

소비패턴 분석 & 소비예측모델

MZ세대 맞춤 카드추천



강예림 · 서다희 · 염은혜 · 이기근 · 진준모 · 최시환 · 홍주오

목차

프로젝트 배경

데이터 분석

소비예측 모델

카드 추천 모델

결론

01

MZ세대란?
MZ세대 소비 특징
소비분석 중요성
프로젝트 목표

02

데이터 소개
데이터 수집 & EDA
시각화 및 분석

03

예측모델 장단점
모델 적용
모델 학습 과정
예측치 그래프

04

카드 추천 알고리즘
카드 혜택 크롤링
알고리즘 실행 결과

05

프로젝트 결론
한계점
개선방안



1. 프로젝트 배경

MZ세대란?

디지털 환경에 익숙한 "디지털 원주민"

밀레니얼 세대

- ✓ 1981년 이후부터 1995년대 출생자들
- ✓ 아날로그와 디지털 문화가 혼재된 환경에서 자란 세대
- ✓ 최초의 글로벌 세대이자 인터넷 시대에서 성장한 첫 세대

Z세대

- ✓ 밀레니얼 세대와 알파 세대 사이에 있는 인구통계학적 집단
- ✓ 1996년 이후부터 2010년대 초 출생자들
- ✓ 어릴 때부터 인터넷과 휴대용 디지털 기술에 노출되며 성장한 최초의 사회세대

MZ세대 소비 특징

나 중심의 소비문화 추구

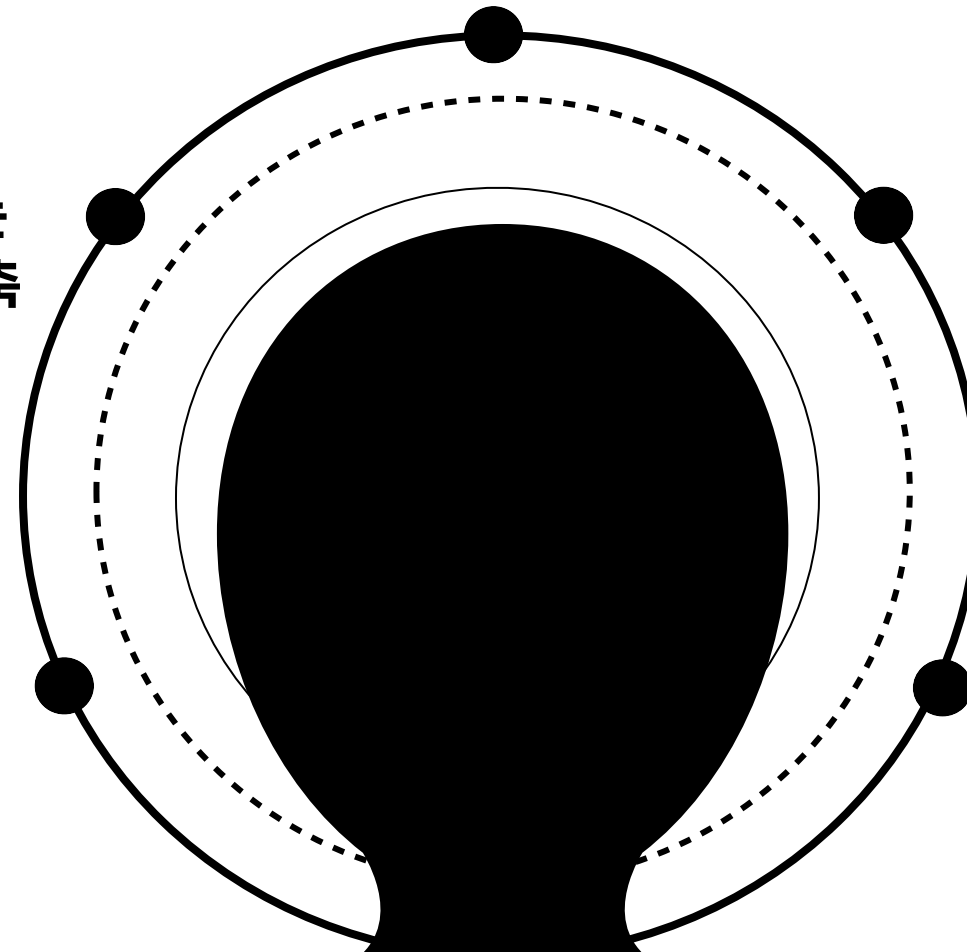
자신의 개성과 신념을 표출하는 워라밸 및 액티비티 소비를 지향
(사회적 가치나 특별한 메시지가 담긴 물건을 구매)

