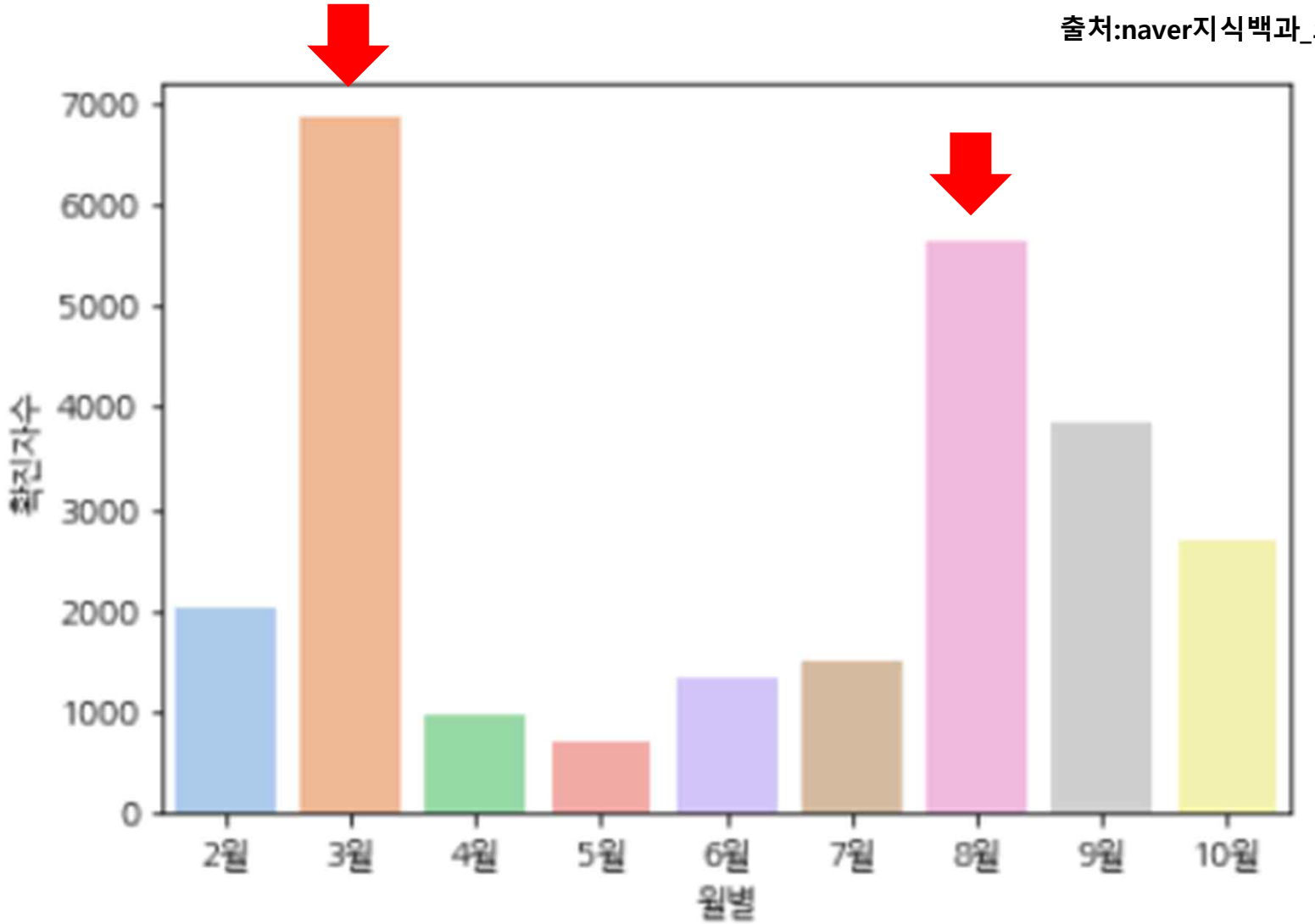
AI알고리즘 활용 카드 사용금액
및
소비 패턴 예측 모델 생성

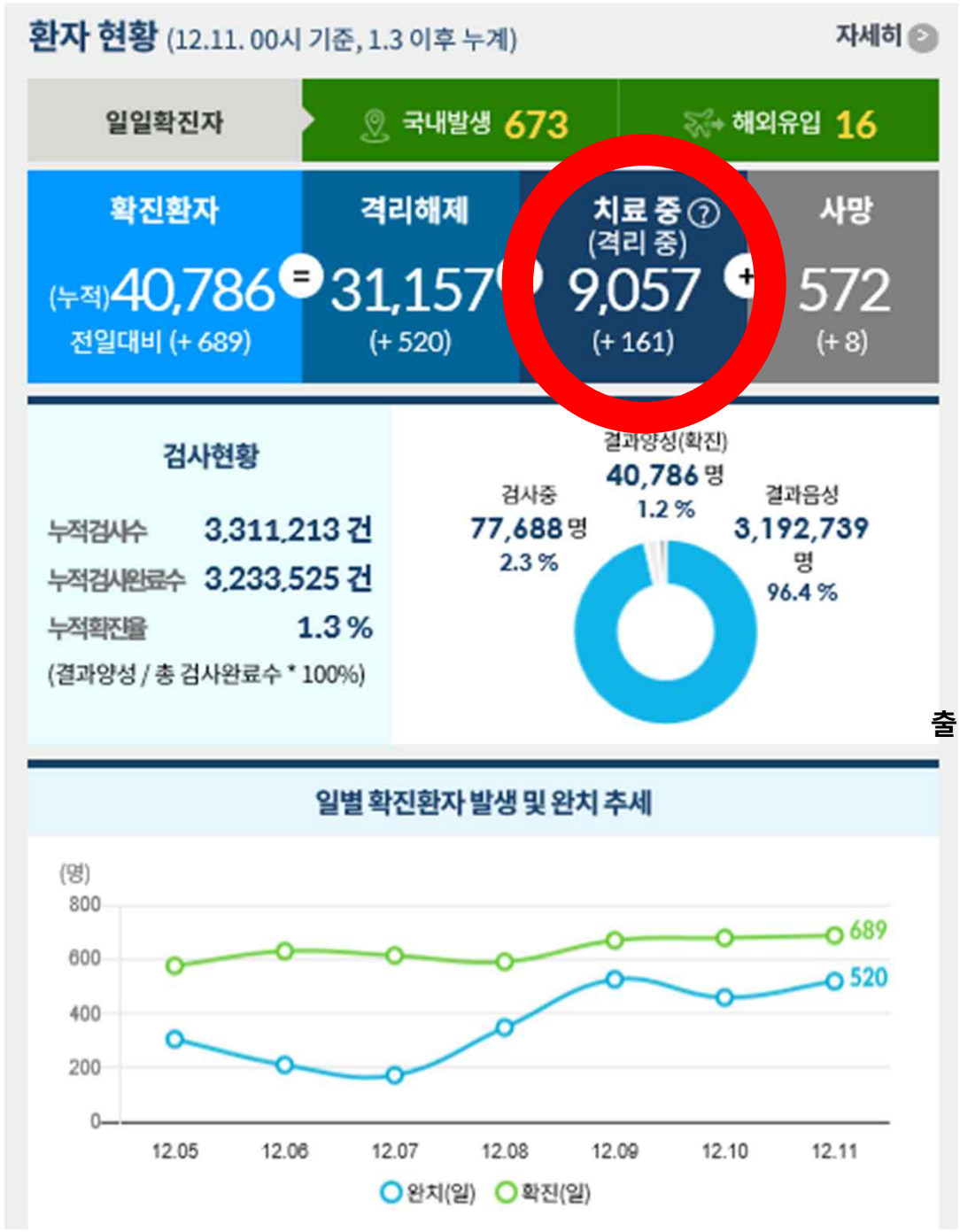
2017010688 김예지
2017010715 허지혜
2019010740 이수빈

목차

1.주제선정 동기 및 데이터 소개

출처:naver지식백과_코로나19





출처:http://ncov.mohw.go.kr/

코로나 자영업 한파, IMF 외환위기 때보다 더 춥다

통계청 2020년 사회동향 코로나 분석

음식·숙박

외환위기

女·20대·

가 가

최훈길 기자

최바울

"올해 음

충격이 19

이다. 학원 문을 닫다 보니 시간강사가 거리로 내몰렸다"며 "소득 격차는 벌어졌고 임시·일용직의 고용 충격이 컸다"고 지적했다.

심각한 것

화돼 양

카드 소비패턴 분석

지역별, 업종별 월간 카드 사용 총액을 예측하는 모델

예측 모델에 따른 지역 경제 위축 및 중소상공인들의 경영난 해소에 도움

Jeju 제주특별자치도
Jeju Special Self-Governing Province

월별 전국민 카드소비내역

	연월	카드이용_시도	카드이용_시군구	업종명	고객거주_시도	고객거주_시군구	연령대	성별	가구생애주기	이용고객수	이용금액	이용건수
0	201901	강원	강릉시	건강보조식품 소매업	강원	강릉시	20s	1	1	4	311200	4
1	201901	강원	강릉시	건강보조식품 소매업	강원	강릉시	30s	1	2	7	1374500	8

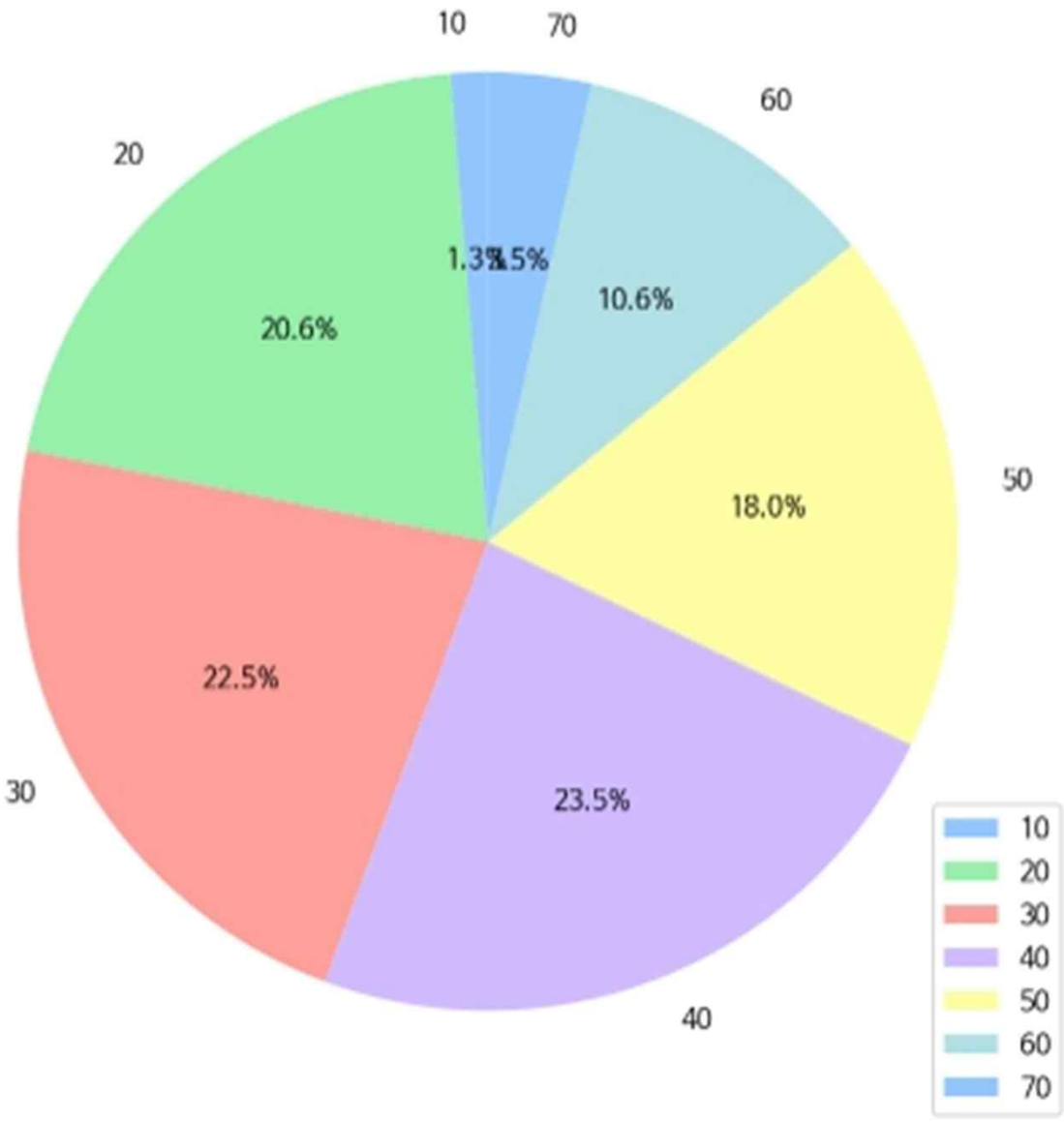
	연월	카드이용_시도	카드이용_시군구	업종명	고객거주_시도	고객거주_시군구	연령대	성별	가구생애주기	이용고객수	이용금액	이용건수
0	201901	강원	강릉시	건강보조식품 소매업	강원	강릉시	20s	1	1	4	311200	4
1	201901	강원	강릉시	건강보조식품 소매업	강원	강릉시	30s	1	2	7	1374500	8

