가 작은 군집
- ✓ 50대 비중이 높고 매출액 일정수준 유지해줌.
- ✓ 어머니/아버지 고객들

- ✓ 오프라인매장 자체가 주 제품이 가정용품, 식품이다.
- ✓ RFM변수를 기반으로 만든 군집에 대해 군집간 구매하는 제품에 차이가 없다. → 구매처 별로 연관성분석을 시행

추천시스템 모델링

온라인 : MF (Matrix Factorization)



추천시스템 모델링

온라인

| Datasets

Explicit Data



사용자가 정확하게 본인이 얼마나 이 Item에 호감이 있는지를 수치로 나타낸 것 (ex : 평점)



Implicit Data



구매내역, 검색내역, 검색 패턴 등 사용자 패턴을 나타내는 데이터

추천시스템 모델링

온라인: ALS with Implicit Data

ALS란?

ALS는 Alternative Least Square로 MF의 목적함수를 최적화하는 기법

Implicit Data의 문제점?



사용자의 호 불호를 정확하게 판단할 수 없다.
(상품을 구매했다고 반드시 그 상품에 호의적인 평가를 내렸다고 볼 수 없음)

추천시스템 모델링

온라인 : ALS with Implicit Data

| ALS Algorithm

MF 수식

$$\min_{q,p} \sum_{(u,i) \in K} (r_{ui} - q_i^T p_u)^2 + \lambda (\|q_i\|^2 + \|p_u\|^2)$$

$$r_{ui} = \text{rating}$$

- ✓ p 와 q 는 latent matrix
- ✓ r 은 rating으로써 preference를 나타냄
- ✓ MF의 목적은 Matrix Complement



ALS with Implicit Data 수식

$$\min_{x,y} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$$

$$p_{ui} = \begin{cases} 1 & r_{ui} > 0 \\ 0 & r_{ui} = 0 \end{cases}, \quad c_{ui} = 1 + \alpha r_{ui}$$

- ✓ MF에서 rating vector r 이 preference vector p 로 바뀜 (선호 비선호의 binary)
- ✓ Preference값을 항상 신뢰 불가능 \rightarrow confidence 개념 도입
- ✓ 수식에서 각각의 u, i 에 대해 confidence c_{ui} 가 곱해짐

추천시스템 모델링

온라인: ALS with Implicit Data

| 분석 흐름도

거래데이터

행동데이터

고객데이터

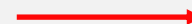
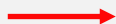
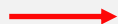
상품데이터

전처리

군집분석

★
USER X ITEM 테이블 생성

ALS 추천시스템



추천시스템 모델링

온라인: ALS with Implicit Data

| User X Item 테이블 생성

clnt_id	trans_id	trans_seq	biz_unit	...
...



고객번호	상품코드	구매횟수
...



