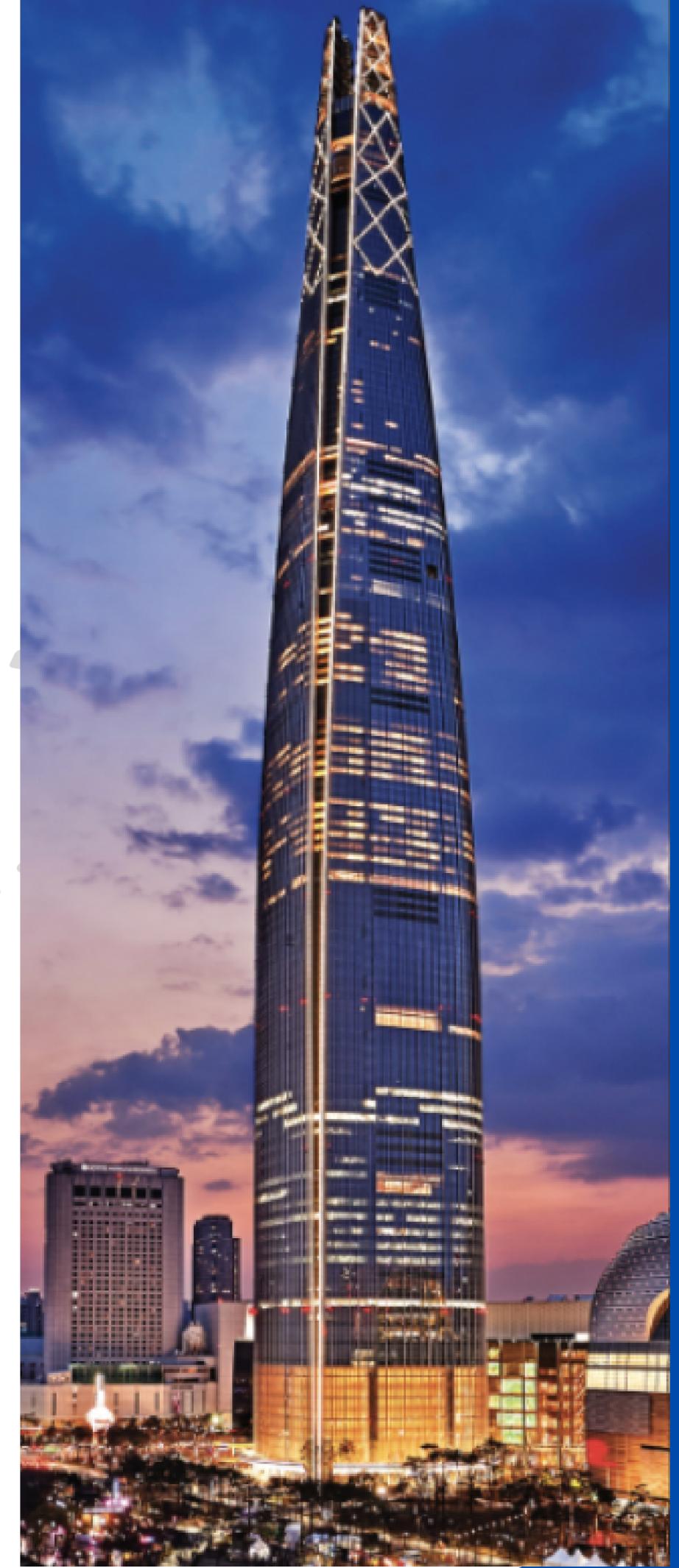


머신러닝을 활용한 위치 기반 점포 관리 시스템

2조 | 장유선 박지훈 이주성 류소리 유민지



Contents

1

기획 배경 및 목표

- 매장 점포 위치 추정
- 품목별 매출 감소 구
- 품목별 매출 감소 점포
- 목표
- 기대효과
- 분석 과정

2

데이터 수집 및 전처리

- 활용 데이터
- 카테고리 재정의
- 필수재/사치재 정의
- 피처 설정
- 데이터 및 레이블 설정
- 등급화 진행

3

모델링 및 분석

- 분류 모델 비교 및 모델 선정
- 분류 결과
- 군집화
- Decision Tree / GridSearchCV
- 분석 내용
- 분석 솔루션

4

결론

- 최종 제안
- 한계점
- Q&A



1

기획 배경 및 목표

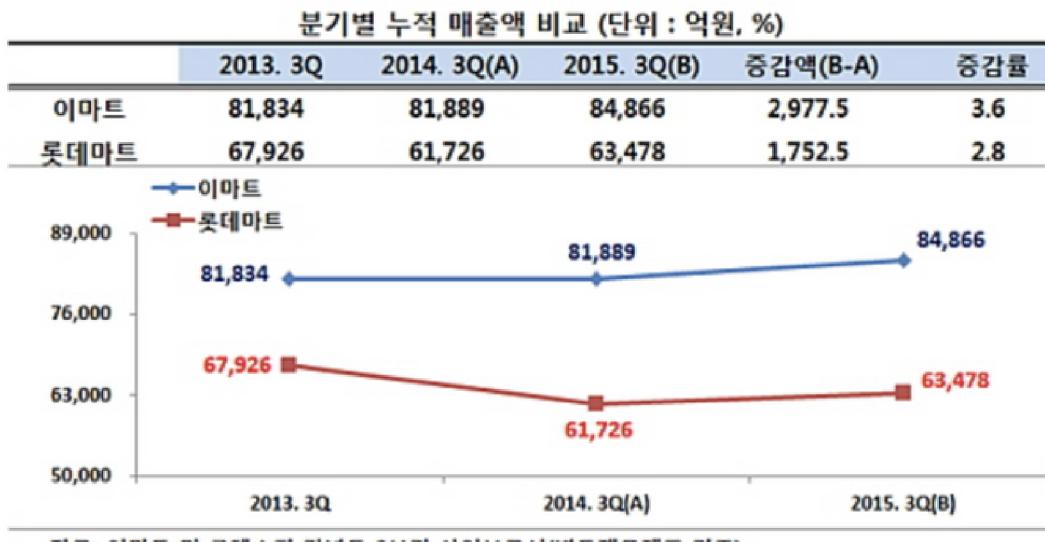
Background

기획 배경

마트, 마트에 완패...외형·손익 격차 커져

[조세일보] 민경증 전문위원

보도 : 2015.12.24 14:05 수정 : 2015.12.24 14:05



롯데쇼핑에서 운영하는 롯데마트의 최근 3년 분기영업실적(누적 기준)이 외형과 손익 모두 경쟁사 이마트에 크게 뒤지면서 해를 거듭할수록 그 격차가 확대되고 있는 것으로 나타났다.

각사 분기보고서에 따르면 별도재무제표 기준 두 회사의 3분기 누적 매출 격차는 지난 2013년 1조3908억원에서 2014년 2조163억원, 올해의 경우엔 2조1388억원으로 확대되는 추세다.

이에 따라 지난 2013년 이마트의 83.0%수준이었던 롯데마트 매출은 2014년엔 75.4%, 올해 3분기 누적으로는 74.8%수준까지 떨어졌다.

수익성의 경우엔 더 참담하다. 두 회사의 3분기 누적 영업이익 격차는 지난 2013년 3799.6억 원, 2014년 4176.5억원, 올 3분기엔 5347.3억원까지 급격히 벌어졌다. 롯데마트가 올 3분기에 306.3억원의 영업 결손을 기록한 탓이다.

종합해 보면 최근 3년 사이 롯데마트 매출은 이마트의 83.0%에서 74.8%로 떨어지고, 영업이익 또한 35.0%수준에서 -6.1%까지 떨어지는 등 외형과 손익 모두 완패를 당했다.

마트, 中 점포 일부 폐점 결정

박시은 기자 | 입력 2015.07.22 10:43 | 댓글 0 | 언어 선택 ▾ Google 번역에서 제공



[일요서울 | 박시은 기자] 롯데마트가 중국 산둥성 내 매장 4곳을 순차적으로 폐점키로 결정했다. 롯데마트는 중국에서 120개 점포를 운영하고 있다.

폐점이 결정된 매장은 칭다오시 2곳, 웨이하이시 1곳, 웨이팡시 1곳 등 모두 4개 매장이다.

업계는 롯데마트가 적자가 심한 영업점을 정리하고, 영업구조를 안정화시키기 위한 효율화 작업에 들어갔다고 보고 있다.

실제로 롯데마트의 중국 매출은 2013년 1조7300억 원 수준에서 지난해 1조5100억 원 수준으로 감소했다.

업계의 한 관계자는 “롯데마트의 중국 할인점 점포수는 2009년 79개에서 2013년 107개로 계속 증가했지만, 2014년에는 103개(신규 출점 4개, 폐점 8개)로 감소했다”며 “당분간 추가 출점 없이 부실 점포를 계속 정리할 것으로 보여 점포수는 추가로 감소할 가능성 있다”고 분석했다.

이에 대해 중국 관영 환구시보 영문판인 글로벌타임스는 지난 20일 롯데마트가 현지 업체를 비롯해 알리바바와 JD닷컴 등 글로벌 전자상거래 업체와의 경쟁에서 밀려 점유율이 하락했기 때문이라고 보도했다.

[2015 유통결산] 대형마트, 변해야 산다…점포 '흐림' 온라인 '맑음'

송고 2015.12.28 10:00 | 수정 2015.12.28 10:07

조호윤 기자 (hcho2014@ebn.co.kr)

저성장 기조 여전…매장 대형화 및 광역화로 집객 기대

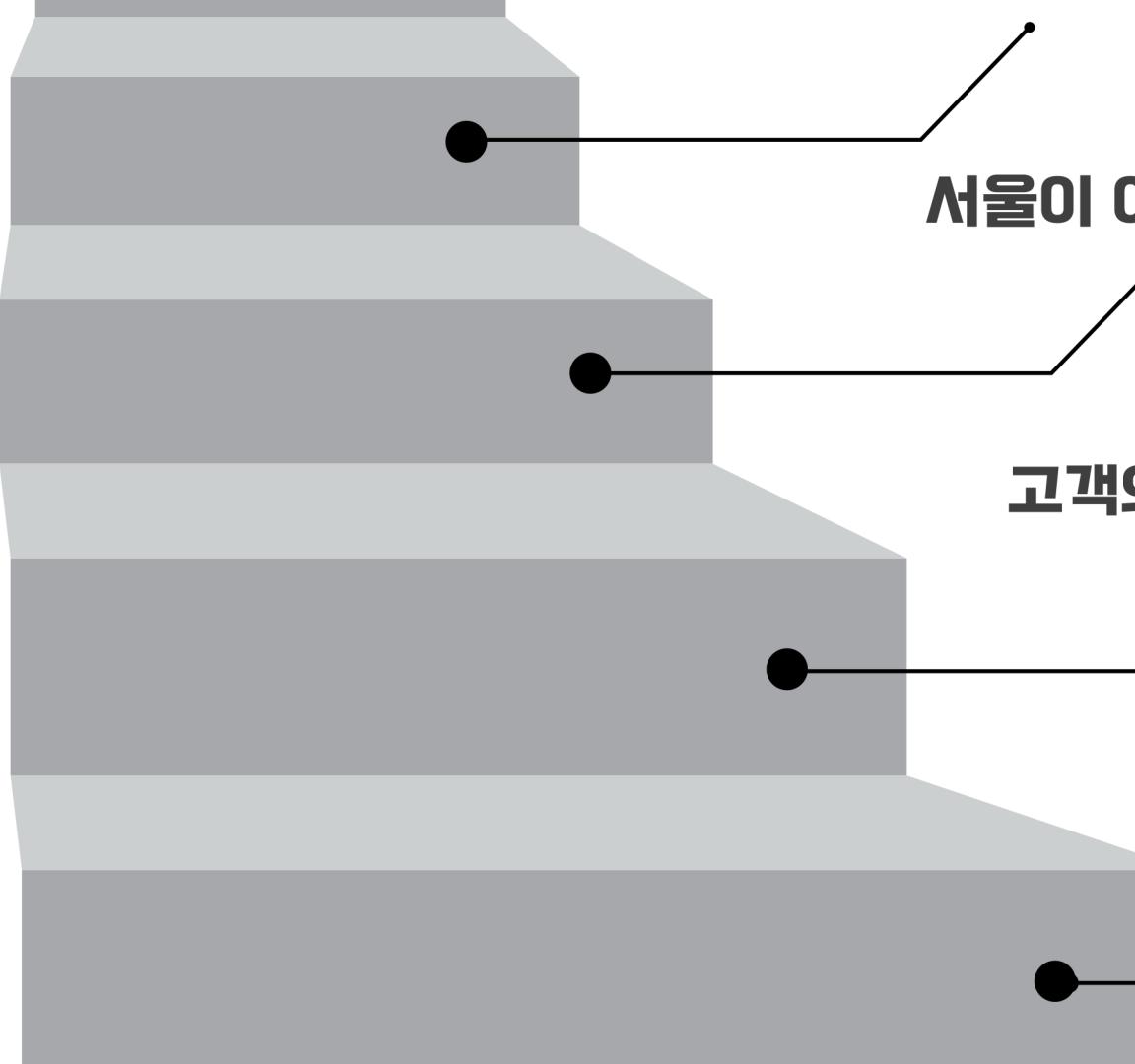
배송전쟁 본격화…온라인몰 구축·차별화 등으로 경쟁력 강화

Background

매장 점포 위치 추정



최종적으로 추려진 점포별 구매 고객들의
최다 거주지역을 해당 점포의 위치로 추정



서울이 아닌 점포 제외(163)