	연월	카드이용_시도	업종명	고객거주_시도	연령대	성별	가구생애주기	이용고객수	이용금액	이용건수
0	201901	강원	건강보조식품 소매업	강원	20s	1	1	4	311200	4
1	201901	강원	건강보조식품 소매업	강원	30s	1	2	7	1374500	8

	연월	카드이용_시도	업종명	고객거주_시도	연령대	성별	가구생애주기	이용고객수	이용금액	이용건수	년	월
0	201901	강원	건강보조식품 소매업	강원	20	1	1	4	311200	4	2019	01
1	201901	강원	건강보조식품 소매업	강원	30	1	2	7	1374500	8	2019	01

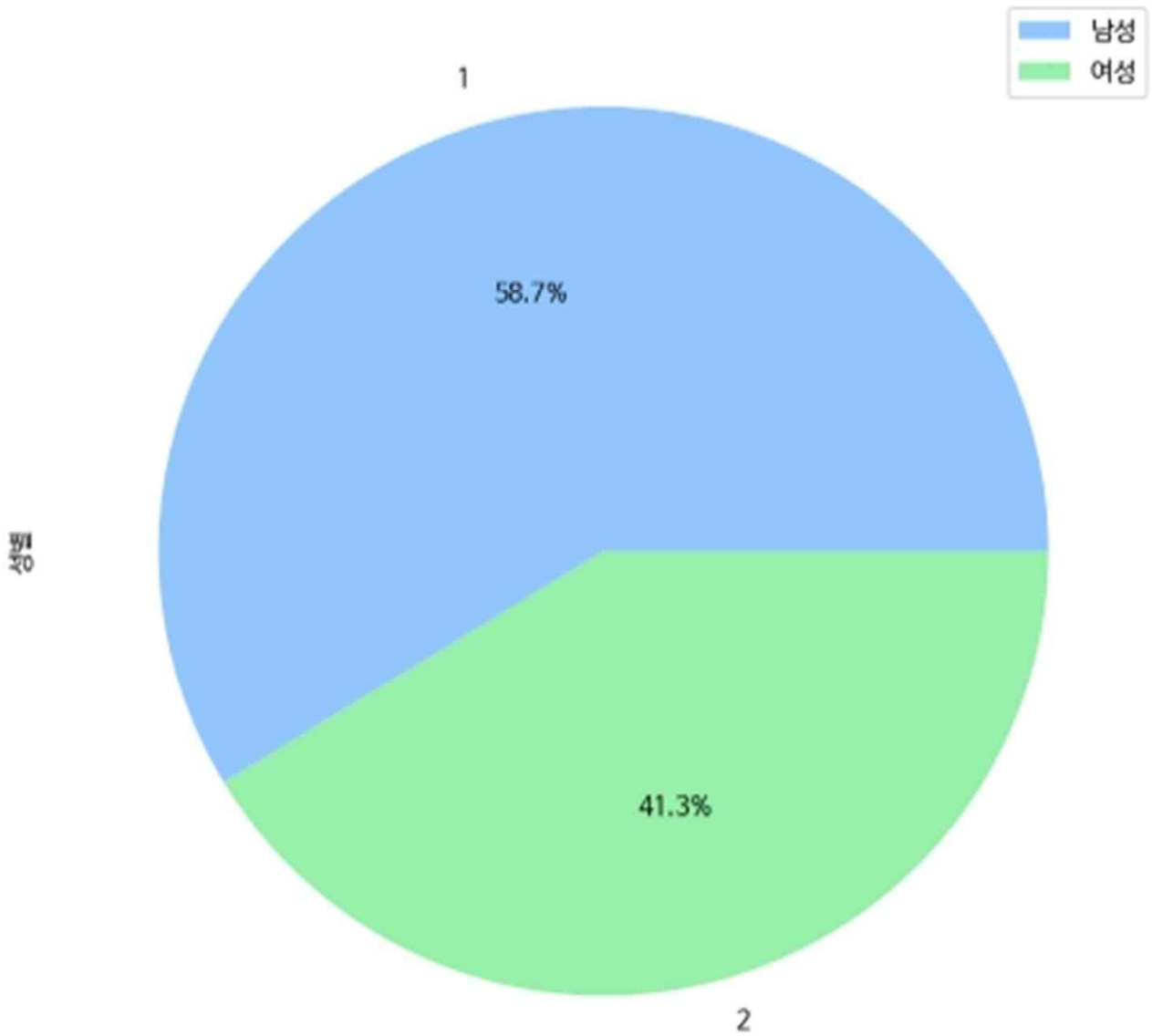
2-1) 연령대

	연령대
10	314674
20	5091675
30	5550519
40	5802447
50	4455687
60	2606168
70	876622



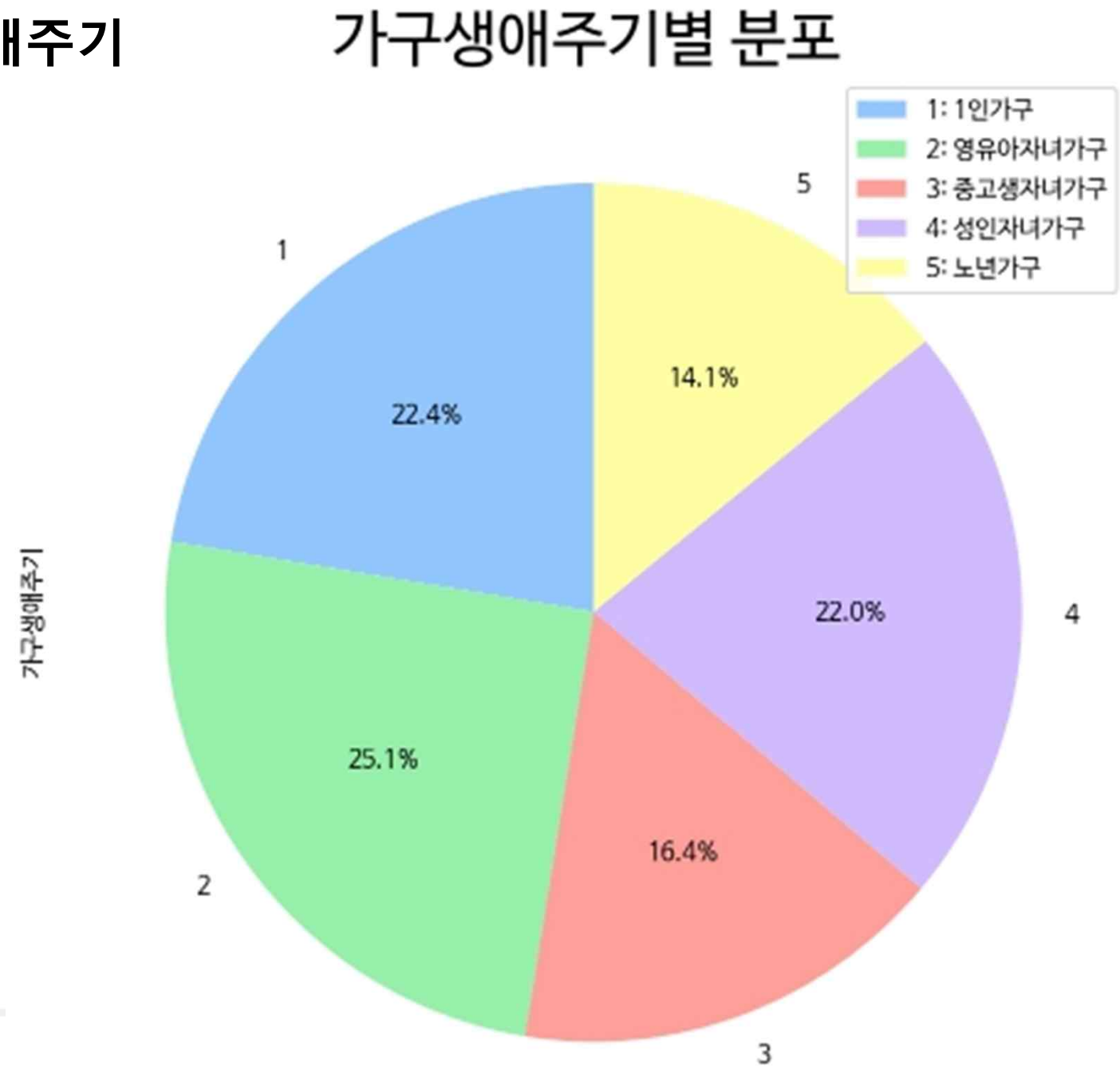
2-2) 성별

	성별
1	14506378
2	10191414



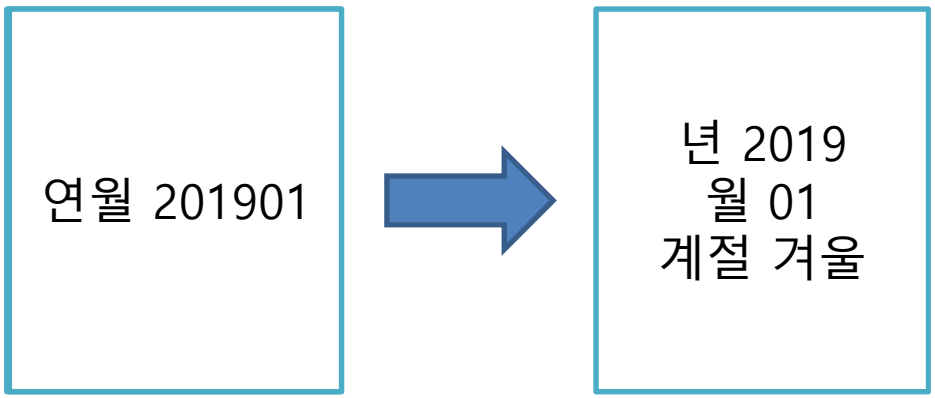
2-3) 가구생애주기

	가구생애주기
1	5526140
2	6188801
3	4054741
4	5445320
5	3482790



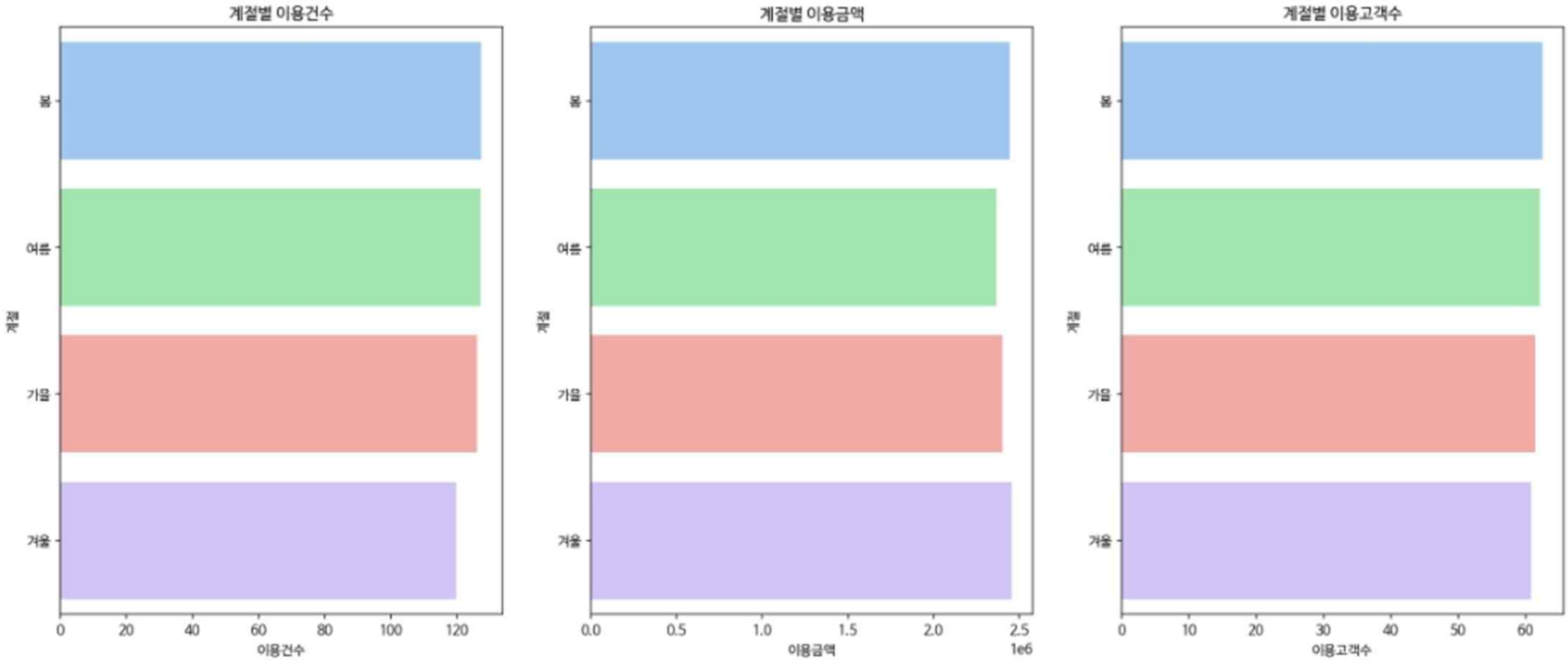