U S E R	ITEM					
	1	0	0	1
	1	0	1	0
	0	0	0	0
	1	1	0	0

- ✓ 구매횟수는 trans_id 기준
- ✓ 추천은 상품 소분류 기준
- ✓ 고객번호, 상품 소분류 기준 groupby 를 이용해 테이블 생성

추천시스템 모델링

온라인

| 추천시스템 결과



Client ID : 2

추천시스템

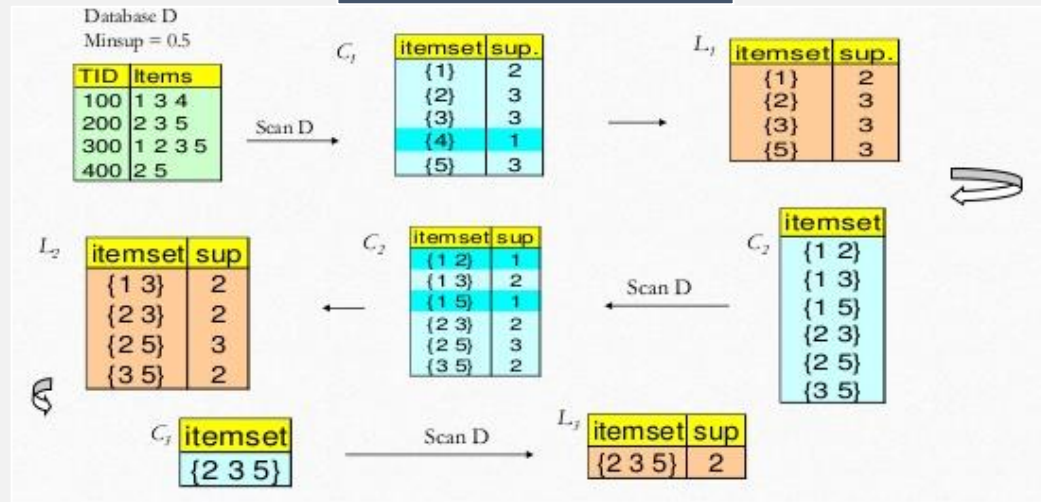
구매목록	추천상품	신뢰도
Coffee Drinks	General Yogurt	1
Canned Vegetable Foods	Infant/Toddlers' Pants	0.939952
Crab Sticks	Frozen Fried Foods	0.887074
Fried Tofu	Bibim Ramens	0.882412
Cream and Condensed milk	Water	0.869237
Infant/ Toddlers' T-shirts/ Tops	Pickled Radishes	0.869237
Men's T-shirts	Fresh Milk	0.868692
Ramens	Chicken Eggs	0.865387

추천시스템 모델링

오프라인 : 연관성분석

| Apriori Algorithm

Apriori Algorithm



출처: <https://www.digitalvidya.com/blog/apriori-algorithms-in-data-mining/>

- ✓ 모든 가능한 항목집합 개수를 줄이는 방식의 연관성 분석(비지도학습)
- ✓ 최소지지도(support)를 지정해주어야 함

평가척도

Rule: $X \Rightarrow Y$

$$Support = \frac{freq(X, Y)}{N}$$

$$Confidence = \frac{freq(X, Y)}{freq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

출처: https://www.saedsayad.com/association_rules.htm

- ✓ 지지도, 신뢰도, 향상도로 연관 규칙의 적절성을 평가함

추천시스템 모델링

오프라인 : 연관성분석

| Apriori Algorithm

Rule: $X \Rightarrow Y$

$Support = \frac{freq(X, Y)}{N}$

$Confidence = \frac{freq(X, Y)}{freq(X)}$

$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$

- ✓ **지지도(support)**
- ✓ 전체 구매이력에서 특정 제품이 등장하는 빈도, 가치치기의 기준
- ✓ 전체 거래 중 제품 A와 제품B를 동시에 포함하는 거래의 비율
- ✓ 지지도가 높을수록 유의미함

- ✓ **신뢰도(confidence)**
- ✓ 연관규칙의 강도
- ✓ 제품 A를 포함하는 거래 중에서 제품 B가 포함된 거래의 비율
- ✓ 신뢰도가 높을수록 유용한 규칙일 가능성이 높음