1차로 추린 점포 각각의
고객의 최다 거주지역(TOP3) 뽑기

서울특별시에 거주하는 고객이
가장 많이 이용한 점포 추리기(168)

고객DEMO 데이터 + 거주지역(우편번호 앞 3자리)

고객번호	성별	연령대	거주지역	시도	시군구
1 00001	M	60세미상	060	서울특별시	강남구
2 00002	M	60세미상	100	경기도	김포시
3 00003	M	60세미상	033	서울특별시	은평구
4 00004	F	60세미상	016	서울특별시	노원구
5 00005	M	60세미상	100	경기도	김포시
6 00006	F	60세미상	240	강원도	철원군
7 00007	F	60세미상	036	서울특별시	서대문구
8 00008	M	60세미상	010	서울특별시	강북구
9 00009	F	60세미상	100	경기도	김포시
10 00010	F	60세미상	024	서울특별시	동대문구
11 00011	M	60세미상	100	경기도	김포시
12 00012	M	60세미상	043	서울특별시	용산구
13 00013	F	60세미상	210	인천광역시	계양구
14 00014	F	60세미상	043	서울특별시	용산구

점포코드 시군구 고객번호

0	A1	김포시	532425
1	A12	강남구	1788753
2	A2	송파구	4965304
3	A20	노원구	2386083
4	A24	강북구	253367
...

추정 점포 위치(서울)

점포코드	점포코드		
시군구	시군구		
강남구	16	서대문구	9
강동구	3	서초구	7
강북구	7	성동구	5
강서구	10	성북구	5
관악구	3	송파구	28
광진구	7	양천구	3
구로구	2	영등포구	5
금천구	4	용산구	3
노원구	12	은평구	4
도봉구	7	종로구	3
동대문구	8	중구	1
동작구	1	중랑구	3
마포구	7		

Background

품목별 매출 감소한 구의 수

서울 자치구 내 품목별 증감율

시군구	대분류명	가공식품	가구/인테리어	교육/문화	기타	디지털/가전	명품	스포츠/레저	신선식품	유아용품	일상용품	패션의류/잡화	행사상품
강남구	0.006846	0.261720	0.298606	0.266203	0.214495	-0.210762	-0.060272	0.090160	-0.085127	0.174023	-0.027413	-0.328211	
강동구	-0.077625	0.991646	0.272681	-0.507933	0.022484	0.089683	0.281412	-0.042164	0.355220	-0.147984	-0.035566	0.144393	
강북구	0.145830	1.697068	0.245713	-0.953625	3.897165	0.163974	0.191390	0.050930	-0.077177	0.019388	0.114330	-0.197020	
강서구	0.273338	0.186212	0.244512	0.929410	-0.835148	0.814793	-0.378959	0.202406	0.221471	0.062056	0.219298	-0.140763	
관악구	-0.038040	-0.312339	0.319537	2.622148	0.065247	-0.485299	-0.098115	0.085496	0.014083	0.064788	-0.176479	0.049276	
광진구	0.044832	1.111164	0.305779	-0.098024	0.503542	0.775108	0.140024	0.088525	-0.218896	0.202434	-0.132258	-0.071305	
구로구	0.121497	-0.630055	0.035738	-0.307115	-0.819866	-0.666149	-0.162549	-0.021026	-0.096152	-0.179120	0.283105	-0.135618	
금천구	-0.026906	0.315777	1.540794	3.377175	1.856362	2.274590	-0.344015	0.115405	-0.273354	-0.273468	0.407987	-0.096031	
노원구	0.083446	0.272998	-0.542361	0.113587	0.229557	1.191200	0.017770	0.084134	-0.211059	0.108463	0.057810	-0.232849	
도봉구	0.052857	0.237986	-0.283579	-0.057526	0.791344	0.347140	-0.081464	0.208774	0.481058	-0.034968	-0.049855	-0.143261	
동대문구	0.091879	0.326205	0.772726	-0.048531	0.273763	-0.262055	0.090656	0.038720	0.043432	0.183671	0.506731	0.091411	
동작구	0.342325	0.248245	0.670831	-0.572525	0.416323	0.033154	0.256944	-0.057836	-0.221532	-0.148650	-0.057718	0.015760	
마포구	0.230613	-0.319910	-0.200831	-0.400288	0.633517	-0.038837	-0.162327	0.078350	0.102370	-0.178492	-0.034217	0.245213	
서대문구	-0.178645	0.779287	0.553486	105.577428	-0.292841	-0.062285	0.363501	0.634554	0.092223	-0.313641	0.231863	-0.447585	
서초구	0.396219	2.273956	0.095680	1.327586	-0.051680	-0.234576	0.166428	0.027692	-0.132323	-0.055098	-0.135692	-0.004708	
성동구	0.024622	0.083141	0.033163	2.316553	-0.079306	0.443643	-0.004813	0.018940	-0.001585	-0.048069	-0.072759	-0.152991	
성북구	0.104813	0.429022	-0.031167	-0.737421	0.071089	0.748639	-0.024613	0.206418	-0.125640	-0.068254	-0.179300	-0.083031	
송파구	0.015379	0.183847	0.226751	2.017081	0.234118	-0.304787	-0.025759	0.052216	-0.033374	0.068945	-0.039731	0.012197	
양천구	-0.216716	-0.102263	-0.388908	0.666667	4.036131	14.163760	1.676988	-0.072683	-0.005389	0.472814	-0.064540	1.185039	
영등포구	0.041240	-0.765227	-0.101851	-0.073329	-0.043658	8.983471	0.008036	-0.010819	0.369213	0.555741	-0.249298	-0.180783	
용산구	-0.001123	-0.428645	1.384451	0.138606	0.474199	-0.445704	0.130254	-0.090406	-0.126535	0.125804	0.231007	-0.130492	
은평구	-0.102610	-0.210648	-0.076988	NaN	-0.373188	-0.161552	-0.353211	-0.070863	0.233285	0.239696	-0.099259	-0.484758	
중랑구	0.072823	0.486857	-0.557869	0.024516	-0.109069	0.080678	-0.110900	0.107064	-0.047178	-0.016474	-0.143088	-0.279654	
중구	-0.146577	1.226858	-0.650034	-0.892473	-0.563595	1.121400	-0.088968	-0.044207	-0.0707388	-0.082380	-0.045828	-0.181370	
중랑구	0.087994	0.244284	-0.094203	8.293518	-0.302351	-0.758485	0.186793	-0.187094	-0.234709	-0.199391	0.097856	-0.528033	

품목별 매출 감소한 구의 수

품목 감소_구_count

11	행사상품	18
8	유아용품	16
10	패션의류/잡화	16
6	스포츠/레저	13
9	일상용품	13
3	기타	11
5	명품	11
2	교육/문화	10
4	디지털/가전	10
7	신선식품	9
0	가공식품	8
1	가구/인테리어	7

서울 자치구 수 : 25개

대부분의 품목에서
약 절반 정도의 자치구에서
매출이 감소함



"그렇다면
점포 차원에서
품목별 매출의 증감은?"



Background

품목별 매출 감소한 점포의 수

[각 점포의 품목별 증감률]

대분류명	가공식품	가구/인테리어	교육/문화	기타	디지털/가전	명품	스포츠/레저	신선식품	유마용품	일상용품	패션의류/잡화	행사상품
점포코드												
A1	0.021849	0.032455	-0.243335	-0.854968	-0.036478	0.188918	0.010656	0.128137	0.616985	-0.110514	0.018667	-0.104271
A12	0.054299	0.231809	3.075342	NaN	0.132945	0.166658	-0.031111	0.041917	-0.082548	0.195984	-0.037285	-0.289504
A2	0.035644	0.307143	2.246490	NaN	0.326162	-0.259639	0.016894	-0.080941	0.011312	0.047404	-0.005055	0.244276
A20	0.075283	0.189046	-0.721394	NaN	0.234627	0.013583	-0.007374	-0.021847	-0.013483	0.191371	0.055811	-0.249010
A24	0.108646	0.193778	0.546480	NaN	2.157022	-0.041780	0.040358	0.558879	-0.120241	-0.100628	-0.003043	0.045845
...
D4	0.184965	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.430409	NaN	NaN	NaN
D5	-0.791797	NaN	-0.759841	NaN	NaN	NaN	NaN	NaN	-0.506475	-0.342723	NaN	NaN
D6	-0.338633	NaN	NaN	NaN	-0.949283	NaN	NaN	NaN	-0.320171	-0.011173	NaN	NaN
D7	0.180840	NaN	0.929425	NaN	-0.623681	NaN	-0.699074	NaN	NaN	0.284933	-0.058163	NaN
D8	-0.628891	NaN	NaN	NaN	-0.800513	NaN	NaN	NaN	NaN	0.383567	NaN	NaN

[품목별 매출 감소 점포의 수]

품목	감소_점포_count	최급점포수
0 가공식품	80	162
9 일상용품	78	161
8 유아용품	72	118
7 신선식품	69	149
10 패션의류/잡화	54	103
6 스포츠/레저	45	73
2 교육/문화	37	68
4 디지털/가전	33	64
11 행사상품	32	60
1 가구/인테리어	29	70
5 명품	5	11
3 기타	4	10

점포 규모에서 항목별 증감률을 살펴본 결과,
대부분의 항목에서 해당 상품을 취급하는 점포 중 절반 정도가 매출이 감소함

Q.

점포의 위치는 적절한가?

Q.

각 점포에 맞는 고객 타겟층을 설정하였는가?

Q.

주 고객층의 수요를 정확하게 파악했는가?

Q.

각 점포마다 다른 수요에 맞게 상품 구색을 갖췄는가?



오프라인 점포의 핵심은 "위치"

알맞은 점포의 위치 선정은 타겟 고객을 설정 및 고객 니즈 기반 점포 운영 및 마케팅의 시작

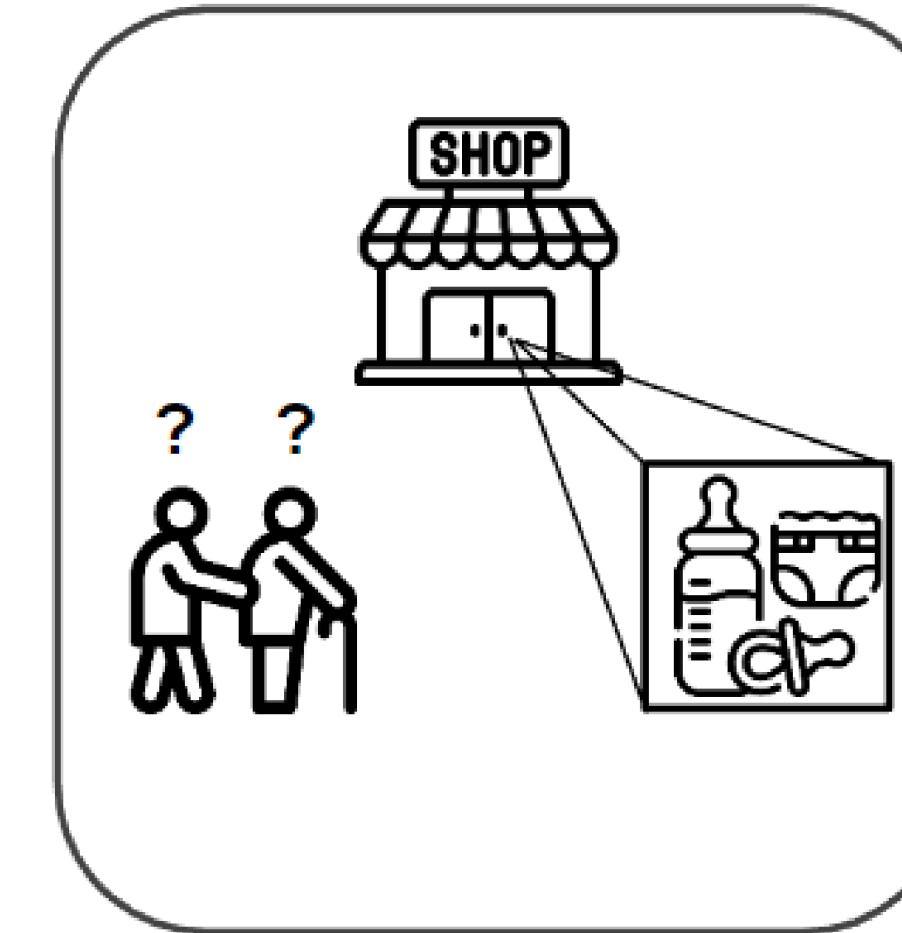


목표

인구 감소 지역에 새로운 점포를 오픈?