2-4) 연월->년, 월, 계절

	연월	카드이용_시도	업종명	고객거주_시도	연령대	성별	가구생애주기	이용고객수	이용금액	이용건수	년	월	계절
0	201901	강원	건강보조식품 소매업	강원	20	1	1	4	311200	4	2019	1	겨울
1	201901	강원	건강보조식품 소매업	강원	30	1	2	7	1374500	8	2019	1	겨울
2	201901	강원	건강보조식품 소매업	강원	30	2	2	6	818700	6	2019	1	겨울
3	201901	강원	건강보조식품 소매업	강원	40	1	3	4	1717000	5	2019	1	겨울
4	201901	강원	건강보조식품 소매업	강원	40	1	4	3	1047300	3	2019	1	겨울

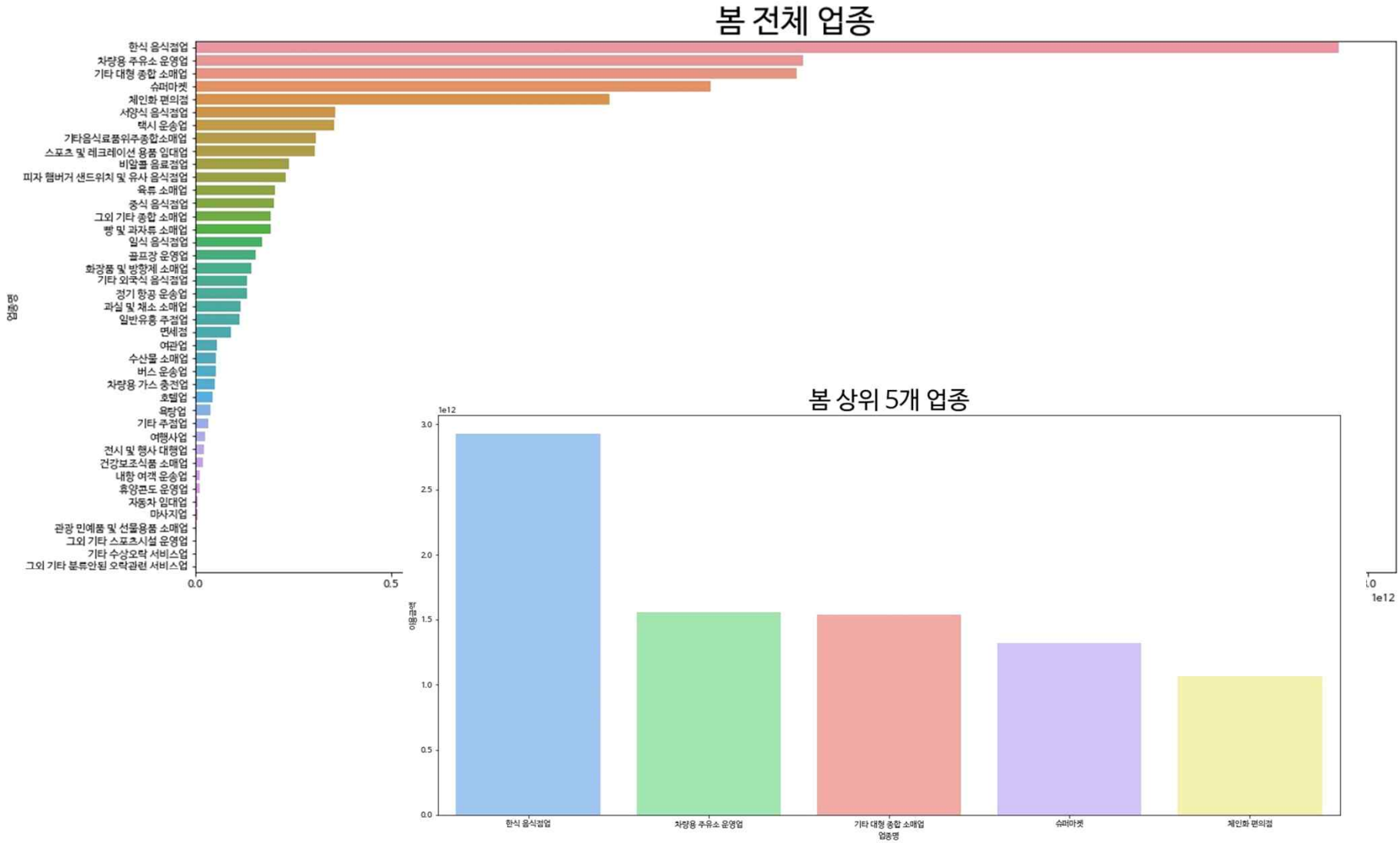


겨울 행의갯수: 4875679
봄 행의갯수: 5096567
여름 행의갯수: 5314345
가을 행의갯수: 5138824

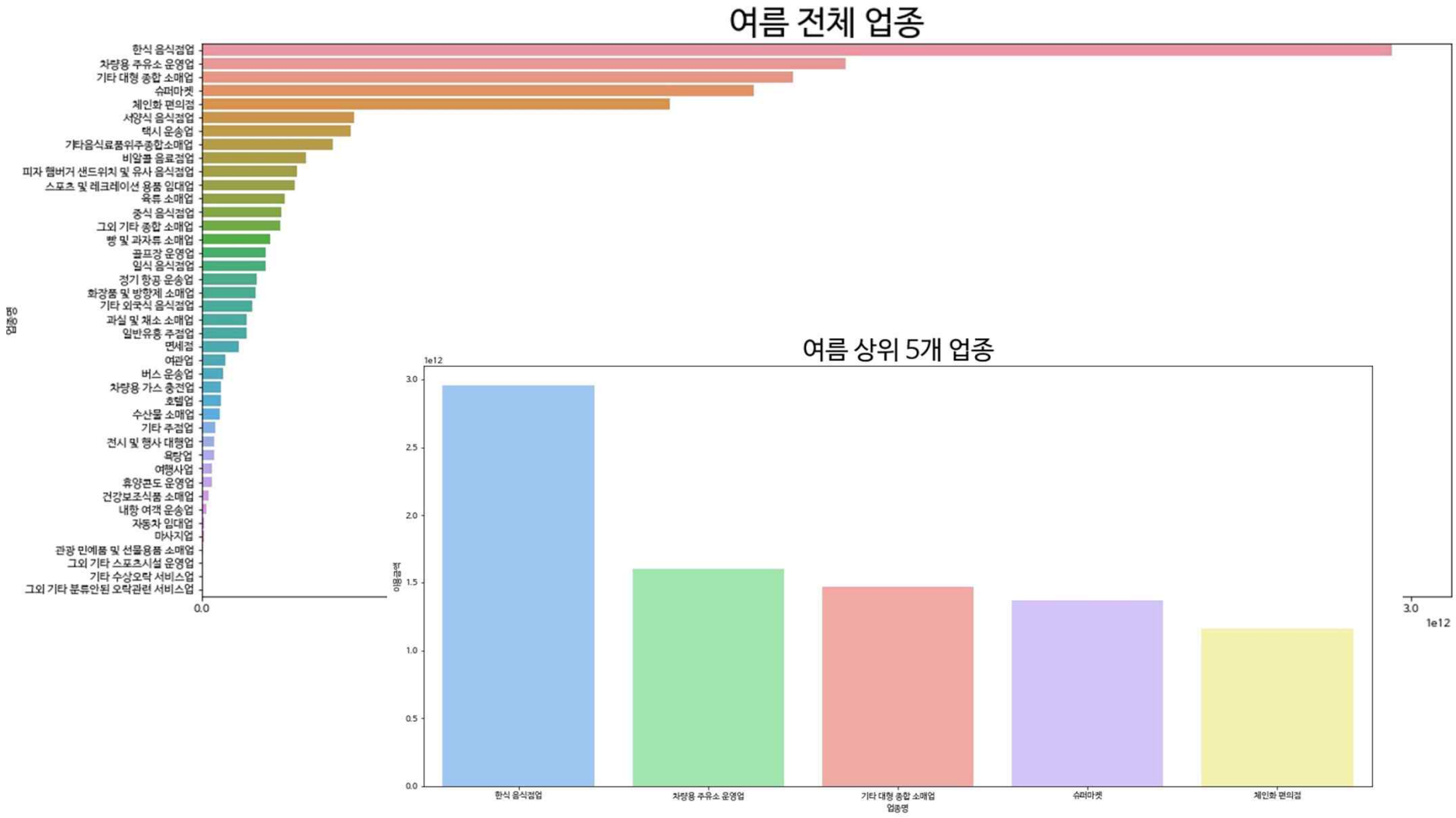
2-4) 계절


계절별 이용건수
봄
여름
가을
겨울
계절
0
20
40
60
80
100
120
이용건수
계절별 이용금액
0.0
0.5
1.0
1.5
2.0
2.5
1e6
이용금액
계절별 이용고객수
0
10
20
30
40
50
60
이용고객수