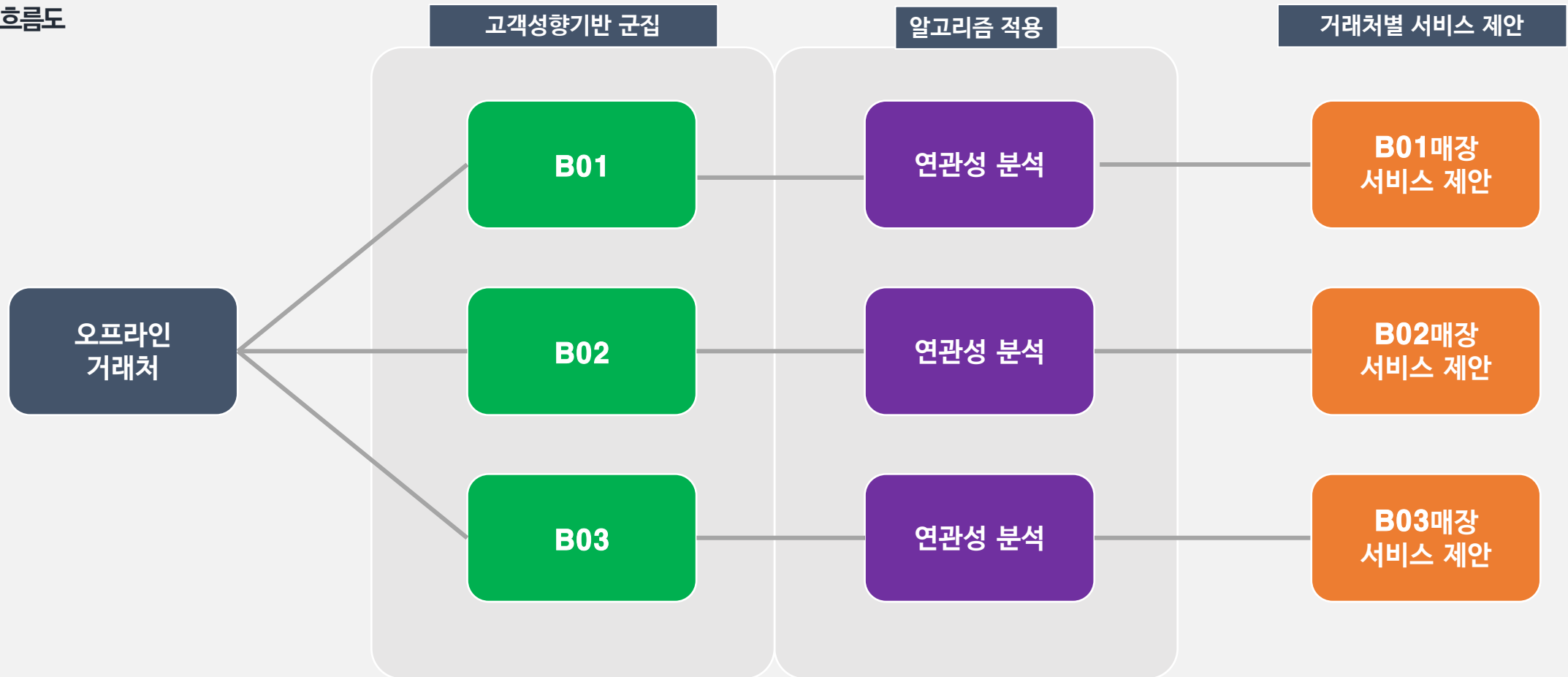
- ✓ **향상도(lift)**
- ✓ 두 제품 간 연관성
- ✓ 제품 B가 임의로 구매될 확률에 비해 제품 A를 구매한 후에 항목 B를 구매할 확률의 비율
- ✓ lift=1이면 두 제품의 구매는 독립, lift>1이면 두 제품의 구매는 연관성이 있음

- ✓ 지지도가 일정 값 이상인 연관 규칙을 먼저 골라낸 후
- ✓ 신뢰도가 일정 값 이상인 연관 규칙을 선정한다.

추천시스템 모델링

오프라인 : 연관성분석

| 분석 흐름도



추천시스템 모델링

오프라인 : 연관성분석

| B01



추천시스템 모델링

오프라인 : 연관성분석

| B02



추천시스템 모델링

오프라인 : 연관성분석

| B03



서비스 제안

온라인 서비스 제안

| 라이프스타일 파악

남성 고객



- ✓ 가전/가구의 구매액이 높다.
- ✓ 7, 8월의 구매액이 높음 (계절적 이유로 소비)
- ✓ A01과 A02에서의 구매액이 높음
- ✓ 인당 구매액은 여성보다 높음
 - 7, 8월에 가전가구 중심의 PUSH 알람 필요
 - A01, A02로의 더 많은 남성회원 유도할 필요 있음

여성 고객



- ✓ 뷰티/의류/음식에서 구매액이 높음
- ✓ 추석 한달 전부터 할인과 이벤트를 통해 더 많은 식품 구매 유도
- ✓ 9월에 구매액 변동성이 큼
 - 9월 매출액이 안정되도록 노력할 필요 있음