노령 인구가 많은 지역에 유아용품 체험 기획?



위와 같은 문제 방지를 위해
위치 및 해당 거주 인구 특성에 기반한 **점포 관리 시스템 개발**

기대효과

폐점
점포
후보
선정

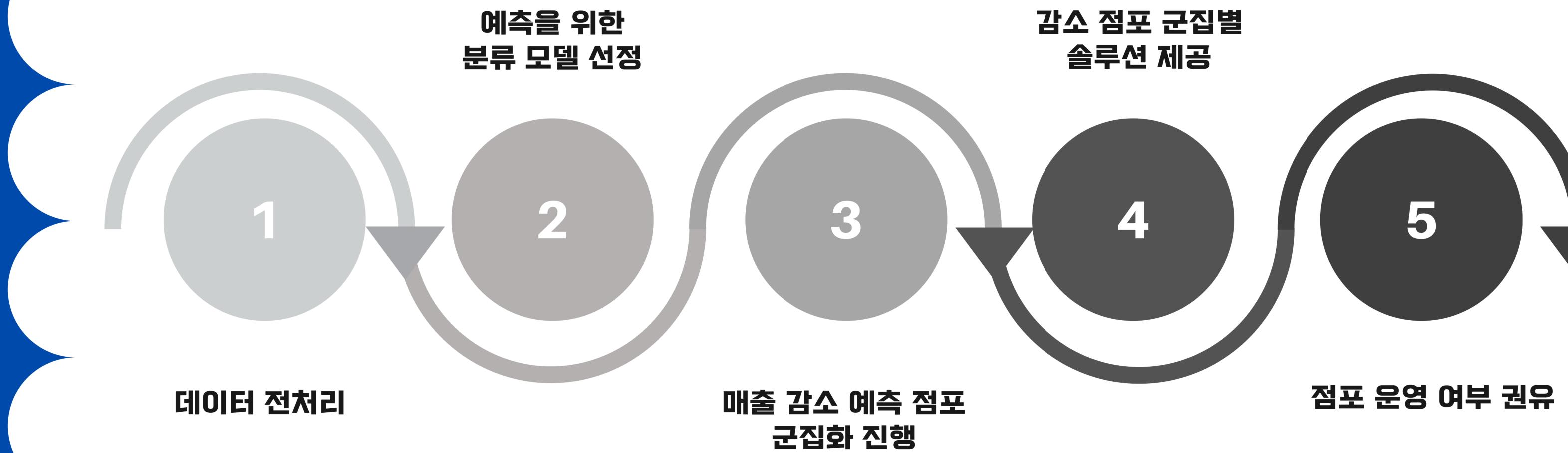
관리
점포
후보
선정

점포
상품
구성
관리



위치 및 상품 관련 효율적인 점포 관리 가능해짐

분석 과정





2

데이터 수집 및 전처리

활용 데이터

데이터명	데이터소스
1인가구(연령별)	https://data.seoul.go.kr/dataList/10995/S/2/datasetView.do
1인가구(거처별)	https://data.seoul.go.kr/dataList/10994/S/2/datasetView.do
아파트 매매	https://data.seoul.go.kr/dataList/10803/S/2/datasetView.do
주택가격지수(매매)	https://data.seoul.go.kr/dataList/801/S/2/datasetView.do
인구밀도	https://data.seoul.go.kr/dataList/10790/S/2/datasetView.do
혼인건수	https://data.seoul.go.kr/dataList/10108/S/2/datasetView.do
지하철역	https://www.data.go.kr/data/15081868/fileData.do?recommendDataYn=Y#layer_data_infomation
서울시 등록외국인 현황 (체류자격별/구별) 통계	http://data.seoul.go.kr/dataList/804/S/2/datasetView.do
서울시 등록외국인 현황 (국적별/구별) 통계	http://data.seoul.go.kr/dataList/803/S/2/datasetView.do
새우편번호 데이터	https://www.epost.go.kr/search zipcode/areacdAddressDown.jsp
서울시 대학 및 전문대학 DB정보	https://data.seoul.go.kr/dataList/OA-12974/S/1/datasetView.do
서울시 학교 총괄 통계	https://data.seoul.go.kr/dataList/199/S/2/datasetView.do
서울시 자치구별 평균 소득	KOSIS
분기별 구별 인구 데이터	https://data.seoul.go.kr/dataList/10718/S/2/datasetView.do
등록 체육시설	https://data.seoul.go.kr/dataList/1391/S/2/datasetView.do
공공문화공간간	https://data.seoul.go.kr/dataList/245/S/2/datasetView.do

카테고리 재정의

대분류	중분류	대분류	중분류	대분류	중분류
신선식품	농산물	유아용품	영유아식품	패션의류/잡화	의류
	축산물		영유아의류		언더웨어
	수산물		영유아용품		신발
	유기농	교육/문화	문구/완구/사무용품		가방
	조미료		악기		액세서리/잡화
	유제품/아이스크림		음반/음원/비디오		패션의류/잡화_기타
	농산가공		서적		스포츠의류/신발/잡화
	축산가공		문화생활		스포츠장비/용품
	수산가공		교육/문화_기타		캠핑용품
	냉동식품		컴퓨터/주변기기		골프용품
가공식품	건강식품	디지털/가전	통신가전	행사상품	식품행사
	과자류		영상가전		의류행사
	음료류		음향가전		가전행사
	빵류/떡류		학습기기		기타행사
	커피/차류		주방가전		명품의류
	주류		계절가전		명품잡화
	외식		생활가전		기타
	가공식품_기타		뷰티/헤어가전	명품	
	위생용품		카메라		
	청소/욕실용품		자동차기기		
일상용품	애완용품		부품/액세서리		
	자동차용품		디지털/가전_기타		
	주방용품	가구/인테리어	가구		
	생활용품		인테리어소품		
	미용용품		원예용품		
	의약품/의료기기		가구/인테리어_기타		
	일용잡화				

필수재, 사치재 정의

대분류명

신선식품

가공식품

유아용품

일상용품

패션의류/잡화

명품

교육/문화

디지털/가전

가구/인테리어

스포츠/레저

기타

구분

필수재

사치재

'의식주' 기준



피처 설정

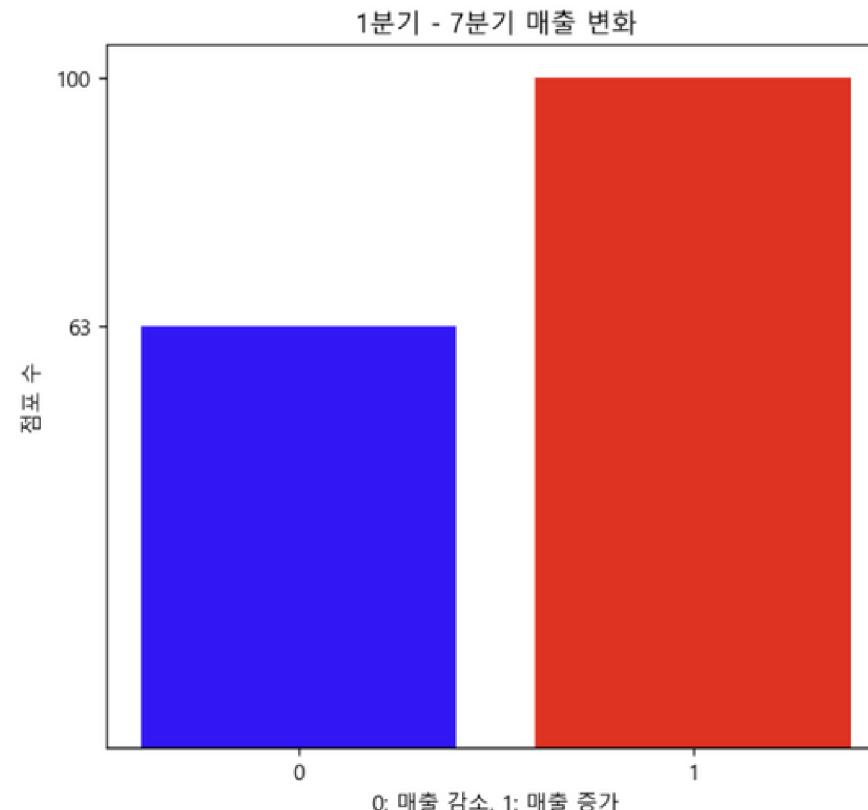
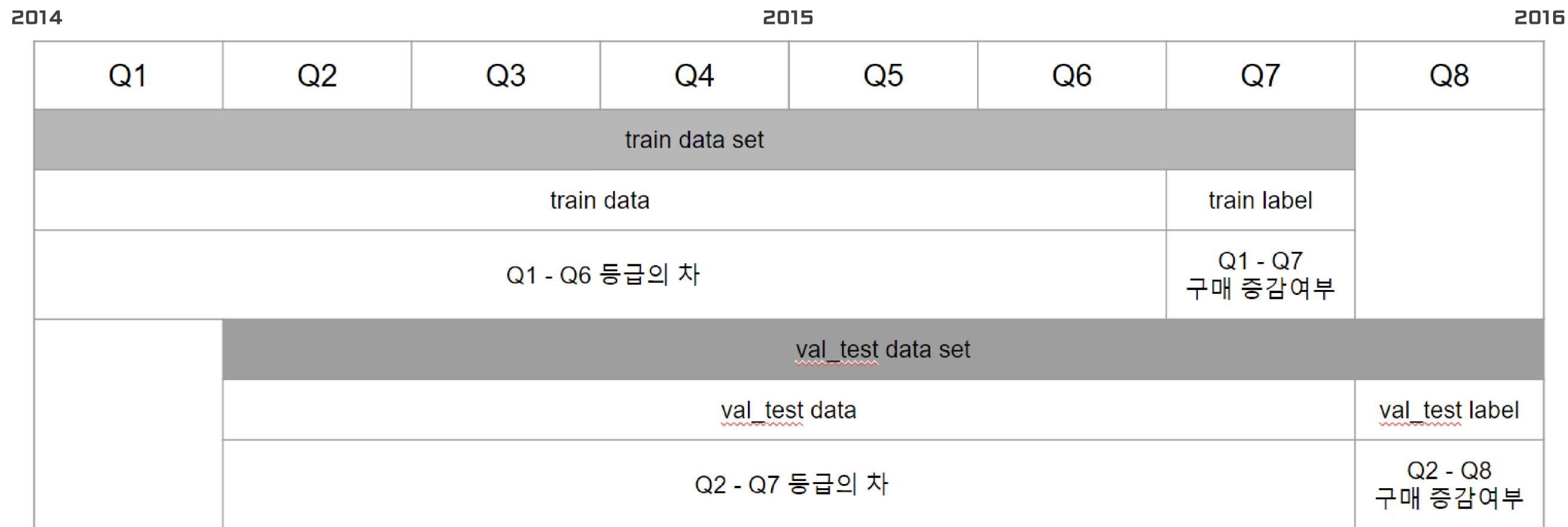
점포 매출에 영향을 주는 변수들을
다각도로 분석하여 유형화

상품 및 매출 관련	
변수 설명	변수명
최다 판매 대분류명	upperclass
최다 판매 중분류명	midclass
최다 판매 소분류명	lowclass
최대판매시간	pertime
필수재의 평균 구매 가격	ESSENTIAL
사치재의 평균 구매 가격	LUXURY
주중 구매 매출 비율	WORK_PCT
주중 구매건수 비율	WK_COUNT
매출 금액	SUM
매출 증감률	RATE
주중 필수재 구매 건수	work_e_count
주중 사치재 구매 건수	work_l_count

고객 관련	
변수 설명	변수명
여성 이용횟수(등급) 여성 구매금액(등급)	f_cntgrade f_amtgrade
남성 이용횟수(등급) 남성 구매금액(등급)	m_cntgrade m_amtgrade
상위고객 평균 이용 점포수	vip_strcount
점포별 고객의 이용 점포수	cust_strcount
소액결제 고객 매출기여도 고액결제 고객 매출기여도	sml_salespct lge_salespct
소액고객 비율 고액고객 비율	sml_custport lge_custport
재방문율	revisit_pct
연령대 비율(가중치와 곱)	age_sales
점포별 평균 구매액	AVG_PUR
점포 이용 횟수	STORE_CNT