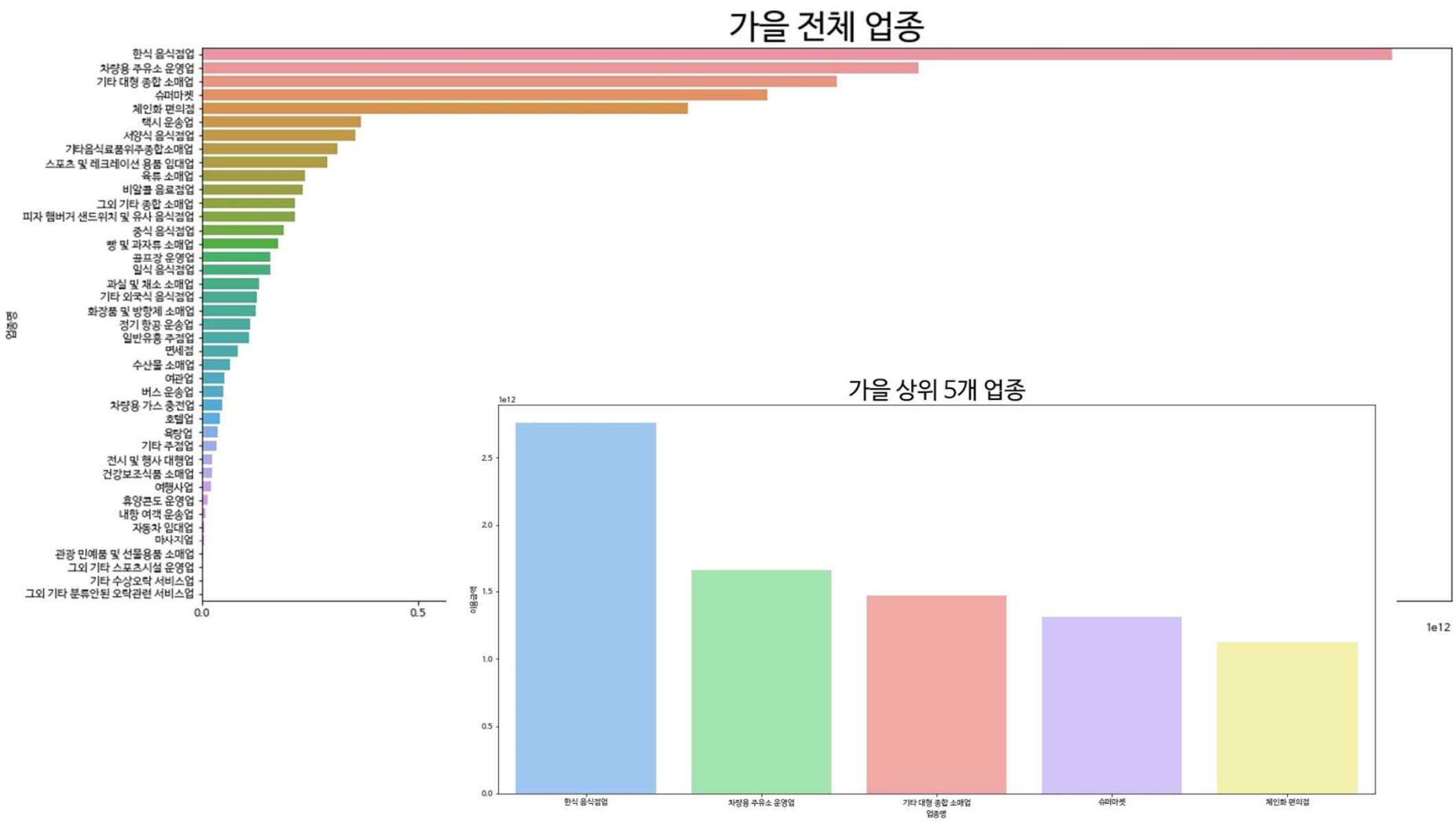
2-4) 계절


봄 전체 업종
한식 음식점업
차량용 주유소 운영업
기타 대형 종합 소매업
슈퍼마켓
체인화 편의점
서양식 음식점업
택시 운송업
기타음식료품위주종합소매업
스포츠 및 레크레이션 용품 임대업
비알콜 음료점업
피자 햄버거 샌드위치 및 유사 음식점업
육류 소매업
중식 음식점업
그외 기타 종합 소매업
빵 및 과자류 소매업
일식 음식점업
골프장 운영업
화장품 및 방향제 소매업
기타 외국식 음식점업
정기 항공 운송업
과실 및 채소 소매업
일반유흥 주점업
면세점
여관업
수산물 소매업
버스 운송업
차량용 가스 충전업
호텔업
욕탕업
기타 주점업
여행사업
전시 및 행사 대행업
건강보조식품 소매업
내항 여객 운송업
휴양콘도 운영업
자동차 임대업
마사지업
관광 민예품 및 선물용품 소매업
그외 기타 스포츠시설 운영업
기타 수상오락 서비스업
그외 기타 분류안된 오락관련 서비스업
업종명
0.0
0.5
봄 상위 5개 업종
1e12
3.0
2.5
2.0
1.5
1.0
0.5
0.0
이용금액
한식 음식점업
차량용 주유소 운영업
기타 대형 종합 소매업
슈퍼마켓
체인화 편의점
업종명

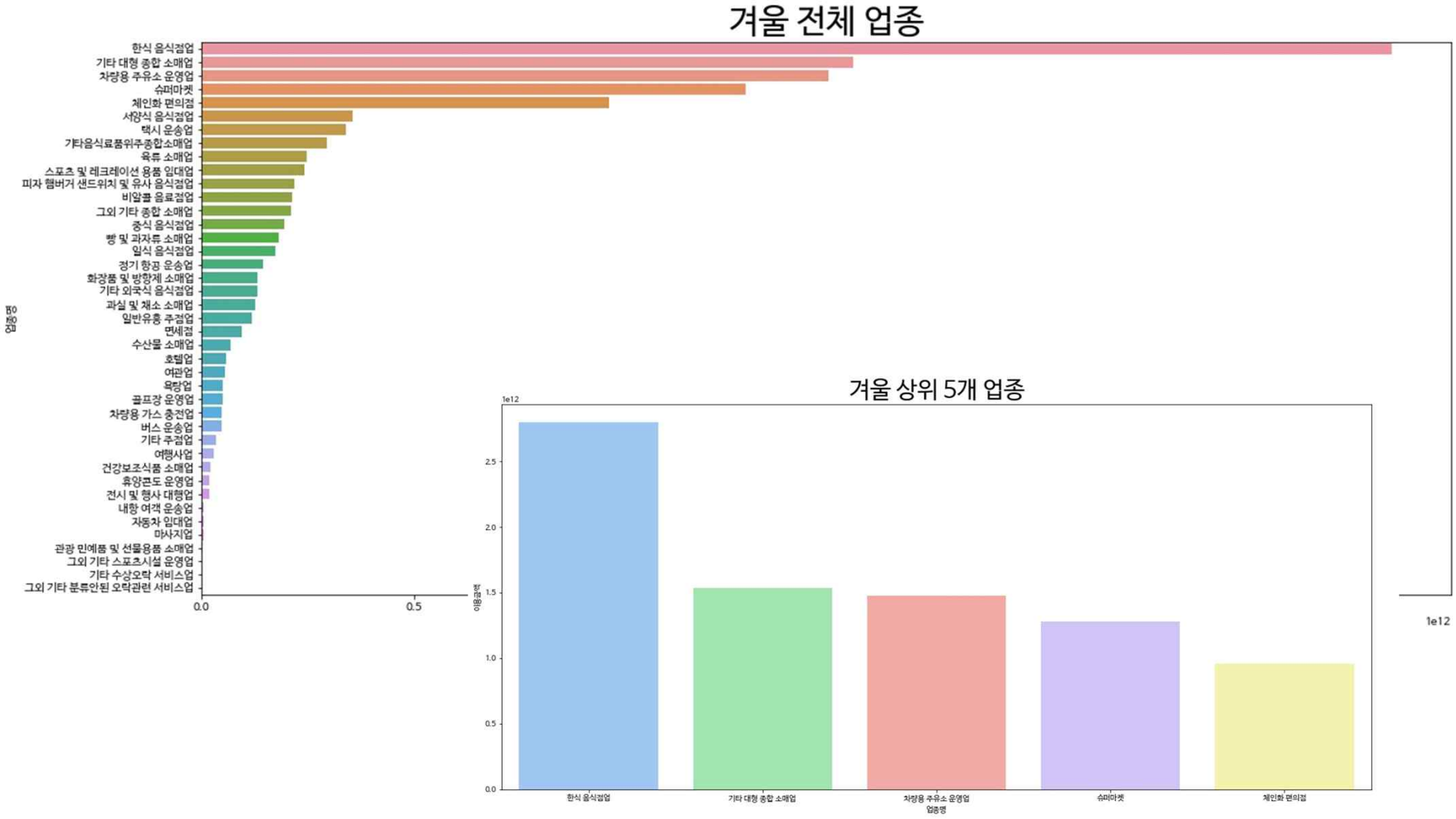
2-4) 계절


여름 전체 업종
한식 음식점업
차량용 주유소 운영업
기타 대형 종합 소매업
슈퍼마켓
체인화 편의점
서양식 음식점업
택시 운송업
기타음식료품위주종합소매업
비알콜 음료점업
피자 햄버거 샌드위치 및 유사 음식점업
스포츠 및 레크레이션 용품 임대업
육류 소매업
중식 음식점업
그외 기타 종합 소매업
빵 및 과자류 소매업
골프장 운영업
일식 음식점업
정기 항공 운송업
화장품 및 방향제 소매업
기타 외국식 음식점업
과실 및 채소 소매업
일반유흥 주점업
면세점
여관업
버스 운송업
차량용 가스 충전업
호텔업
수산물 소매업
기타 주점업
전시 및 행사 대행업
욕탕업
여행사업
휴양콘도 운영업
건강보조식품 소매업
내항 여객 운송업
자동차 임대업
마사지업
관광 민예품 및 선물용품 소매업
그외 기타 스포츠시설 운영업
기타 수상오락 서비스업
그외 기타 분류안된 오락관련 서비스업
업종명
0.0
3.0
1e12
여름 상위 5개 업종
1e12
3.0
2.5
2.0
1.5
1.0
0.5
0.0
매출액
한식 음식점업
차량용 주유소 운영업
기타 대형 종합 소매업
슈퍼마켓
체인화 편의점
업종명