서비스 제안

온라인 서비스 제안

| 라이프스타일 파악

A01 & A02



- ✓ 의류/뷰티/패션 압도적
- ✓ A03에 비해 각종 가전 가구 구매액 비율이 높음
- ✓ 많은 남성회원 구입 유도 필요

A03



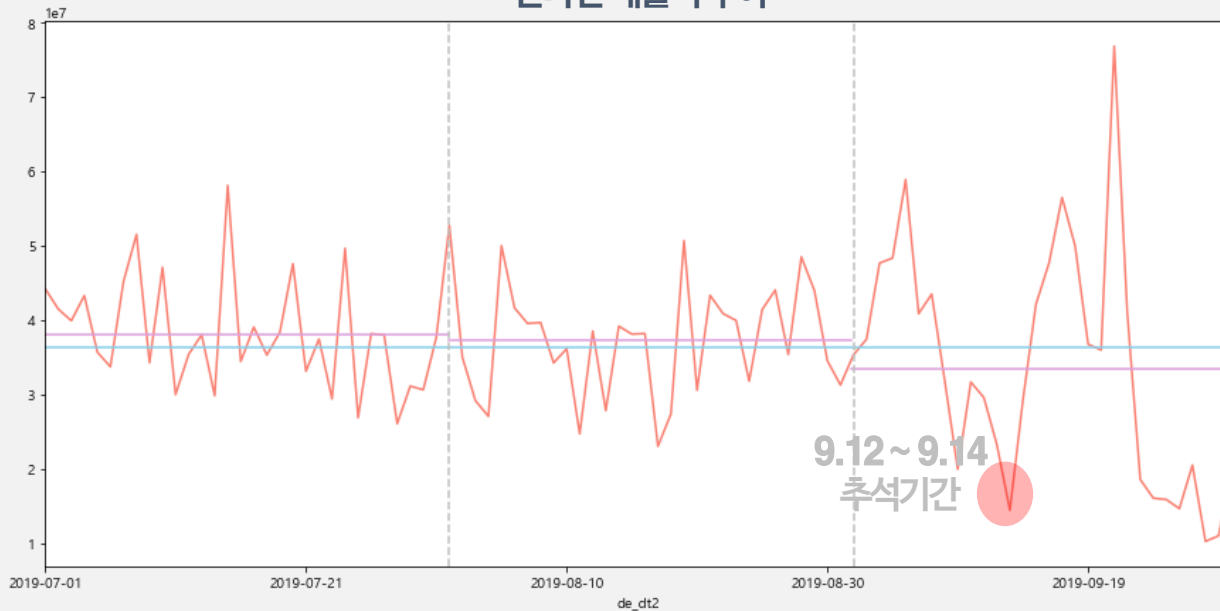
- ✓ 대부분의 구매상품이 '음식'
- ✓ 대부분 40대 여성의 비율이 높음
- ✓ 추석 연휴기간에 프로모션 필요
→ 40대 여성을 타겟으로 마케팅 필요