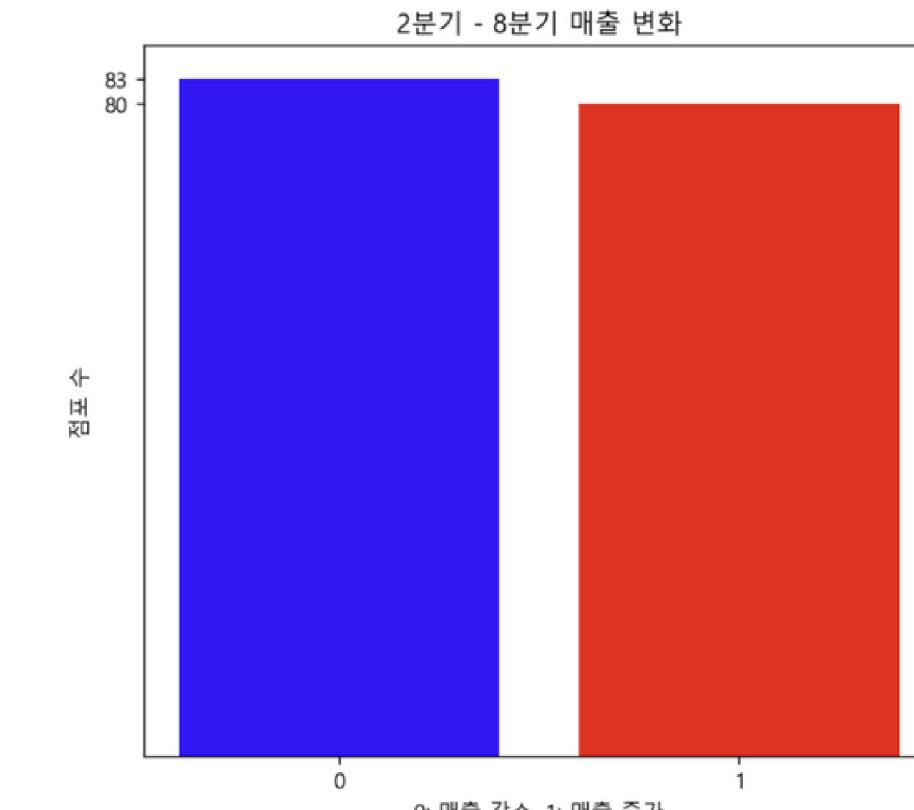
지역 관련	
변수 설명	변수명
인구 증감	인구 증감
자동차 등록수	car_regi
지역 연령대별 순위와 고객 이용 횟수 비율의 비율 (메이저, 마이너)	AGE_REGION_RATE
제휴사	제휴사

데이터 및 레이블 설정



매출 증가 점포

1-7분기 > 2-8분기

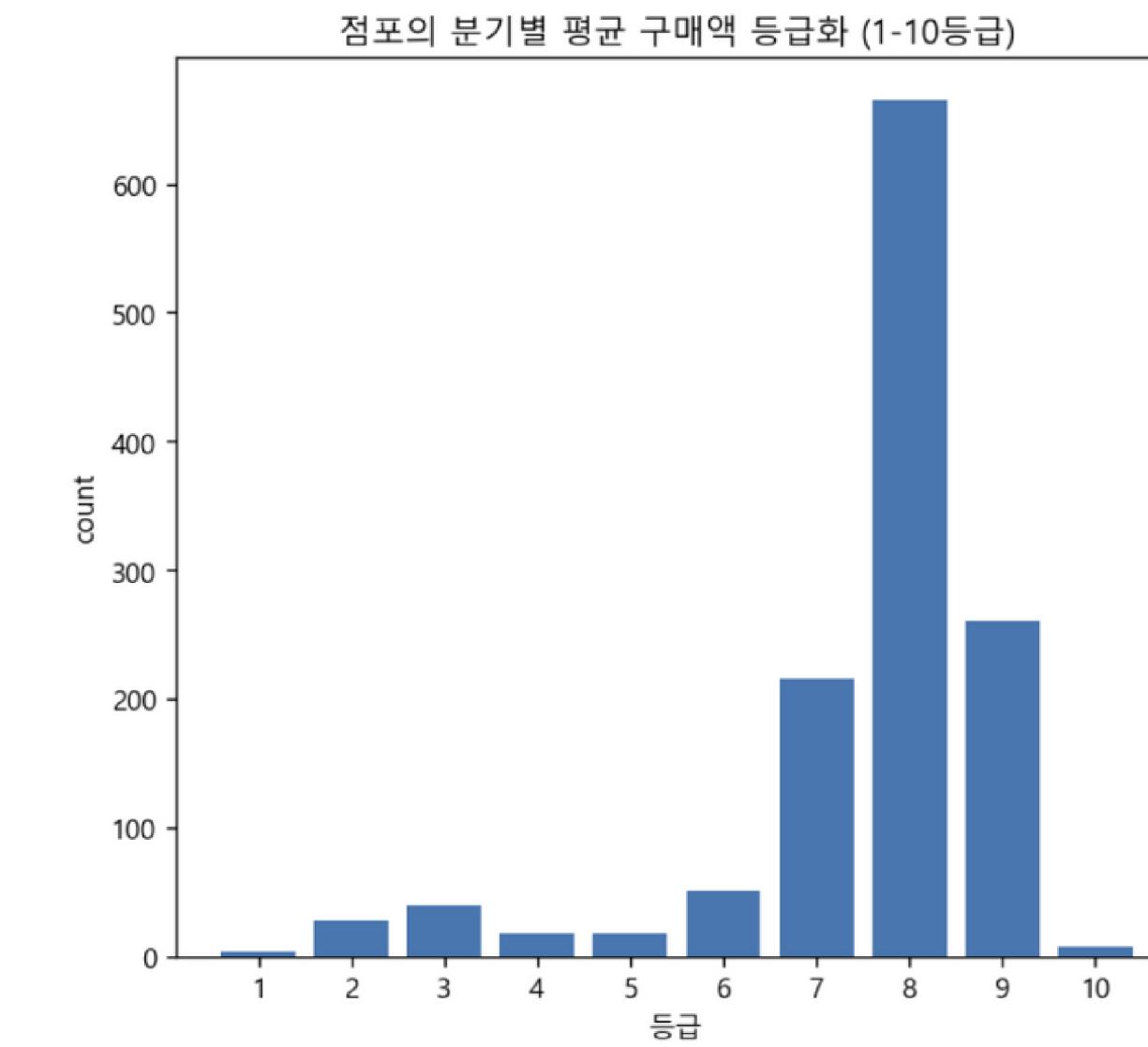


등급화 진행

```
CREATE TABLE AVG_PUR_CUST
AS SELECT A.점포코드, C.RANK - A.RANK AS X_TRAIN, D.RANK - B.RANK X_VAL_TEST
FROM AVG_PUR_RANK A, (SELECT * FROM AVG_PUR_RANK WHERE 분기 = '2nd') B,
(SELECT * FROM AVG_PUR_RANK WHERE 분기 = '6th') C, (SELECT * FROM AVG_PUR_RANK WHERE 분기 = '7th') D
WHERE A.분기 = '1st' AND A.점포코드 = B.점포코드 AND B.점포코드 = C.점포코드 AND C.점포코드 = D.점포코드
ORDER BY A.점포코드;
```

	점포코드	분기	평균구매액
0	A1	1st	142540
1	A1	2nd	136499
2	A1	3rd	121924
3	A1	4th	164597
4	A1	5th	140180
...
1299	D8	4th	6163
1300	D8	5th	7667
1301	D8	6th	8354
1302	D8	7th	10135
1303	D8	8th	6665

	등급	count
1	3	
2	28	
3	40	
4	17	
5	17	
6	51	
7	216	
8	665	
9	260	
10	7	



3

모델링 및 분석

분류 모델 비교 및 모델 선정

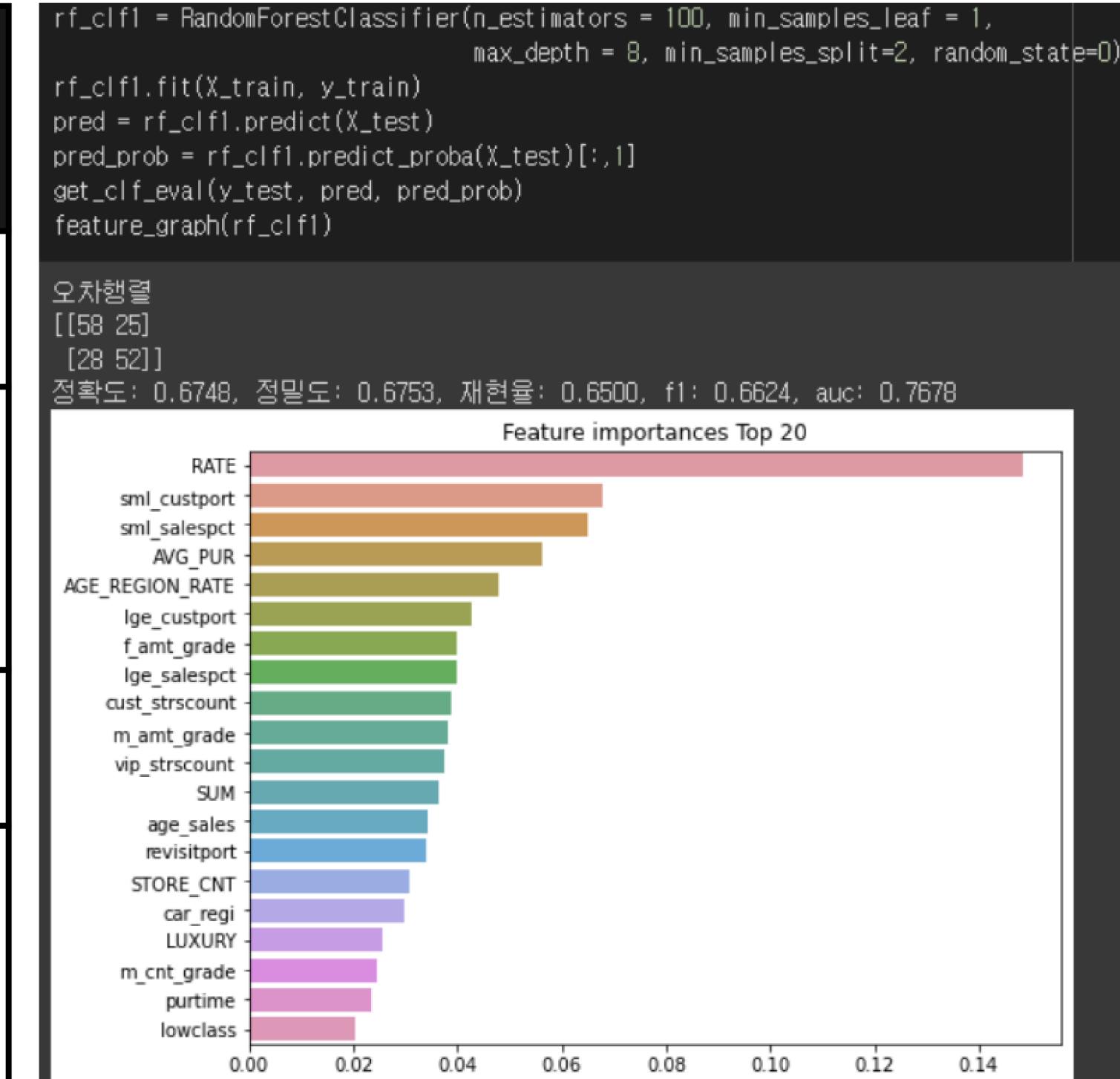


Pycaret으로 분류 모델 TOP3 추린 후, 평가 지표 비교

Evaluation	Random Forest	Logistic Regression	Light GBM
Accuracy	0.6939	0.7301	0.6258
F1	0.7692	0.6716	0.6667
Precision	0.7143	0.8333	0.5922
Recall	0.8333	0.5625	0.7625



가장 성능 좋다고 판단한 Random Forest로 선정 후
feature importances 확인



분류 결과

[매출 감소 예측 점포]

	점포코드	y_y	predict
0	A1	1	0
1	A12	1	0
2	A2	1	0
3	A20	1	0
4	A24	1	0
...
81	D12	1	0
82	D2	0	0
83	D4	0	0
84	D5	0	0
85	D7	0	0

86 rows × 3 columns



[매출 증가 예측 점포]

	점포코드	y_y	predict
0	B11	1	1
1	B3	1	1
2	B4	0	1
3	B6	0	1
4	B9	0	1
...
72	D10	1	1
73	D11	1	1
74	D3	1	1
75	D6	1	1
76	D8	0	1