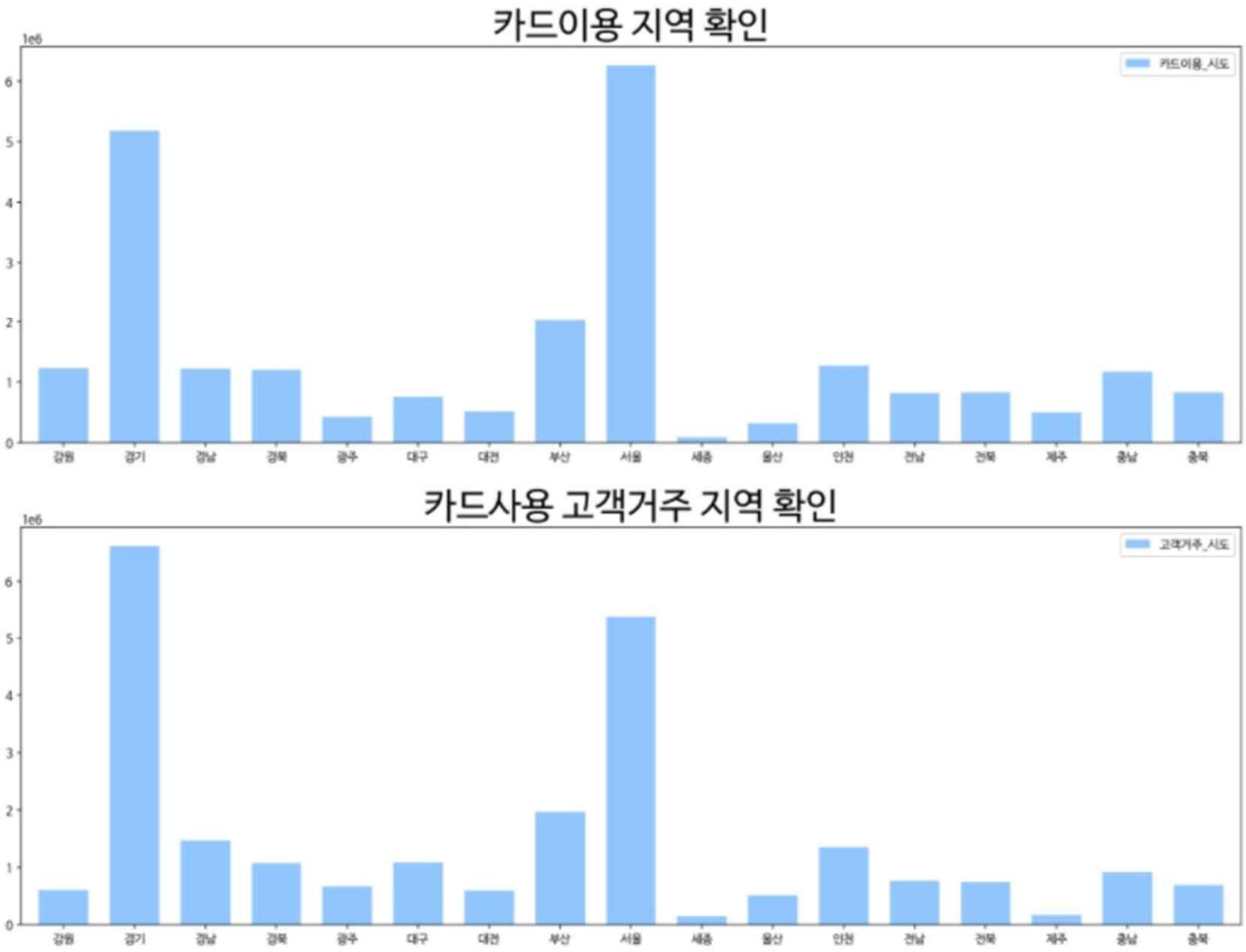
2-4) 계절


가을 전체 업종
한식 음식점업
차량용 주유소 운영업
기타 대형 종합 소매업
슈퍼마켓
체인화 편의점
택시 운송업
서양식 음식점업
기타음식료품위주종합소매업
스포츠 및 레크레이션 용품 임대업
육류 소매업
비알콜 음료점업
그외 기타 종합 소매업
피자 햄버거 샌드위치 및 유사 음식점업
중식 음식점업
빵 및 과자류 소매업
골프장 운영업
일식 음식점업
과실 및 채소 소매업
기타 외국식 음식점업
화장품 및 방향제 소매업
정기 항공 운송업
일반유흥 주점업
면세점
수산물 소매업
여관업
버스 운송업
차량용 가스 충전업
호텔업
요양업
기타 주점업
전시 및 행사 대행업
건강보조식품 소매업
여행사업
휴양콘도 운영업
내항 여객 운송업
자동차 임대업
마사지업
관광 민예품 및 선물용품 소매업
그외 기타 스포츠시설 운영업
기타 수상오락 서비스업
그외 기타 분류안된 오락관련 서비스업
업종명
0.0
0.5
1e12
가을 상위 5개 업종
1e12
2.5
2.0
1.5
1.0
0.5
0.0
이용금액
한식 음식점업
차량용 주유소 운영업
기타 대형 종합 소매업
슈퍼마켓
체인화 편의점
업종명

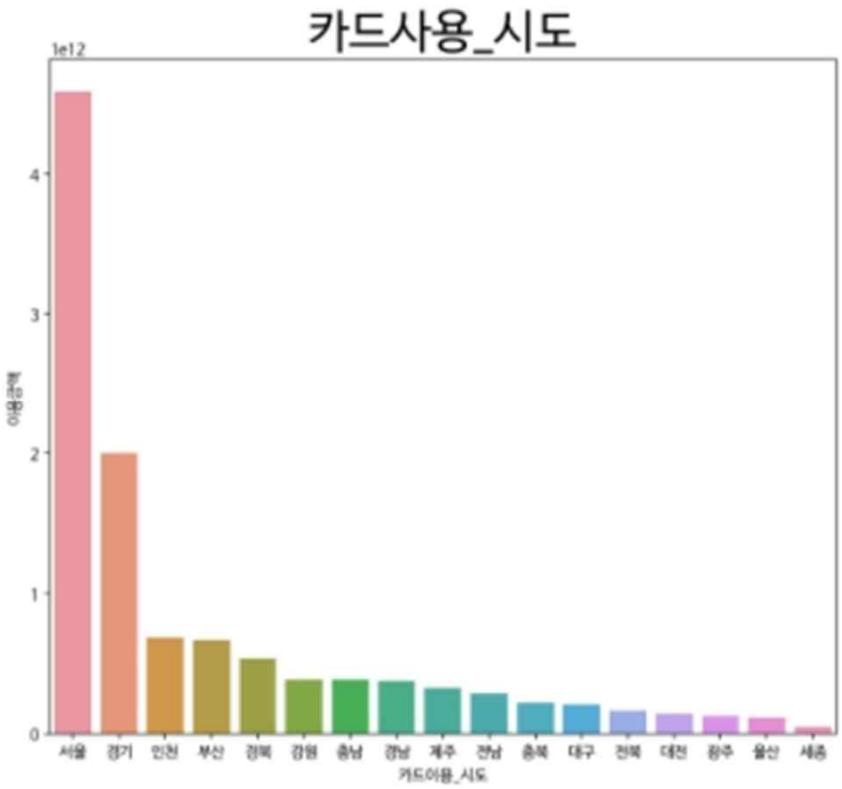
2-4) 계절

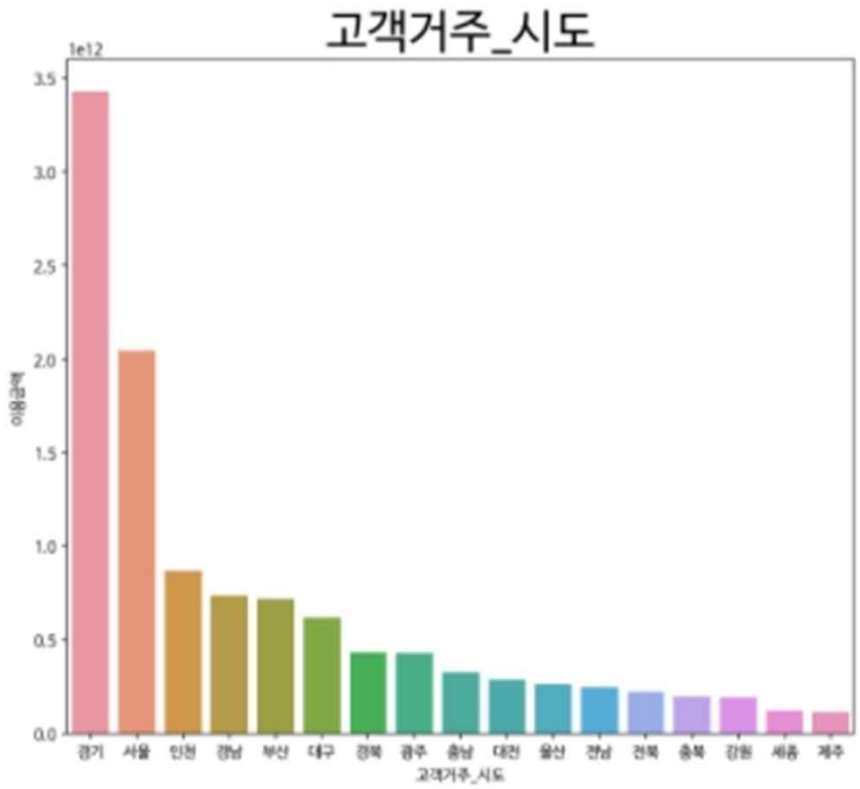

겨울 전체 업종
한식 음식점업
기타 대형 종합 소매업
차량용 주유소 운영업
슈퍼마켓
체인화 편의점
서양식 음식점업
택시 운송업
기타음식료품위주종합소매업
육류 소매업
스포츠 및 레크레이션 용품 임대업
피자 햄버거 샌드위치 및 유사 음식점업
비알콜 음료점업
그외 기타 종합 소매업
중식 음식점업
빵 및 과자류 소매업
일식 음식점업
정기 항공 운송업
화장품 및 방향제 소매업
기타 외국식 음식점업
과실 및 채소 소매업
일반유흥 주점업
면세점
수산물 소매업
호텔업
여관업
욕탕업
골프장 운영업
차량용 가스 충전업
버스 운송업
기타 주점업
여행사업
건강보조식품 소매업
휴양콘도 운영업
전시 및 행사 대행업
내항 여객 운송업
자동차 임대업
마사지업
관광 민예품 및 선물용품 소매업
그외 기타 스포츠시설 운영업
기타 수상오락 서비스업
그외 기타 분류안된 오락관련 서비스업
업종명
0.0
0.5
1e12
겨울 상위 5개 업종
1e12
2.5
2.0
1.5
1.0
0.5
0.0
한식 음식점업
기타 대형 종합 소매업
차량용 주유소 운영업
슈퍼마켓
체인화 편의점
업종명