서비스 제안

온라인 서비스 제안

| 라이프스타일 파악

온라인 매출액 추이



기간



- ✓ 추석기간 매출 최하위
→ 추석기간 동안만의 프로모션과 이벤트 진행
→ 매출액 회복
- ✓ 특히 A03을 중점적으로 마케팅 진행

서비스 제안

온라인 서비스 제안

| 고객성향 파악

VIP 고객



VIP

- ✓ 검색 비율 ↓
- ✓ 제품 추가 비율 ↑
- ✓ App 사용 비율 ↑

이탈고객 & 보통고객



- ✓ 검색 비율 ↑
 - ✓ 제품 추가 비율 ↓
 - ✓ App 사용 비율 ↓
- 고객들에게 App 설치 권장 및 검색 시스템 개선

서비스 제안

온라인 서비스 제안

| 추천시스템

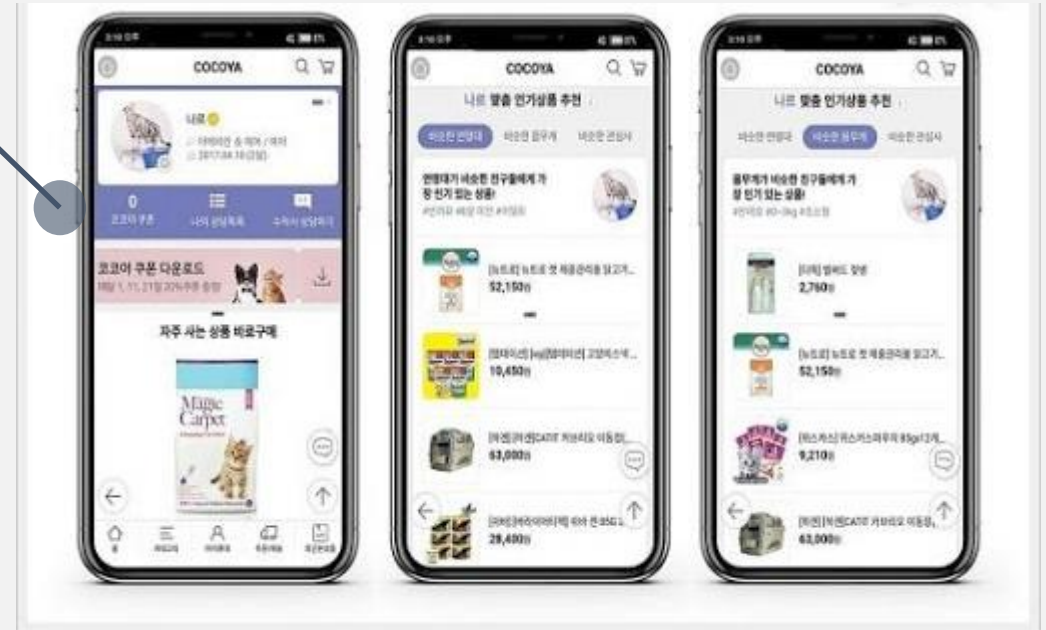
개인 맞춤 상품 PUSH 알람

다른 고객이 함께 본 상품

1/3



쇼핑 사이트에서 상품 추천 제시



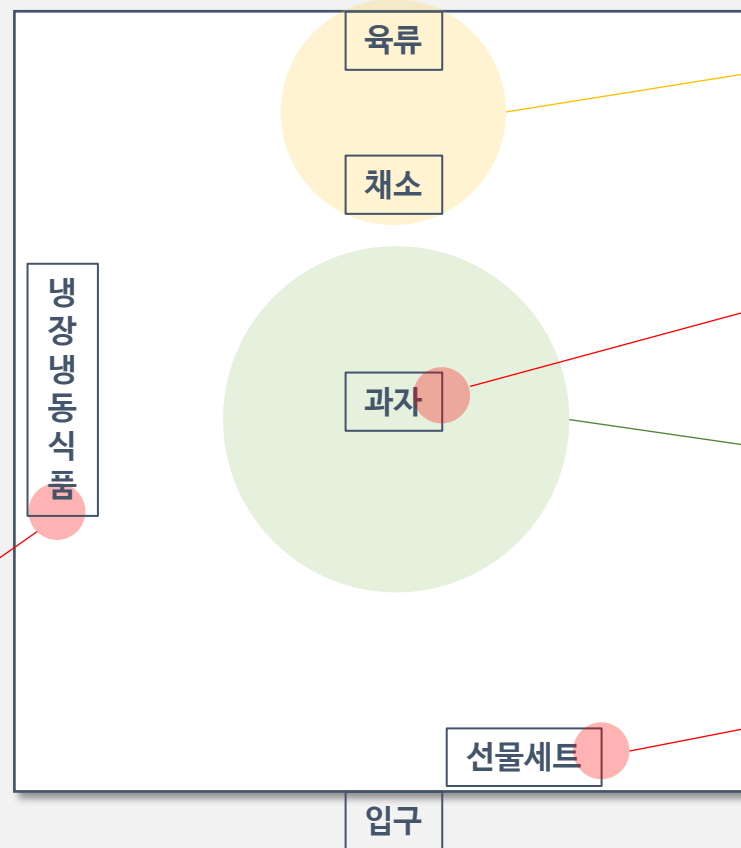
서비스 제안

오프라인 서비스 제안

| B01

- ✓ 식료품 판매 매장
- ✓ 과일, 채소의 매출액이 가장 많은 매장
- ✓ 추석 선물세트가 매출액의 영향을 많이 받은 매장
- ✓ 특히 추석 시즌에만 방문하는 고객들을 사로잡아야 함

✓ 냉장, 냉동식품은 같이 진열한다.