77 rows × 3 columns



매출 감소 예측 점포 수 : 86
매출 증가 예측 점포 수 : 77



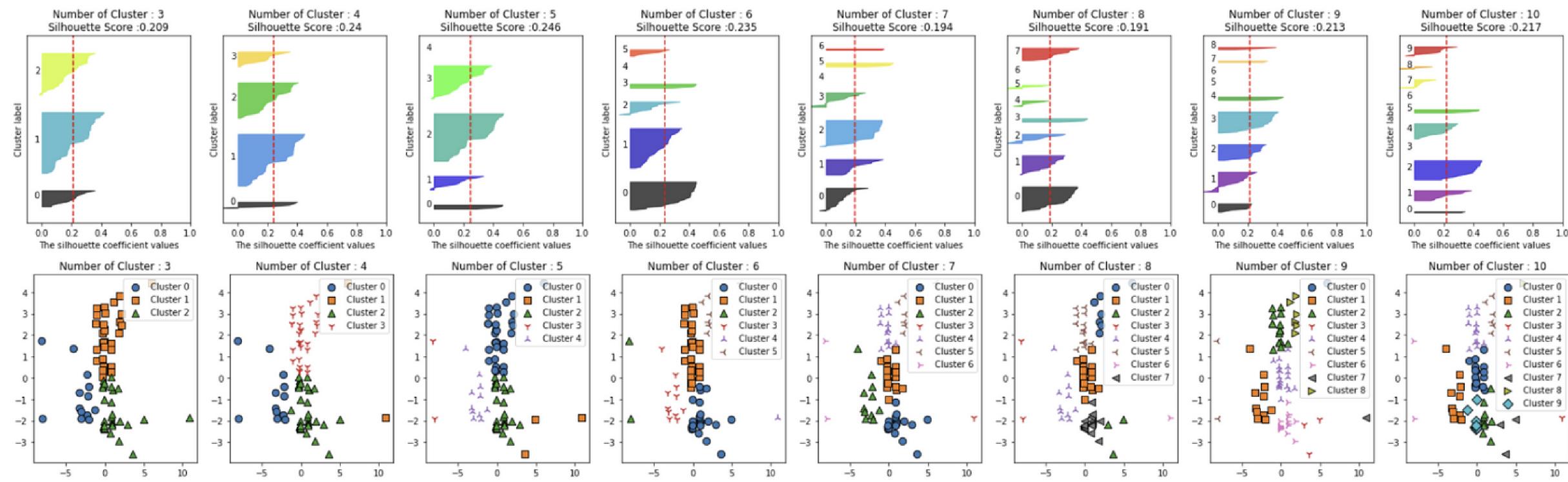
앞서 피처 중요도로 확인한
TOP 10 피처와 지역 관련
외부 데이터를 사용해
매출 증가 예측 점포 및
매출 감소 예측 점포별
군집화 진행

군집화

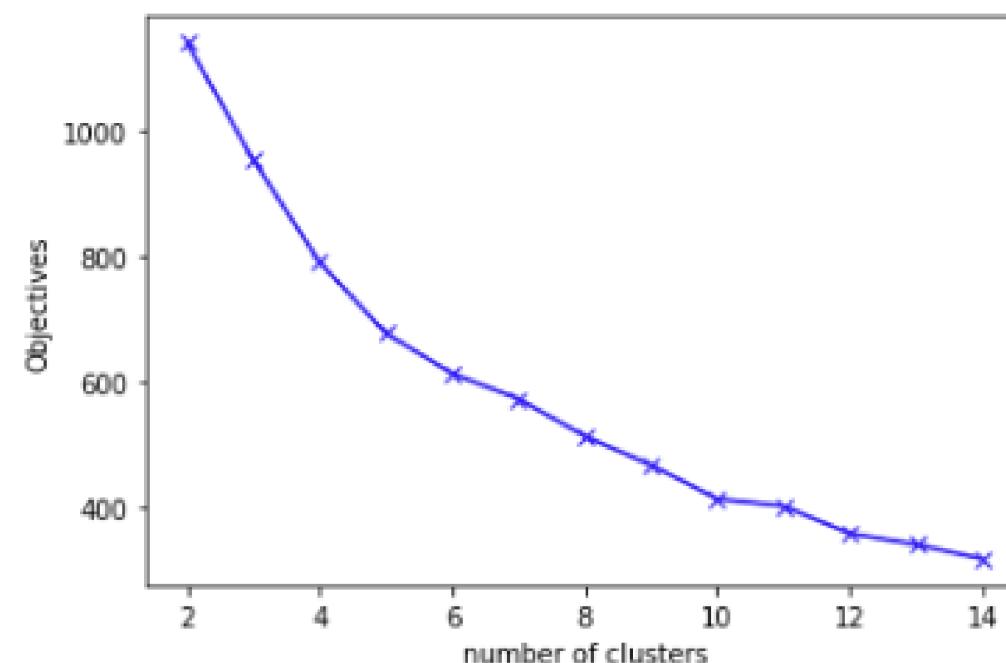
• 매출 증가 예측 점포 (77개)



실루엣 계수



elbow 기법



군집 개수 : 47개

77 rows × 12 columns

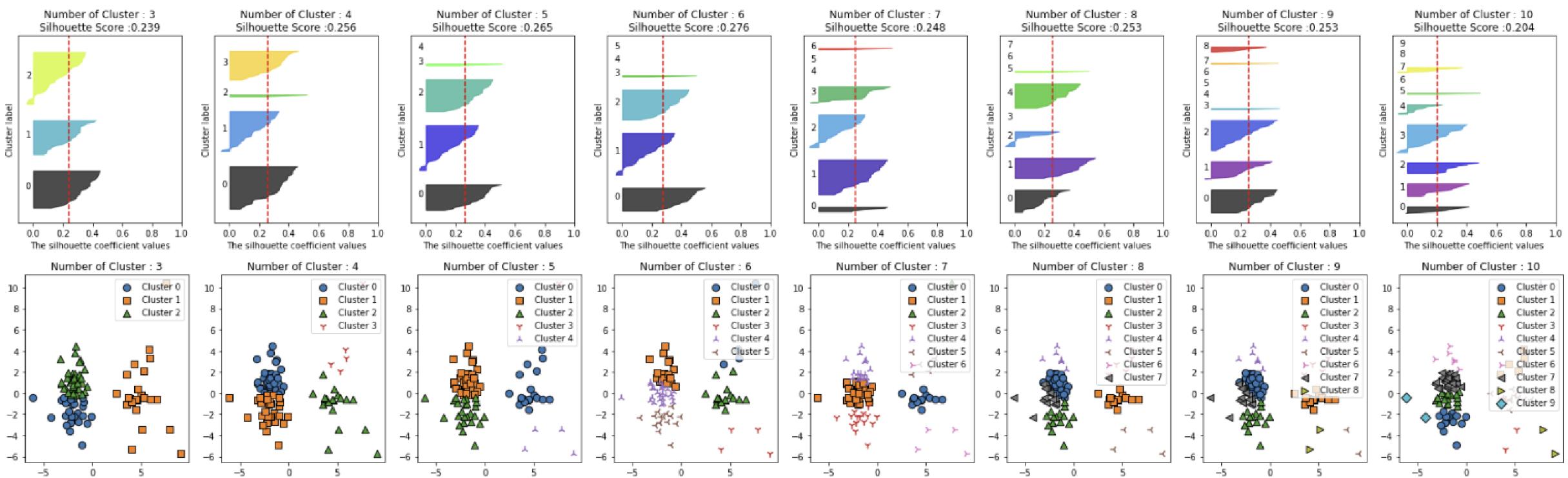
점포코드	시군구	toddler_grade	property_mean_grade	elderly_rate	1인가구수	AVG_PUR	LUXURY	ESSENTIAL	AGE_REGION_RATE	f_amt_grade	cluster_labels
0	B11	동대문구	4		4	2	2	-1	1.0	-1.0	1
1	C343	동대문구	4		4	2	2	0	0.0	0.0	9
2	C573	동대문구	4		4	2	2	-2	0.0	0.0	2
3	B3	영등포구	3		2	4	3	0	-2.0	0.0	1
4	C14	영등포구	3		2	4	3	-2	-3.0	-3.0	2
...
72	C242	동작구	3		4	3	2	0	0.0	0.0	3
73	C303	양천구	2		4	5	4	0	0.0	0.0	2
74	C402	용산구	4		4	2	4	0	0.0	0.0	2
75	C540	종로구	5		2	1	5	0	2.0	0.0	2
76	C95	성동구	4		3	3	4	-1	-4.0	0.0	1

군집화

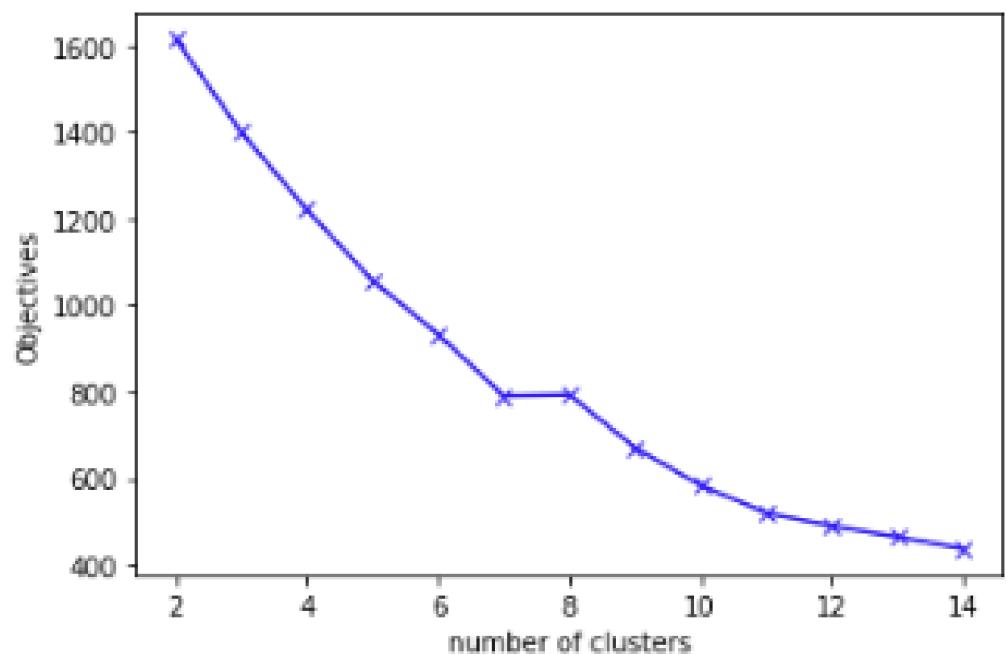
• 매출 감소 예측 점포 (86개)