2-5) 카드이용_시도, 고객거주_시도

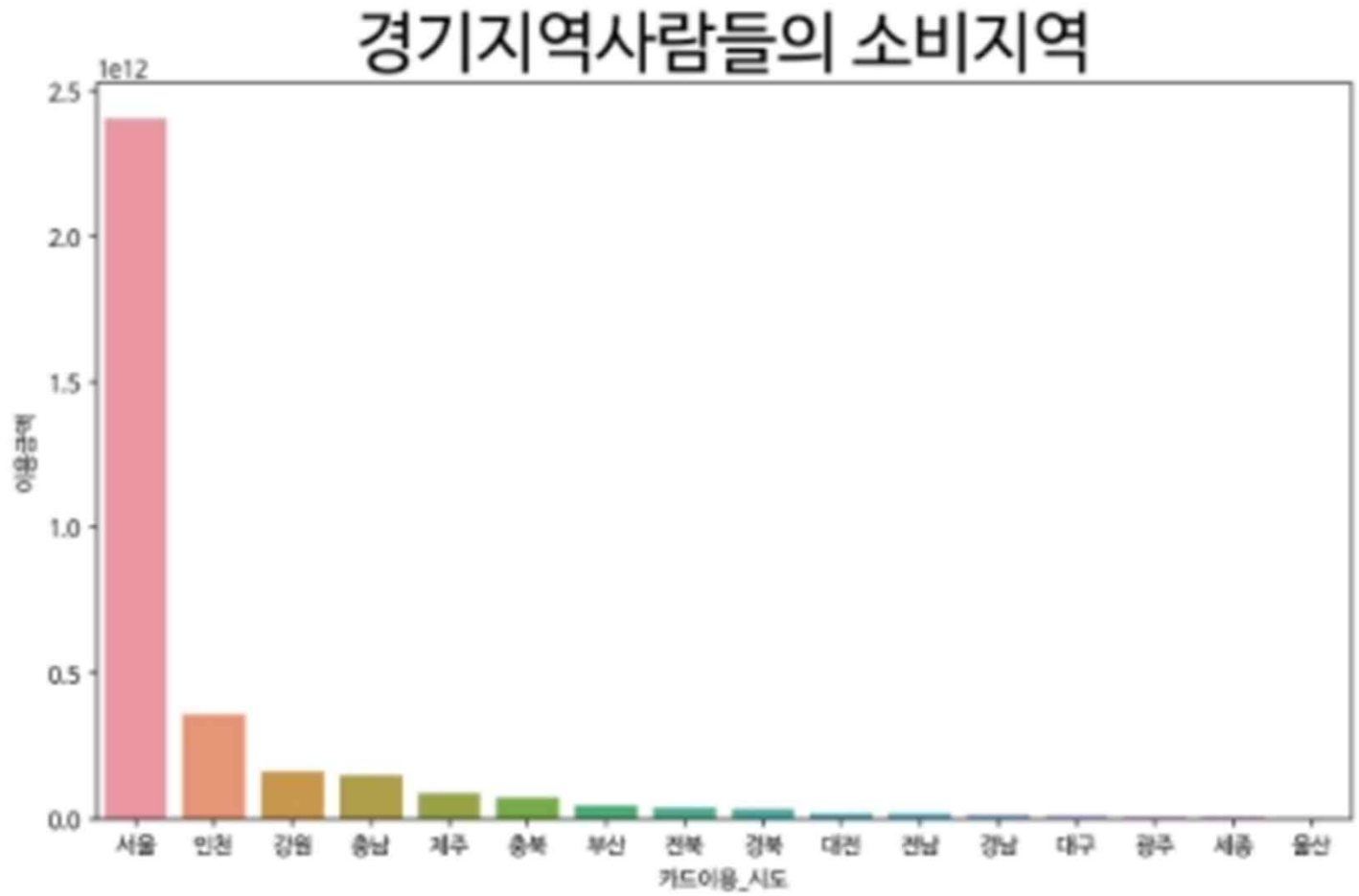


2-5) 카드이용_시도, 고객거주_시도



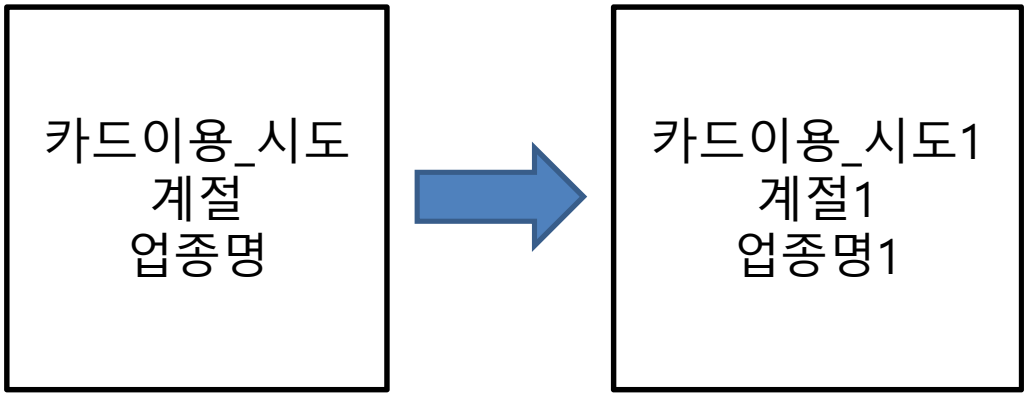


2-5) 카드이용_시도, 고객거주_시도



3-0) 데이터셋 만들기

	연월	카드이용_시도	업종명	고객거주_시도	연령대	성별	가구생애주기	이용고객수	이용금액	이용건수	년	월	계절	계절1	카드이용_시도1	업종명1
0	201901	강원	건강보조식품 소매업	강원	20	1	1	4	311200	4	2019	1	겨울	4	1	1
1	201901	강원	건강보조식품 소매업	강원	30	1	2	7	1374500	8	2019	1	겨울	4	1	1
2	201901	강원	건강보조식품 소매업	강원	30	2	2	6	818700	6	2019	1	겨울	4	1	1
3	201901	강원	건강보조식품 소매업	강원	40	1	3	4	1717000	5	2019	1	겨울	4	1	1
4	201901	강원	건강보조식품 소매업	강원	40	1	4	3	1047300	3	2019	1	겨울	4	1	1



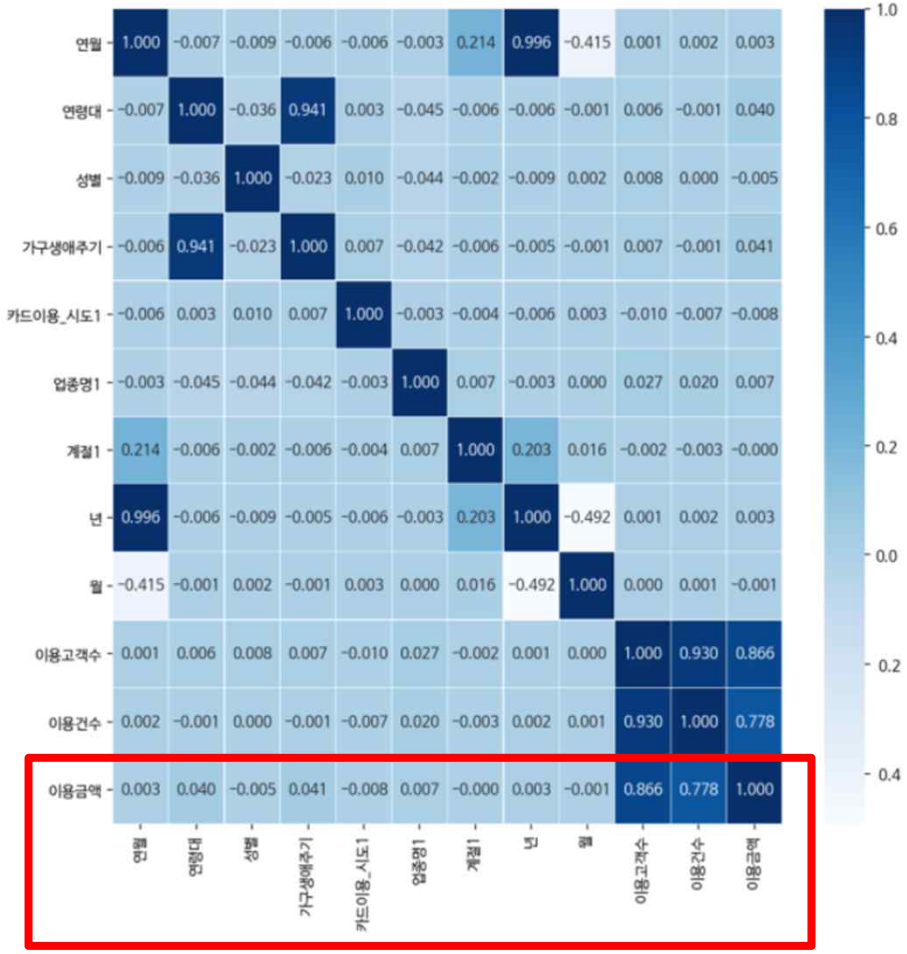
3-0) 데이터셋 만들기

	연월	연령대	성별	가구생애주기	카드이용_시도1	업종명1	계절1	년	월	이용고객수	이용건수	이용금액
0	201901	20	1	1	1	1	4	2019	1	4	4	311200
1	201901	30	1	2	1	1	4	2019	1	7	8	1374500
2	201901	30	2	2	1	1	4	2019	1	6	6	818700
3	201901	40	1	3	1	1	4	2019	1	4	5	1717000
4	201901	40	1	4	1	1	4	2019	1	3	3	1047300

독립변수

종속변수

3-0) 데이터셋 만들기-상관계수분석



$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

상관계수
두 변수의 상관 관계의 정도를 나타내는 수치

상관 계수가 0.4 이상이면
선형성이 있다고 판단

독립 변수(x) : 이용고객수,이용건수
종속 변수(y) : 이용금액

3-1) Linear Regression

$$\begin{cases} a_0 n + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum y x_1 \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum y x_2 \end{cases}$$

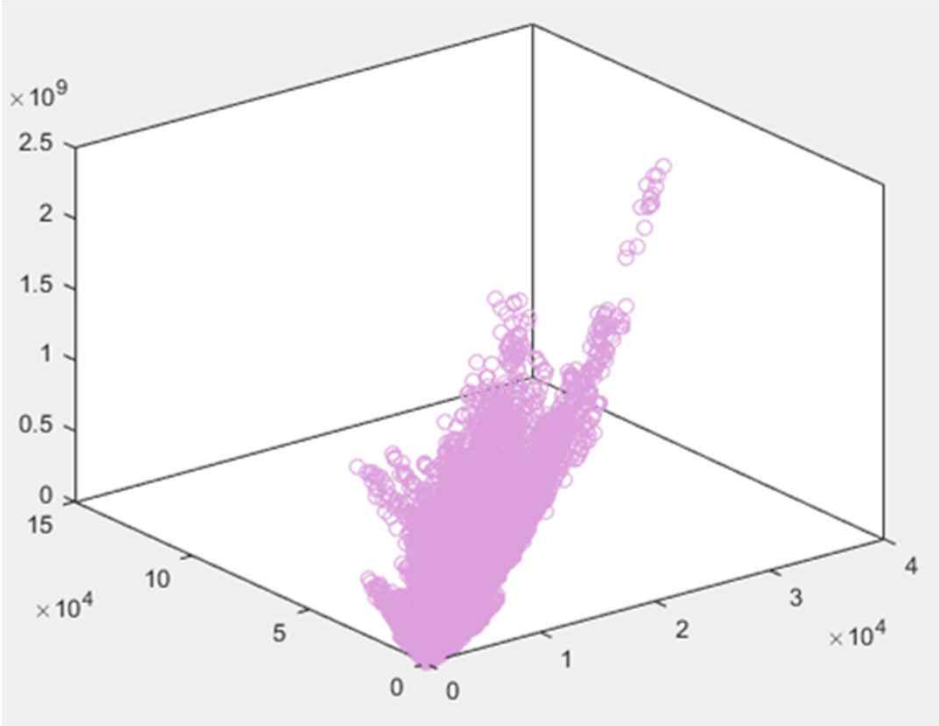
$$\boldsymbol{A * X = b}$$

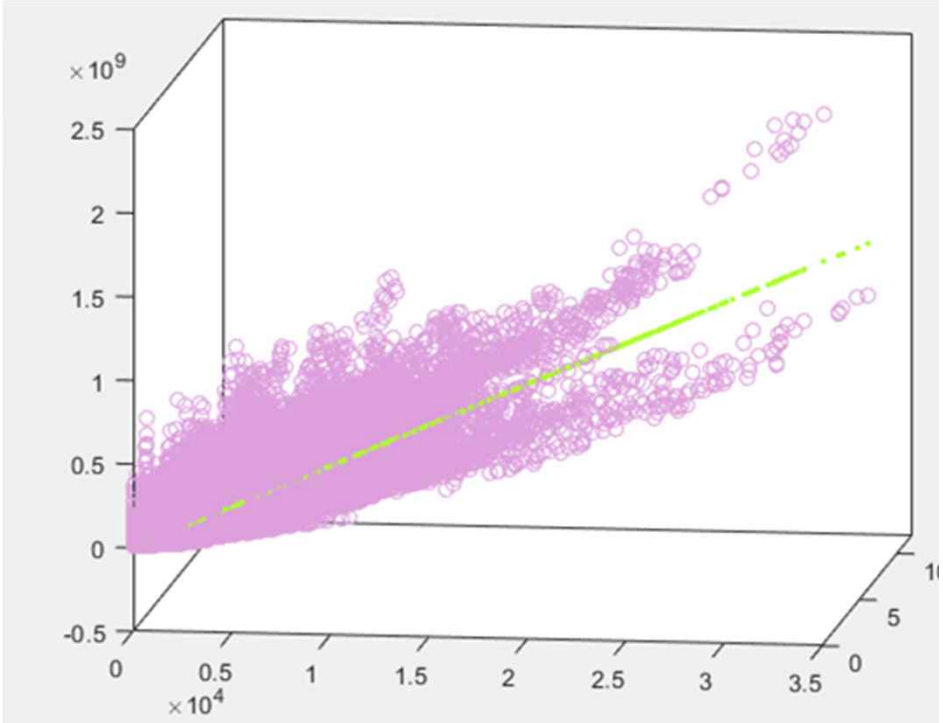
$$\begin{bmatrix} n & sum\ x_1 & sum\ x_2 \\ sum\ x_1 & sum\ x_1^2 & sum\ x_1 x_2 \\ sum\ x_2 & sum\ x_1 x_2 & sum\ x_2^2 \end{bmatrix} * \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} sum\ y \\ sum\ yx_1 \\ sum\ yx_2 \end{bmatrix}$$