✓ 고기와 채소는 가까이 진열한다.

✓ 과자를 매장 가운데에 두고, 젤리, 사탕 등 가벼운 간식거리도 같이 진열한다.

✓ 과자 주위에 기타 먹거리를 배치한다.

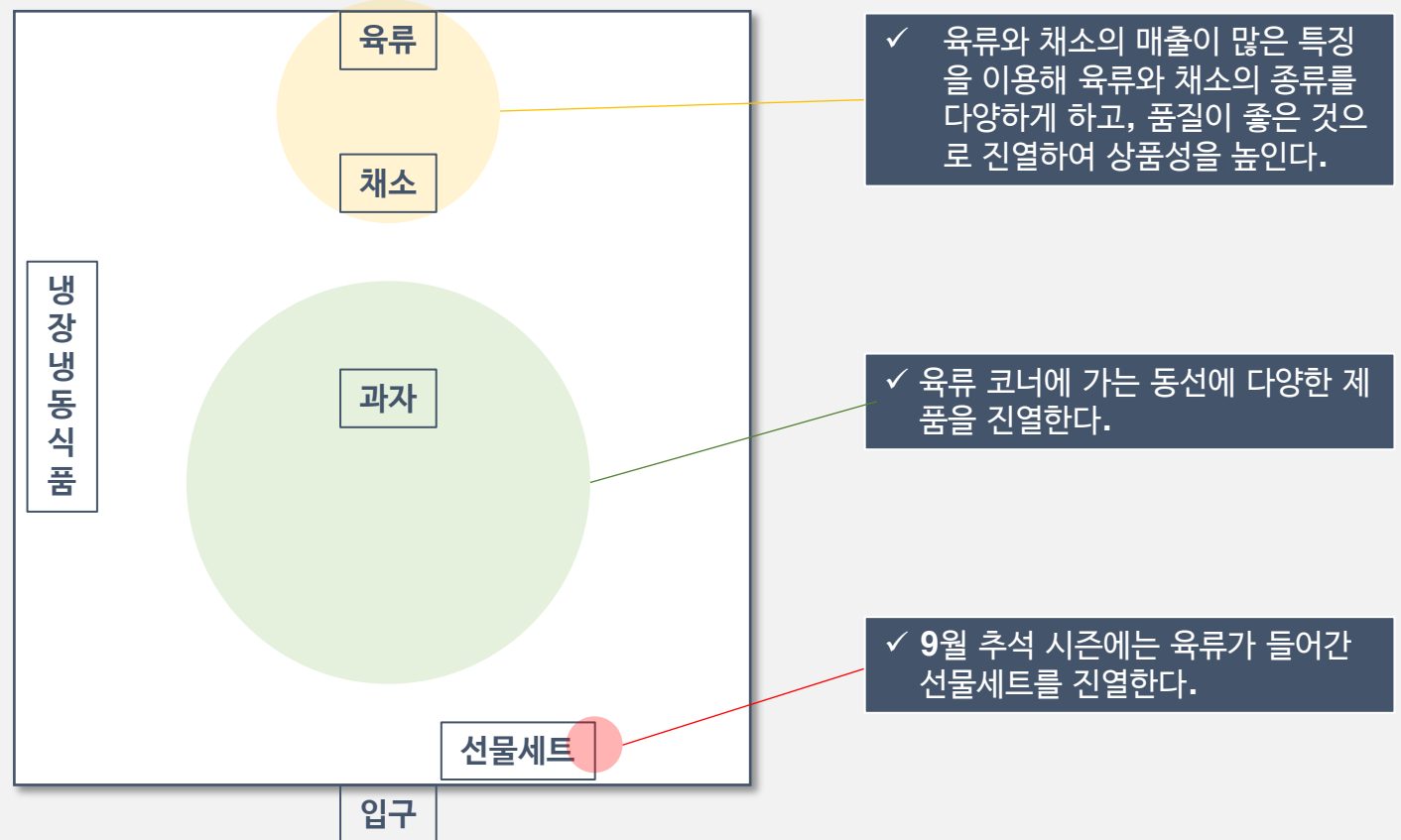
✓ 9월에는 추석 전에 유동인구가 많은 매장 입구 앞에 추석 선물세트를 진열한다.