실루엣 계수



elbow 기법



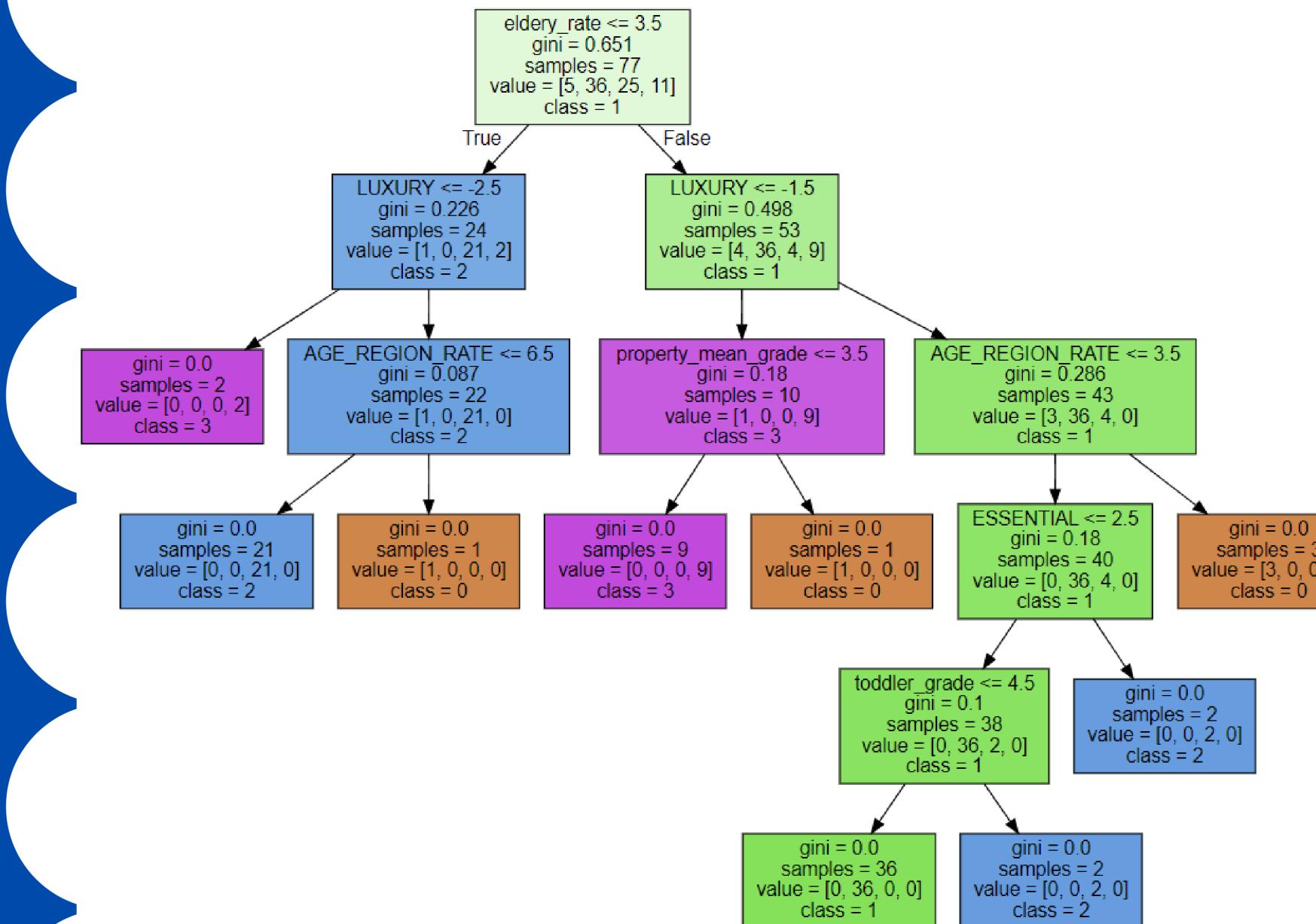
군집 개수 : 4개

점포코드	시군구	toddler_grade	property_mean_grade	elderly_rate	1인가구수	AVG_PUR	LUXURY	ESSENTIAL	AGE_REGION_RATE	f_ant_grade	cluster_label
0	A1	서대문구	4	4	2	3	0	-1.0	0.0	9	0
1	C331	서대문구	4	4	2	3	0	-1.0	-1.0	9	0
2	C475	서대문구	4	4	2	3	0	0.0	0.0	9	1
3	A12	강남구	2	1	5	2	0	0.0	-1.0	10	0
4	C317	강남구	2	1	5	2	0	-1.0	3.0	9	1
...
81	C330	은평구	2	5	2	3	0	2.0	1.0	2	0
82	C34	은평구	2	5	2	3	0	2.0	0.0	9	0
83	C342	도봉구	4	5	2	4	0	-1.0	-1.0	9	0
84	C406	도봉구	4	5	2	4	0	0.0	0.0	3	0
85	C532	도봉구	4	5	2	4	1	4.0	1.0	2	0

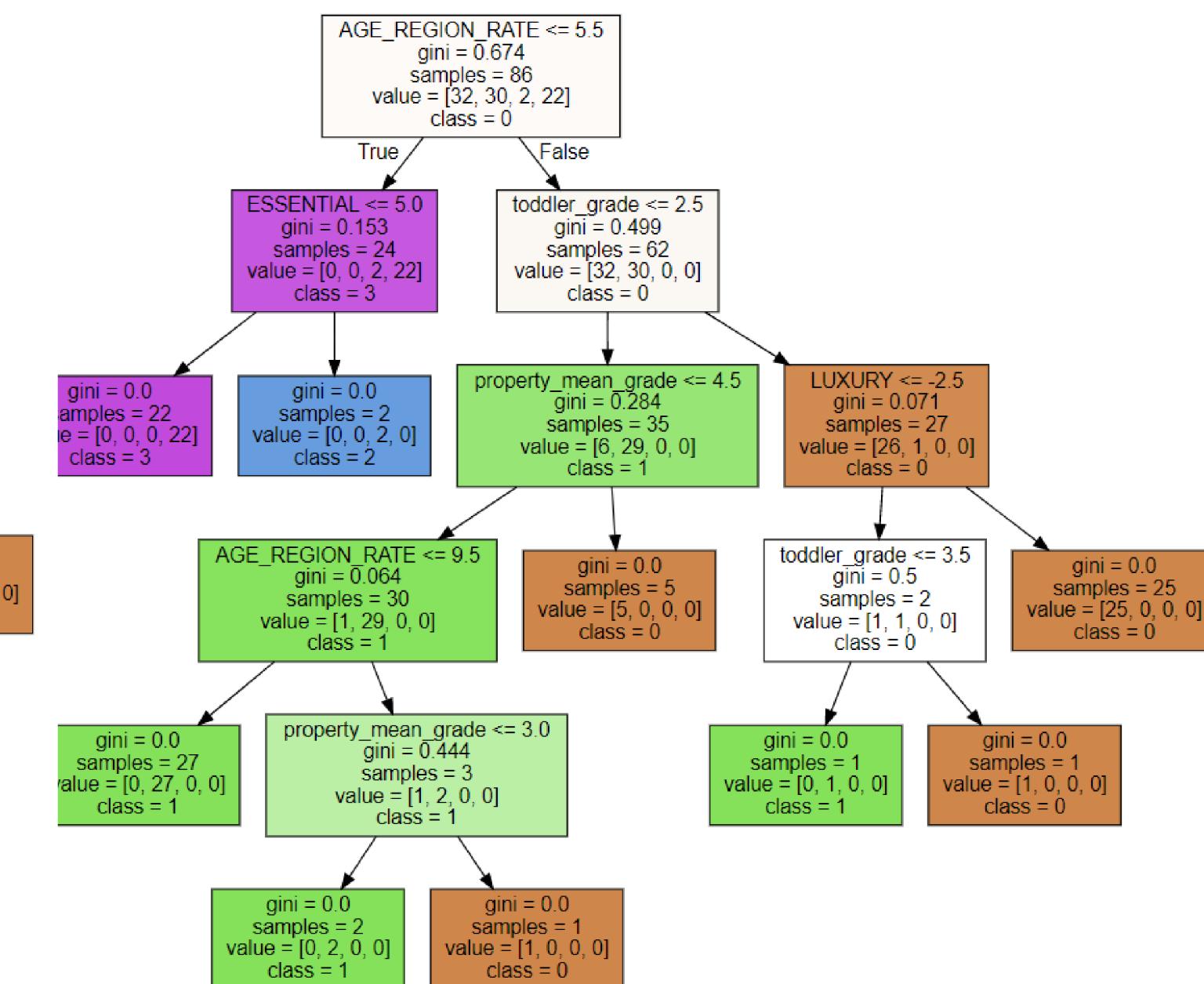
86 rows × 12 columns

Decision Tree

매출 증가 예측 점포



매출 감소 예측 점포



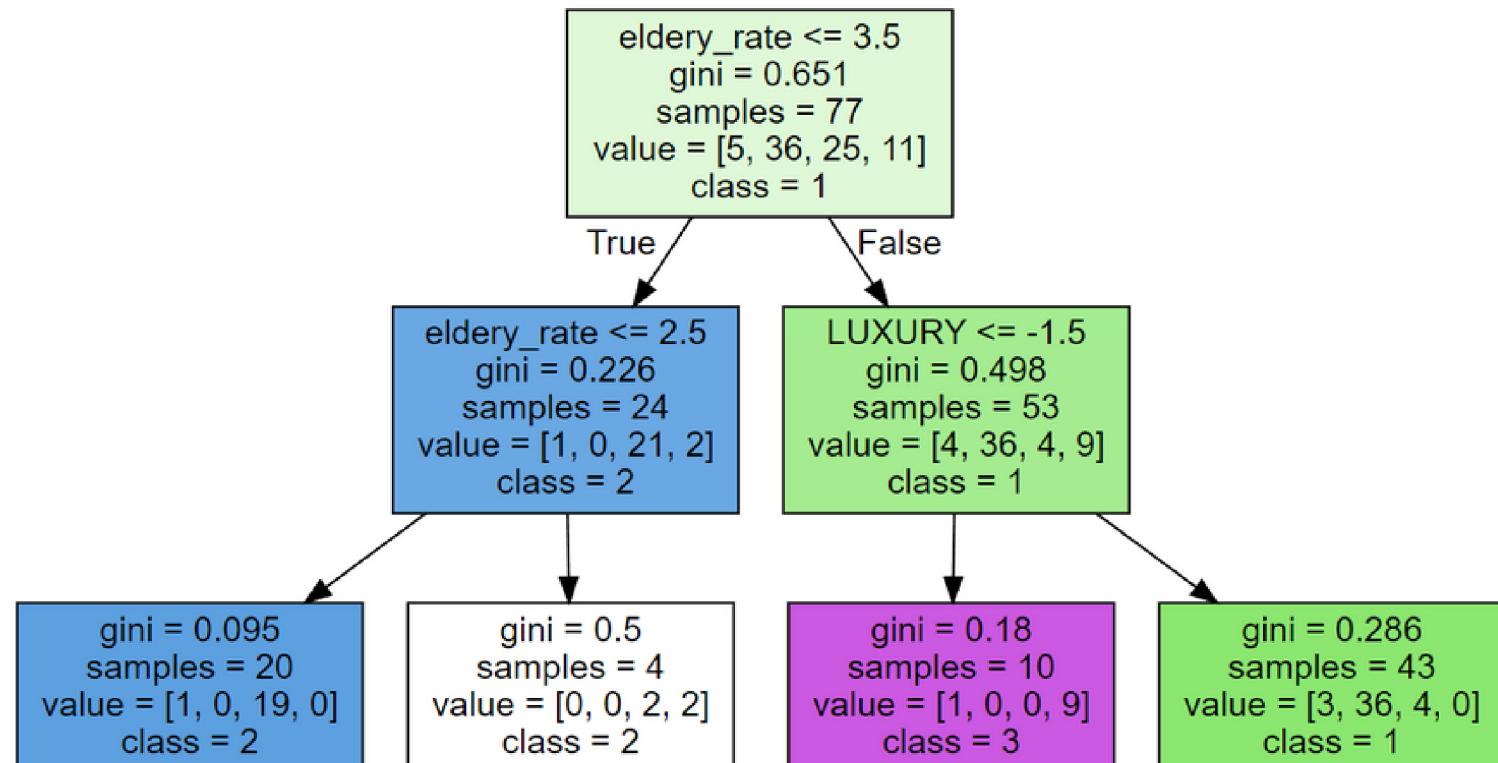
GridSearchCV

매출 증가 예측 점포

```
from sklearn.model_selection import GridSearchCV
param={
    'min_samples_leaf':[4,5,6,7,8,9],
    'max_depth':[2,3,4,5,6,None],
    'min_samples_split':[5,6,7,8,9,10,15]
}
dt_clf_cv = DecisionTreeClassifier()
gs = GridSearchCV(dt_clf_cv, param_grid= param, cv = 5, refit=True)

gs.fit(X_increase, y_increase)
dt_clf_cv = gs.best_estimator_
print("train_score : {}".format(dt_clf_cv.score(X_increase, y_increase)))
#print("test_score : {}".format(dt_clf_cv.score(X_increase_k5, y_increase_k5)))
print(gs.best_score_)
print(gs.best_params_)

train_score : 0.8571428571428571
0.8433333333333334
{'max_depth': 2, 'min_samples_leaf': 4, 'min_samples_split': 5}
```

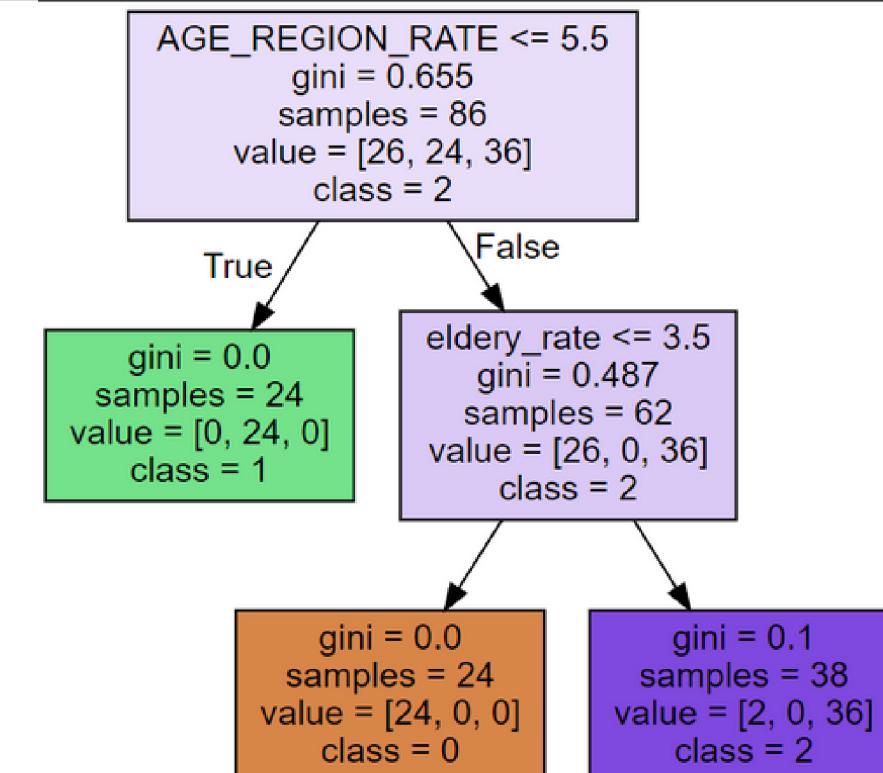


매출 감소 예측 점포

```
from sklearn.model_selection import GridSearchCV
param={
    'min_samples_leaf':[4,5,6,7,8,9],
    'max_depth':[2,3,4,5,6,None],
    'min_samples_split':[5,6,7,8,9,10,15]
}
dt_clf_cv = DecisionTreeClassifier()
gs = GridSearchCV(dt_clf_cv, param_grid= param, cv = 5, refit=True)

gs.fit(X_decrease, y_decrease)
dt_clf_cv = gs.best_estimator_
print("train_score : {}".format(dt_clf_cv.score(X_decrease, y_decrease)))
#print("test_score : {}".format(dt_clf_cv.score(X_decrease_k5, y_decrease_k5)))
print(gs.best_score_)
print(gs.best_params_)

train_score : 0.9767441860465116
0.9764705882352942
{'max_depth': 2, 'min_samples_leaf': 4, 'min_samples_split': 5}
```