$$\boldsymbol{X = A^{-1} * b}$$

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} n & sum\ x_1 & sum\ x_2 \\ sum\ x_1 & sum\ x_1^2 & sum\ x_1 x_2 \\ sum\ x_2 & sum\ x_1 x_2 & sum\ x_2^2 \end{bmatrix}^{-1} * \begin{bmatrix} sum\ y \\ sum\ yx_1 \\ sum\ yx_2 \end{bmatrix}$$

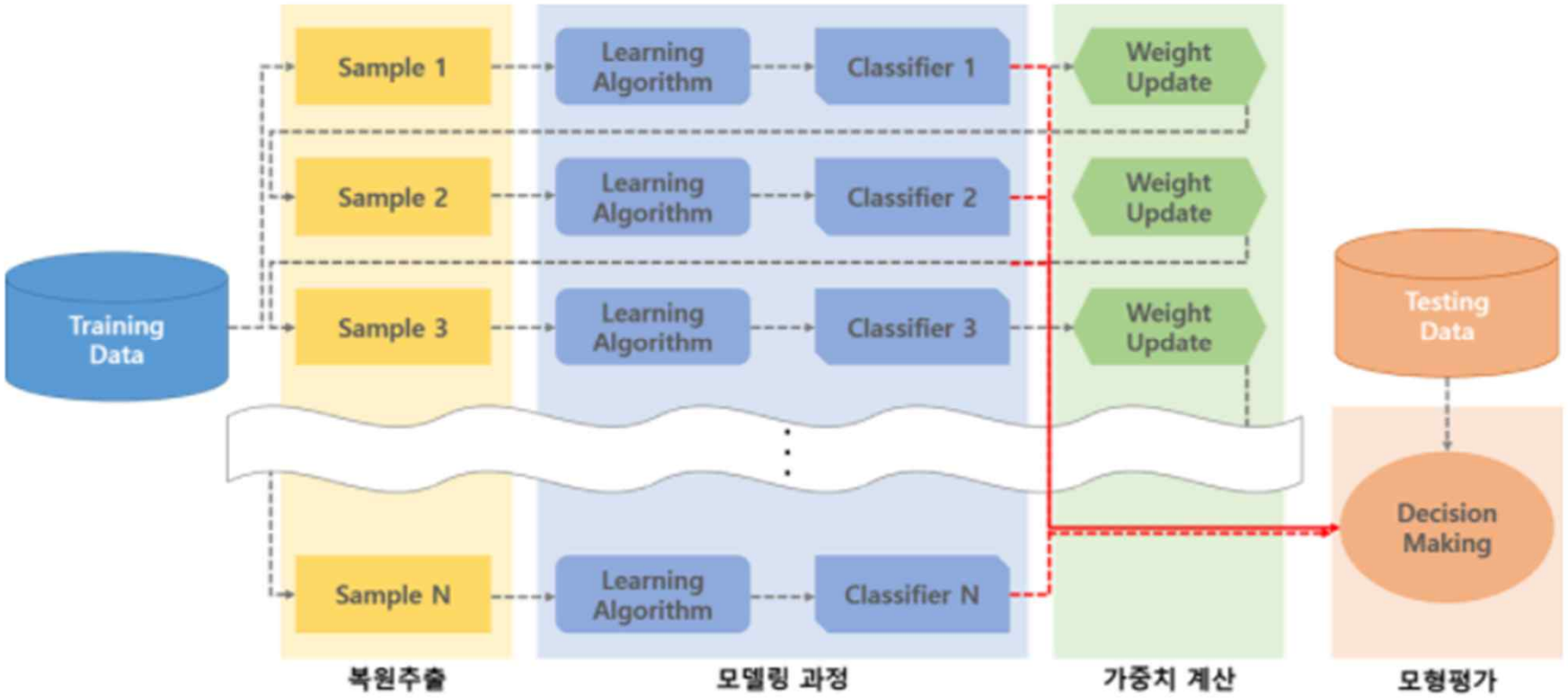
3-1) Linear Regrssion





3-2) XGBoost

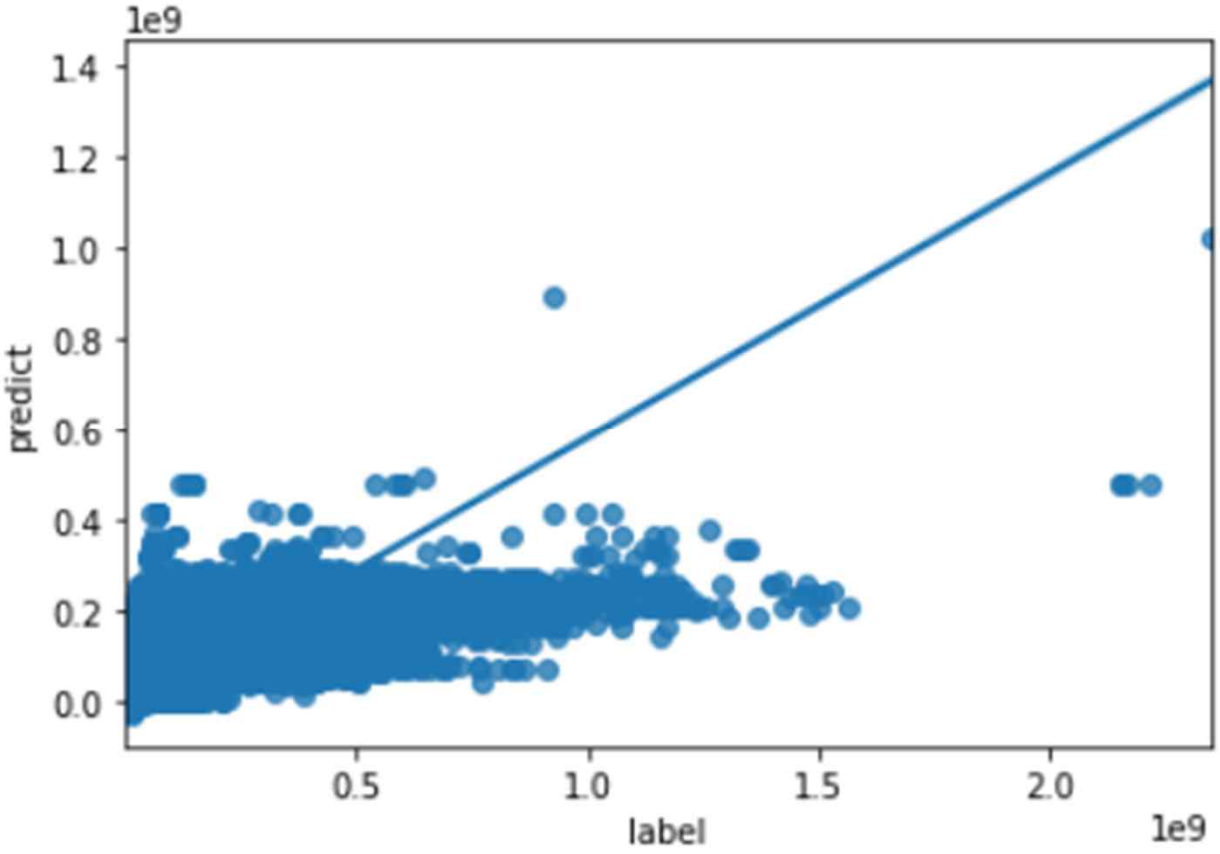
Gradient Boost



Gradient Boost : 여러 개의 약한 학습기를 순차적으로 학습-예측하면서 잘못 예측한 데이터에 가중치 부여를 통해 오류를 개선해 나가면서 학습하는 방식이다.
(가중치를 부여할 때 Gradient Decenting기법을 이용한다)

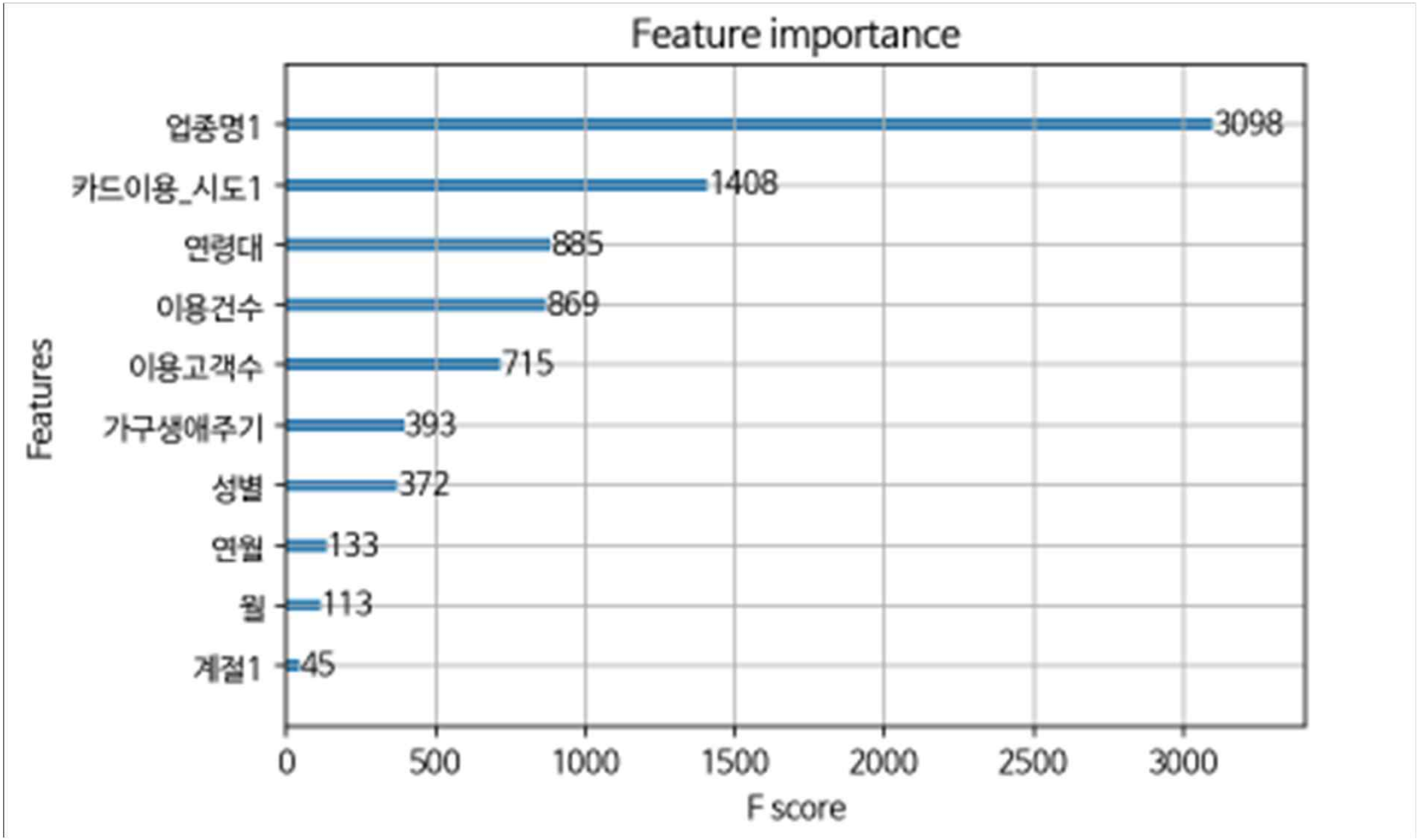
3-2) XGBoost

독립 변수 : 전체
종속 변수 : 이용금액
독립 변수 중요도 순위는 그래프이다.