서비스 제안

오프라인 서비스 제안

| B02

- ✓ 식료품 판매 매장
- ✓ 육류와 채소의 매출액이 많은 매장
- ✓ 50대가 주요 고객들
- ✓ 방문 빈도는 높지만 구매액이 높지 않은 특징이 있으므로 육류, 채소의 상품성을 높게 유지하고 그 외 제품의 매출액을 올릴 수 있도록 함
- ✓ 방문 빈도가 높은 고객들이 이탈하지 않도록 단골고객 관리 전략을 세우고 실천함
- ✓ 추석 선물세트가 매출액의 영향을 받지 않음

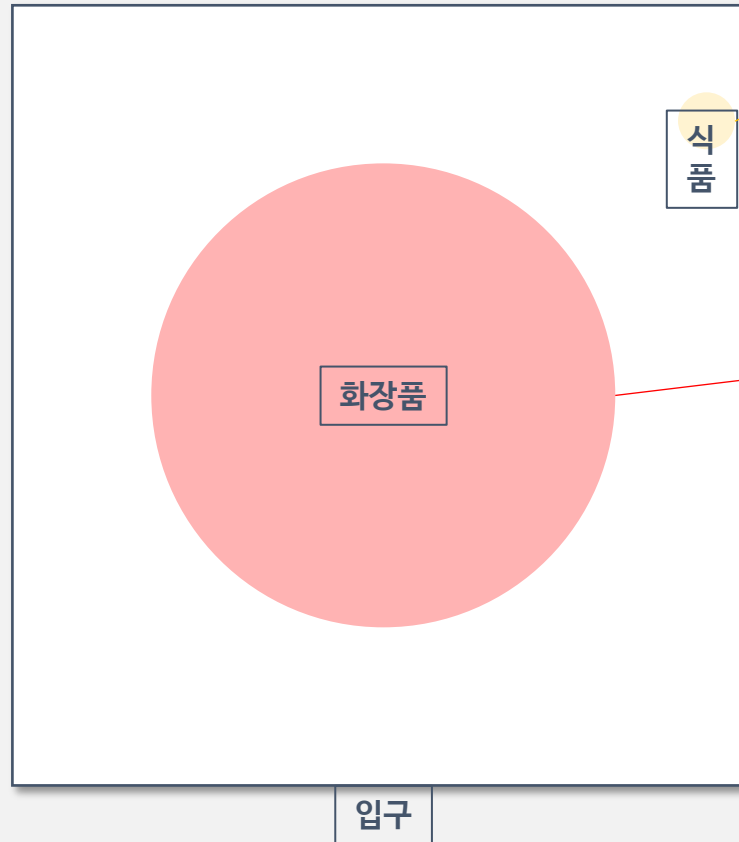


서비스 제안

오프라인 서비스 제안

| B03

- ✓ 화장품 판매 매장
- ✓ 20대 고객의 비중이 높음
- ✓ 방문 주기가 긴 특징이 있으므로 단골 고객을 만들거나 주기적 할인을 통해 고객의 방문 주기를 짧게 함
- ✓ 전체적으로 매출액이 적은 특징이 있으므로 한번씩 매장의 매출액을 올릴 수 있는 전략을 세움



✓ 간단한 식품을 매장 가장자리에 진열한다.

✓ 다양한 화장품 제품을 매장 전체에 진열한다.

감사합니다
Thank You!