분석내용

군집별 특징 분석

Increase 점포

0

주 고객층
=
해당 구의
최다 거주 연령층

1

필수재 구입이
증가하는 경향

2

노령층 많음

3

사치재 구입이
증가하는 경향

Decrease 점포

0

사치재 구입이
거의 동일하게
유지되는 경향

1

영유아 많음

2

필수재 구입이
줄어드는 경향

3

주 고객층
≠
해당 구의
최다 거주 연령층

분석 솔루션 _점포별 상품 추천(인구 특성 군집)

1 비교 가능한 점포 있는지 확인(Increase 0 / Decrease 3)

```
def product_recommendation_system(x):
    if final_df.query(f"점포코드 == '{x}'")['I_D'].unique()[0] == 'I':
        print(f'입력된 점포코드 : {x} \nSTATUS : I \n')

    else:
        print(f'입력된 점포코드 : {x} \nSTATUS : D \n')

    # decrease 3과 increase 0
    if final_df.query(f"점포코드 == '{x}'")['cluster_labels'].unique()[0] == 3:
        d_sigungu_3 = final_df.query(f"점포코드 == '{x}'").시군구.unique()[0]

    # 비교 가능한 구 있는지 확인
    if final_df.query(f"I_D == 'I' & cluster_labels == 0 & 시군구 == '{d_sigungu_3}'").empty == False:
        in_df = final_df.query(f"I_D == 'I' & cluster_labels == 0 & 시군구 == '{d_sigungu_3}'")

    # 비교할 점포 리스트 뽑아보기
    c_store_list = in_df['점포코드'].unique()
    print(f'같은 지역구 내 매출 증가 점포({x}) : ', len(c_store_list))
    print(" ")
    print(c_store_list, '\n')

    # 같은 제휴사로만 처리기
    compare_list = []

    for i in c_store_list:
        if x[0] == i[0]:
            compare_list.append(i)
        else:
            continue

    if len(compare_list) == 0:
        print('같은 지역구 내 매출이 증가한 점포 중 동일 제휴사 점포가 없습니다.\n')
    else:
        print('동일 제휴사 점포: ', compare_list)
        print('\n')
```

입력 점포의 매출이
증가했는지 감소했는지
살펴보기

감소점포일 때,
입력 점포와 같은 지역구인
점포 중 매출 증가 점포 유무 확인

해당 점포의 제휴사가
입력 점포와 동일한지 확인

분석 솔루션 _점포별 상품 추천(인구 특성 군집)

2 비교 가능한 점포의 상품 분류 비교

```
# 추려진 점포 리스트의 중분류명과 대상 점포의 중분류명 순위 비교
i_midclass = final_df.query(f"점포코드 == {compare_list}").groupby(['중분류명_재'])[['구매금액']].sum().reset_index().sort_values('구매금액', ascending = False)[:5]
x_midclass = final_df.query(f"점포코드 == '{x}'").groupby(['중분류명_재'])[['구매금액']].sum().reset_index().sort_values('구매금액', ascending = False)[:5]

# 관리 필요한 중분류명 리스트 뽑기
weak_midclass = []

for i_class in i_midclass['중분류명_재'].tolist():
    if i_class not in x_midclass['중분류명_재'].tolist():
        weak_midclass.append(i_class)
    else:
        continue

print('관리 필요한 중분류명 리스트 :', weak_midclass)
print('\n')

# 관리 필요한 중분류명에서 추천할 상품명 추출하기
product_recommend = []

for midclass in weak_midclass:
    i_products = final_df.query(f"점포코드 == {compare_list} & 중분류명_재 == '{midclass}'").groupby(['중분류명_재', '소분류명'])[['구매금액']].sum().reset_index().sort_values('구매금액', ascending = False)
    x_products = final_df.query(f"점포코드 == '{x}' & 중분류명_재 == '{midclass}'").groupby(['중분류명_재', '소분류명'])[['구매금액']].sum().reset_index().sort_values('구매금액', ascending = False)

    product_li = []
    midclass_dict = {}

    for pdt in i_products['소분류명'].tolist():
        if pdt not in x_products['소분류명'].tolist():
            product_li.append(pdt)
        else:
            continue

    if product_li:
        midclass_dict[f'{midclass}'] = product_li[:10]
    else:
        midclass_dict[f'{midclass}'] = '해당 중분류 내 품목은 적절하나 중분류 차원의 관리가 필요합니다.'

    product_recommend.append(midclass_dict)

print('추천 상품 리스트 : ')
for i in range(len(weak_midclass)):
    print(product_recommend[i])
```

2-1. 매출 증가 점포의 중분류 TOP10 중 입력 점포 TOP10에 없는 중분류를 관리가 필요한 대상으로 설정

2-2. 관리 필요 중분류에서 매출 증가 점포와 비교 했을 때 입력 점포에 없는 상품 소분류 중 최대 10개 상품 추천

분석 솔루션 _점포별 상품 추천(인구 특성 군집)

3 상품 추천 리스트 추출

• 비교 가능 매장 있는 입력 점포

```
product_recommendation_system('C142')
```

입력된 점포코드 : C142
STATUS : D

같은 지역구 내 매출 증가 점포(C142) : 1

['C74']

동일 제휴사 점포: ['C74']

관리 필요한 중분류명 리스트 : ['생활용품', '주방가전', '수산물', '미용용품']

추천 상품 리스트 :

```
{ '생활용품': [ '샴푸/린스세트', '액상세탁세제', '구강청정제', '바디워시', '액상표백제', '특수용세탁세제', '세탁비누', '부분세탁제', '장갑류', '접착제' ] }  
{ '주방가전': [ '인덕션/가스레인지' ] }  
{ '수산물': [ '기타어류', '갈치', '랍스터', '기타포장생선', '기타해초류', '새우', '전복', '톳', '가자미', '파래' ] }  
{ '미용용품': [ '페이셜클렌저', '헤어에센스', '헤어왁스', '풋케어', '면봉/화장솜', '크림/밤/오일' ] }
```

• 비교 가능한 매장 없는 입력 점포

```
product_recommendation_system('D1')
```

입력된 점포코드 : D1
STATUS : D

같은 지역구 내 매출 증가 점포(D1) : 1

['C412']

같은 지역구 내 매출이 증가한 점포 중 동일 제휴사 점포가 없습니다.

분석 솔루션 _점포별 상품 추천(인구 특성 군집)

3

상품 추천 리스트 추출

- 관리 필요 중분류 내에 상품 품목이 증가 점포와 동일한 점포

```
product_recommendation_system('C308')
```

입력된 점포코드 : C308

STATUS : D

같은 지역구 내 매출 증가 점포(C308) : 1

['C412']

동일 제휴사 점포: ['C412']

관리 필요한 중분류명 리스트 : ['생활용품', '조미료', '음료류', '건강식품']

추천 상품 리스트 :

```
{'생활용품': '해당 중분류 내 품목은 적절하나 중분류 차원의 관리가 필요합니다.'}  
{'조미료': '해당 중분류 내 품목은 적절하나 중분류 차원의 관리가 필요합니다.'}  
{'음료류': '해당 중분류 내 품목은 적절하나 중분류 차원의 관리가 필요합니다.'}  
{'건강식품': ['일반비타민']}
```



1. 소분류(품목) 재고 관리가 부족한 경우

- 재고가 고객들의 수요보다 부족한지 확인
- 상품 DP가 적절한지 확인

2. 다른 외부 요인으로 인한 매출 감소

- 외부 요인 파악 후 관련 솔루션 도출

분석 솔루션2 _점포별 상품 추천(상품 특성 군집)

1 감소 점포 중 상품 특성으로 군집된 점포 확인

	점포코드	시군구	toddler_grade	property_mean_grade	eldery_rate	1인가구수	AVG_PUR	LUXURY	ESSENTIAL	AGE_REGION_RATE	f_amt_grade	cluster_label	2
0	B10	구로구	2	3	4	3	5	-1.0	12.0	1	0	2	
1	B13	송파구	1	2	5	2	2	1.0	7.0	1	0	2	

필수재 구입이
줄어드는 경향

2 확인한 점포 벡터화 ----- increase 점포 전체 벡터화

```
dec_B10 = np.array(dec_cluster2.iloc[0, 2:])
dec_B10

array([2, 3, 4, 3, 5, -1.0, 12.0, 1, 0, 2], dtype=object)

dec_B13 = np.array(dec_cluster2.iloc[1, 2:])
dec_B13

array([1, 2, 5, 2, 2, 1.0, 7.0, 1, 0, 2], dtype=object)
```

```
test = []

for i in range(0, 77):
    a = np.array(increase_cluster.iloc[i , 2:])
    test.append(a)

test
...
len(test)
77
```

3 코사인 유사도 높은 점포 확인

```
for i in range(0, 77):
    print(cosine_similarity([test[i].reshape(-1, )], [dec_B13]))
...

jj=[]
for i in range(0, 77):
    j = cosine_similarity([test[i].reshape(-1, )], [dec_B13])
    jj.append(j)

jj.index(max(jj))

52
jj[52] # C478
array([[0.88900089]])
```

```
# B10과 코사인유사도 높은 순 increase 점포
increase_cluster.iloc[[52, 4, 11], :2]
```

점포코드	시군구
52	C478 광진구
4	B9 서대문구
11	C127 노원구

같은 군집화 유형으로
매출이 감소한 점포라서 그런지
감소 점포 두 곳과 코사인유사도
TOP3 모두 같은 구

```
# B13과 코사인유사도 높은 순 increase 점포
increase_cluster.iloc[[52, 4, 11, 70, 57, 6, 74], :2]
```

점포코드	시군구
52	C478 광진구
4	B9 서대문구
11	C127 노원구
70	C96 강남구
57	C506 송파구
6	C109 송파구
74	D3 송파구

분석 솔루션2 _점포별 상품 추천(상품 특성 군집)

4

감소 점포(B10) 분석

점포	대분류명	구매건수	매출금액
B10	유아용품	1230	484854870
B10	가공식품	32461	315422806
B10	일상용품	6853	243564770
B10	신선식품	11094	89254681
B10	행사상품	732	69178018
B10	패션의류/잡화	1796	49341730
B10	교육/문화	1701	36912440
B10	스포츠/레저	569	17791470
B10	가구/인테리어	307	13433340
B10	기타	74	3745040
B10	디지털/가전	20	1376600

5

비슷한 증가 점포(C478, B9, C127) 분석

점포	대분류명	구매건수	매출금액	점포	대분류명	구매건수	매출금액	점포	대분류명	구매건수	매출금액
C478	가공식품	8840	26909110	B9	가공식품	108107	646883286	C127	가공식품	11606	40987661
C478	신선식품	2246	11765876	B9	일상용품	28535	589562800	C127	신선식품	7818	38443126
C478	일상용품	966	3529900	B9	신선식품	49388	364561245	C127	일상용품	2636	10926700
C478	디지털/가전	8	267660	B9	유아용품	2012	159414750	C127	행사상품	19	1143418
C478	유아용품	34	211960	B9	행사상품	2810	87278908	C127	패션의류/잡화	87	997840
C478	패션의류/잡화	31	164260	B9	스포츠/레저	908	28930824	C127	유아용품	114	923500
C478	행사상품	4	109700	B9	교육/문화	2202	23694088	C127	가구/인테리어	8	256780
C478	가구/인테리어	5	46780	B9	가구/인테리어	757	21466130	C127	디지털/가전	10	226500
C478	교육/문화	24	46130	B9	디지털/가전	20	1238090	C127	스포츠/레저	13	84320
C478	스포츠/레저	6	22200	B9	기타	88	1146050	C127	교육/문화	10	25420
				B9	명품	14	555000				

점포 소분류명

B10 분유

B10 봉지면

B10 방향탈취제

B10 감자스낵

B10 종합세트

점포 소분류명

C478 일반우유

C478 생수

C478 국산맥주

C478 바아이스크림

C478 콘아이스크림

VS

6 솔루션 제안

B10 점포는 영유아가 많은 지역에 위치해 있어 유아 관련 용품이 많이 팔리나, 영유아 수가 점차 줄어듦으로 인하여 필수재의 구입이 줄어드는 것으로 보임



영유아 관련 용품을 줄이고, 가공식품 중심의 상품 구색 갖추기 필요!