결정 계수 : 0.6003

3-2) XGBoost



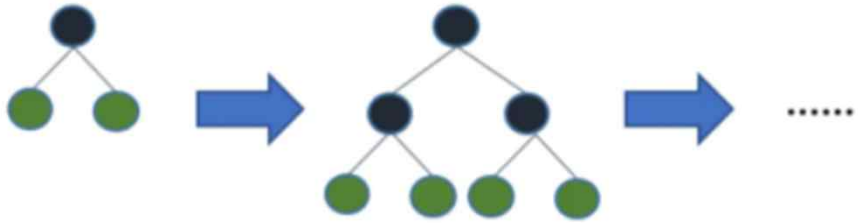
결정 계수 : 0.6003

3-2) XGBoost- 결정계수 계산법

$$R^2 = 1 - \frac{\sum(t_i - y_i)^2}{\sum(t_i - \overline{t_i})^2}$$

t_i : 실제값
y_i : 예측값
$\overline{t_i}$: 평균값

3-3) LightGBM



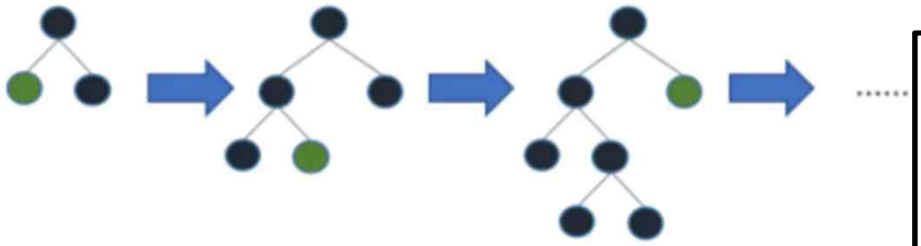
일반적인 boosting 기법

Gradient boosting 방식의 프레임 워크

차이점

일반적인 부스팅 기법은 tree가 수평적으로 확장되는데 LightGBM은 Tree가 수직적으로 확장

동일하게 확장될 때 일반적인 부스팅 기법보다 손실을 줄일 수 있다.

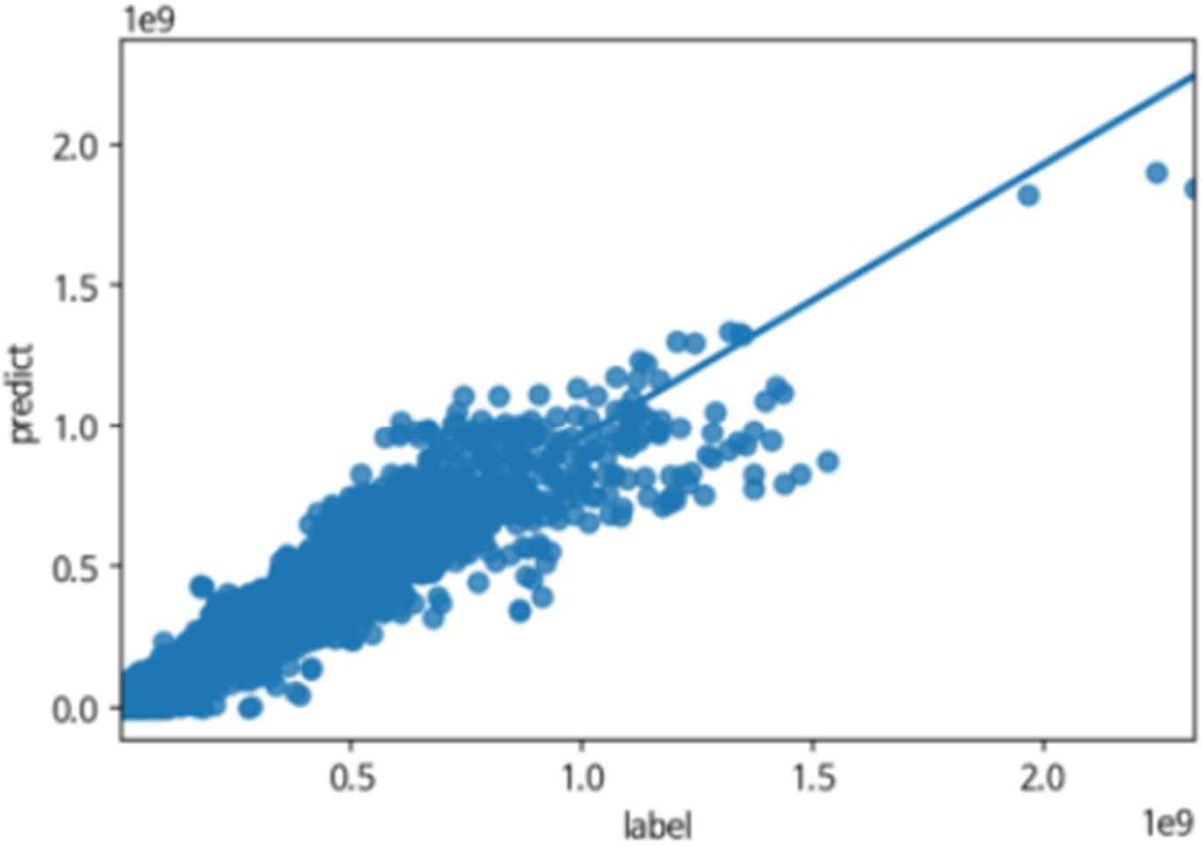


LightGBM 기법

LightGBM 장점

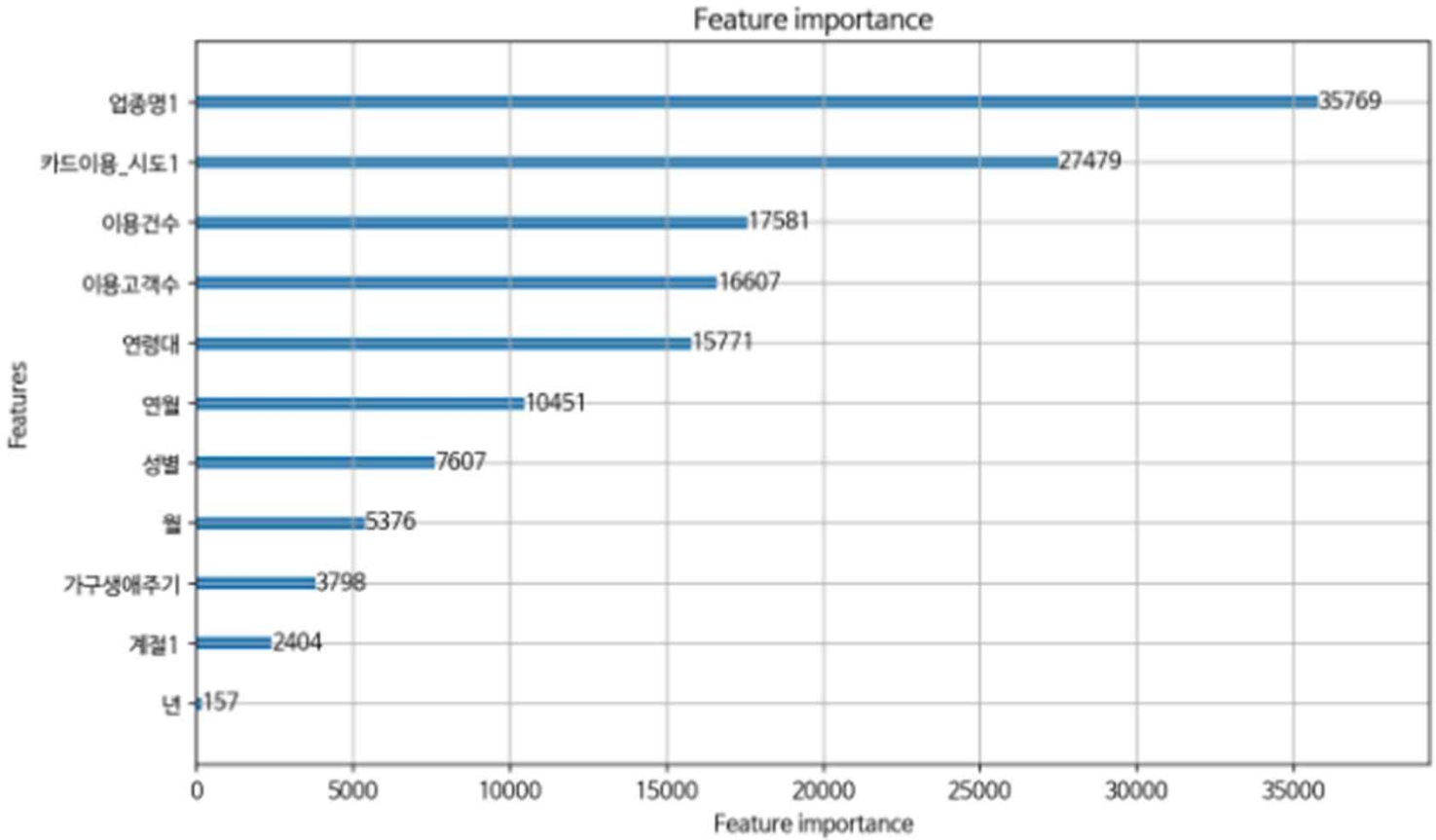
속도가 빠름
큰 사이즈의 데이터를 다룰 때 적은 메모리 차지

3-3) LightGBM



결정 계수 : 0.9676

3-3) LightGBM


Feature importance
업종명1
카드이용_시도1
이용건수
이용고객수
연령대
연월
성별
월
가구생애주기
계절1
년
35769
27479
17581
16607
15771
10451
7607
5376
3798
2404
157
Features
0
5000
10000
15000
20000
25000
30000
35000
Feature importance

4-1) 결과

선형회귀모델	XGBoost	LightGBM

사람들이 카드 소비를 덜 하게 되어 가계가 어려워진 상황이다.
예측금액에 따라 지원금을 차등 지급하는 방식의 정책을
만들 수 있을 것이다.

4-2) 아쉬운점 & 보완점

코로나 이후의 데이터는
1,2,3월 뿐이어서
그 이후의 데이터가 있다면
좀 더 잘 예측을 할 것 같다.

모델을 돌릴 때 생각보다 많은 시간이 소요되어 다양한 모델을 사용할 수 없었고, 파라미터를 다양하게 조절해볼 수 없어 기본 형태로만 돌려봤다. 파라미터를 조절하면 더 좋은 모델이 나올 것 같다.

감사합니다