분석 솔루션2 _점포별 상품 추천(상품 특성 군집)

4 감소 점포(B13) 분석

총매출금액	
점포	중분류명
B13	농산물 572581889
	영유아식품 408308280
	의류 404165650
	축산물 402061209
	미용용품 258881050
	유제품/아이스크림 210524540
	농산가공 206734275
	커피/차류 170811030
	과자류 157420460
	수산물 156706712

5 비슷한 증가 TOP3 점포 (C478, B9, C127) 분석

총매출금액	
점포	중분류명
C478	농산물 7871360
	유제품/아이스크림 5475040
	가공식품_기타 3395080
	축산물 3241640
	음료류 3069750
	과자류 2350790
	수산가공 2326420
	빵류/떡류 2266270
	주류 2137800
	농산가공 1506380

총매출금액	
점포	중분류명
B9	미용용품 231186590
	농산물 175933191
	생활용품 163785570
	영유아식품 137507200
	축산물 118402926
	의류 90711430
	유제품/아이스크림 79724930
	미용용품 72673820
	과자류 69374670
	농산가공 65239944

총매출금액	
점포	중분류명
C127	농산물 24835185
	축산물 11369248
	유제품/아이스크림 6078330
	농산가공 5039257
	가공식품_기타 4218210
	수산가공 4196794
	조미료 3377900
	위생용품 3261900
	주류 3208660
	빵류/떡류 2779360

6 솔루션 제안

B13 점포는 비슷한 증가 점포와 비교하여 영유아식품, 의류, 커피/차류, 수산물의 판매 비중이 높은 것으로 확인됨



현재 잘 팔리는 상품과 비슷하지만
매출이 증가한 점포에서 잘 팔리는 시너지 좋은 상위권 상품을
눈에 잘 띠는 곳에 함께 진열하여 판매 증진!

예) 농산물 or 축산물 or 수산물 +주류, 커피/차류 + 빵류/떡류





결론

효과성

```
product_recommendation_system('C142')
```

입력된 점포코드 : C142

STATUS : D

같은 지역구 내 매출 증가 점포(C142) : 1

['C74']

동일 제휴사 점포: ['C74']

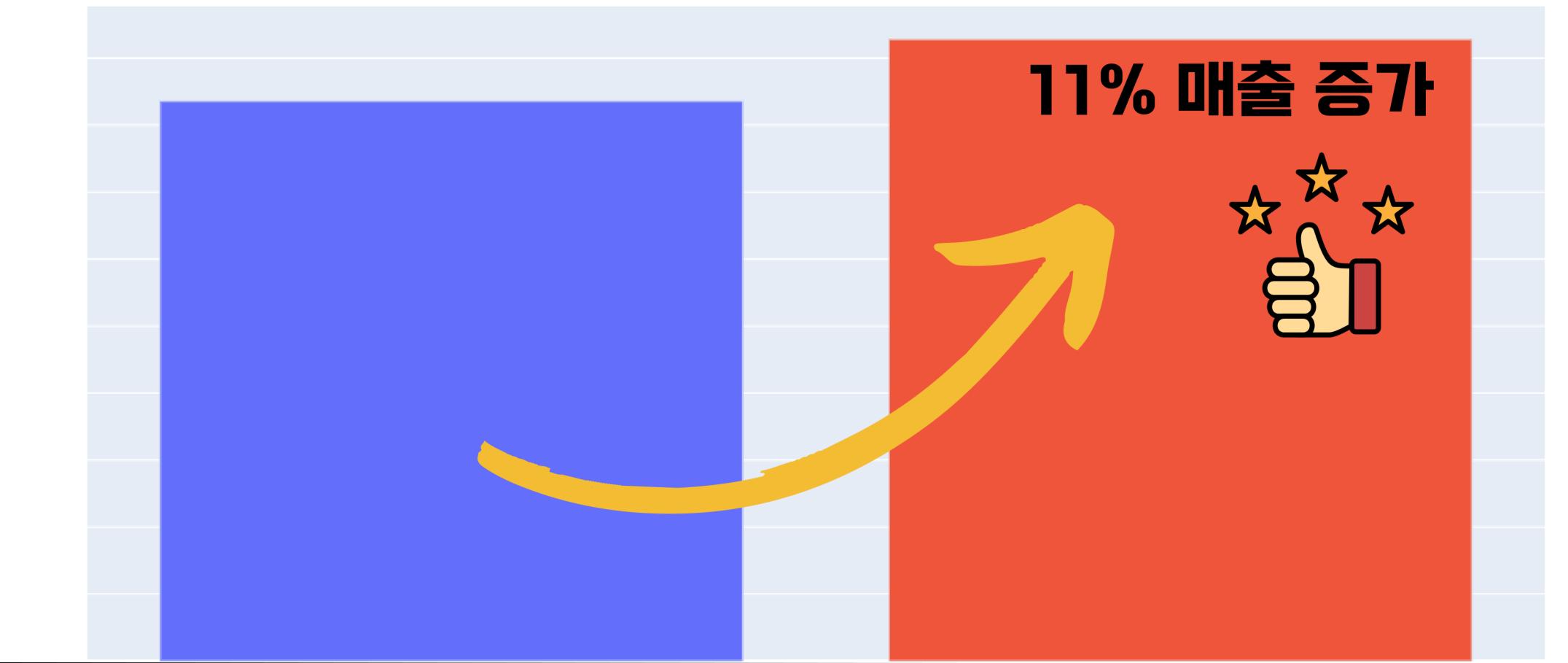
관리 필요한 중분류명 리스트 : ['생활용품', '주방가전', '수산물', '미용용품']

추천 상품 리스트 :

```
{'생활용품': ['샴푸/린스세트', '액상세탁세제', '구강청정제', '바디워시', '액상표백제', '특수용세탁세제', '세탁비누', '부분세탁제', '장갑류', '접착제']}  
{'주방가전': ['인덕션/가스레인지']}
```

```
{'수산물': ['기타어류', '갈치', '랍스터', '기타포장생선', '기타해초류', '새우', '전복', '톳', '가자미', '파래']}
```

```
{'미용용품': ['페이셜클렌저', '헤어에센스', '헤어왁스', '풋케어', '면봉/화장솜', '크림/밤/오일']}
```

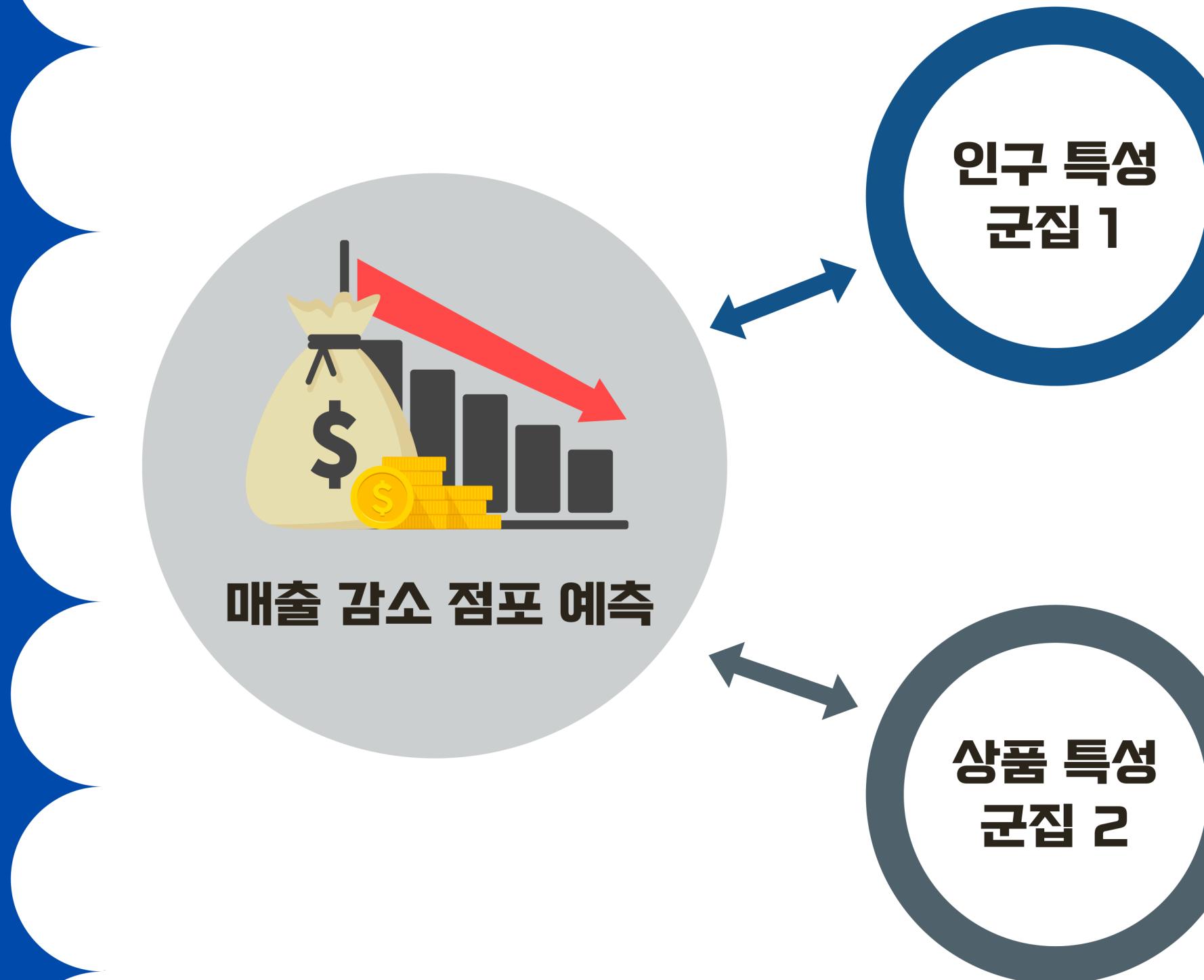


C142 솔루션 적용 전

C142 솔루션 적용 후

최종 제안

매출 감소 점포 군집 특성별 솔루션 제안



주 고객층이 해당 점포가 위치하는 구의 거주하고 있는
최다 연령층과 일치하지 않은 군집

1. 같은 자치구 내 매출이 증가한 점포 유무 확인
2. 비교할 점포가 있다면 비교 분석 후 상품 추천
3. 관리 필요 중분류 내에 상품 품목이 증가 점포와 동일한
점포라면,

- 가. 소분류(품목) 재고 관리가 부족한 경우
 - 재고가 고객들의 수요보다 부족한지 확인
 - 상품 DP가 적절한지 확인

- 나. 다른 외부 요인으로 인한 매출 감소
 - 외부 요인 파악 후 관련 솔루션 도출

상품의 특정 상위 카테고리 매출이 점차 증감하는
추세의 군집

1. 감소 점포와 증가 점포의 코사인 유사도 측정
2. 코사인유사도 높은 점포(TOP3)와 매출 분석
3. 매출 증감의 원인을 찾아 맞춤 솔루션 도출

한계점

한계점

- 점포 위치 추정 시 정확성 떨어짐
 - 구 단위로만 위치를 추정할 수 있었음
 - 구체적인 지역의 인구 특성 파악하기 어려움
- 위치와 매출 사이의 관계성이 떨어질 수 있음
 - 같은 구 내에서도 매출이 증가한 점포와 감소한 점포 존재
 - 구 단위가 크다보니 위치와 점포의 매출 사이의 관계성이 떨어짐
- 매출이 증가한 점포와 감소한 점포의 상품 구성의 큰 차이가 없는 경우도 있음
 - 매출 감소 점포에서는 중분류 순위가 낮지만 매출 증가 점포 취급 품목과 동일한 경우
 - 지역의 인구 특성만으로 점포 품목 관리 기준을 정하는 일이 어려울 수 있음
- 매출 감소의 원인 중에는 데이터에서 찾을 수 없는 불명확한 요인들이 많음
 - 예) 직원의 불친절한 서비스, 점포의 청결 상태 등

Q

A