새롭고 남다른 경험을 추구하는
고유의 소비특성을 구축

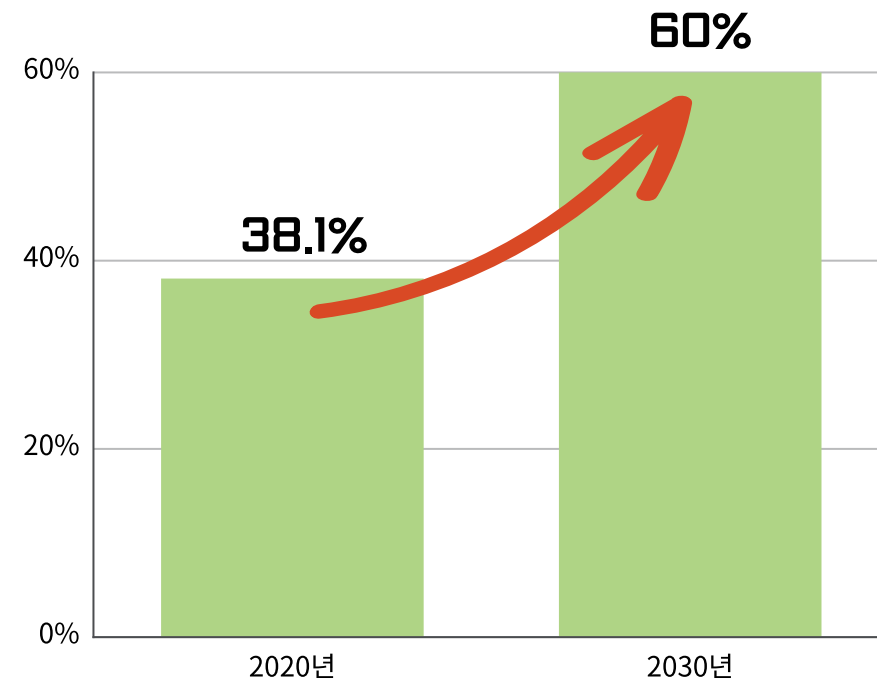
솔직하고 자기표현이 강하며,
디지털과 온라인 트렌드에 민감

소유보다는 공유를 중시하는 소비 특징
(공유 경제, 중고거래, 렌트 서비스 등)

현재의 즐거움을 중시하고,
구매 결정 요인이 가격이 아닌
개개인의 취향이 우선되는 경우가 많음

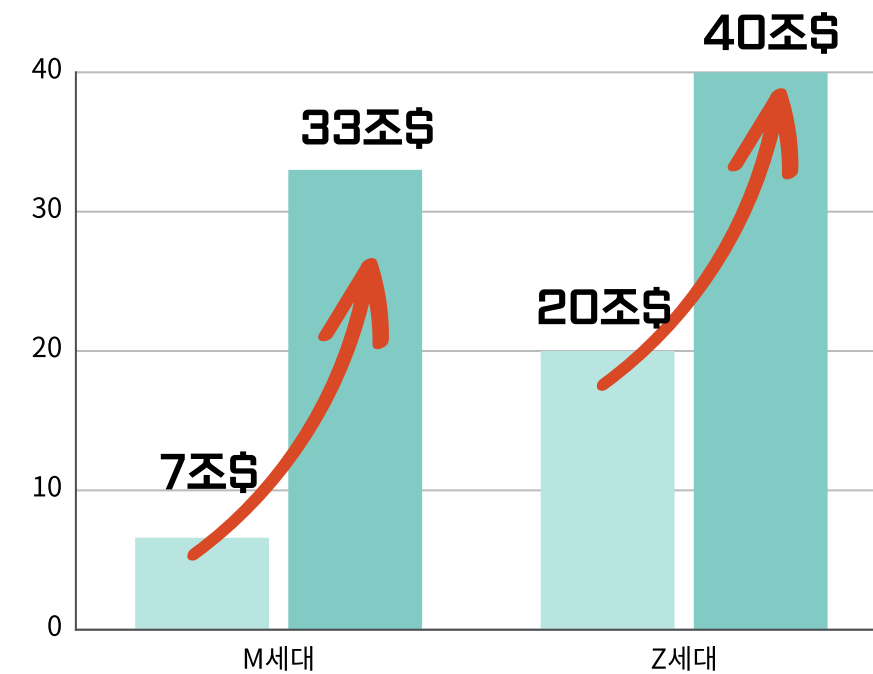


MZ세대 소비분석 중요성



생산연령 인구 비중 확대

2030년 기준 국내 생산연령 인구(15~64세)
→ 약 60%를 차지 예상



MZ세대 소득 증가

세계 MZ세대의 소득
→ 총소득의 60%를 차지할 것으로 예상

→ **부와 소비의 중심 MZ세대로 이동**

Project 목적

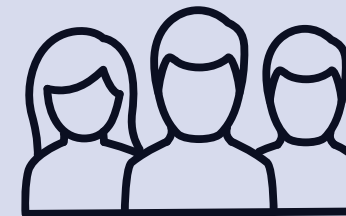
연령별 가계부채 비중¹⁾²⁾

	30대 이하	40대	50대	60대 이상
2018년	25.6	28.7	28.1	17.6
2019년	24.9	29.5	27.8	17.8
2020년	27.0	28.2	25.9	18.9
2021년	27.1	29.0	25.4	18.5
2022.1/4	27.5	28.0	25.4	19.0

주: 1) 기말 기준

2) 전체 가계대출잔액 대비

자료: 한국은행 시산(가계부채DB)



- ☑ 율로, 카푸어, 영끌족 등 저축이 빈약한 소비 특성
- ☑ 지속적인 인플레이션과 고금리, 그리고 물가상승
- ☑ 가계부채 중 2030 청년층의 빚의 차지비중 증가

MZ세대의 효율적인 소비를 위한 도움 필요



소비예측 및 카드추천





2. 데이터 분석

데이터 수집 & EDA

서울특별시 빅데이터 캠퍼스

캠퍼스소개 빅데이터 이용 신청 이용자

나의 데이터경쟁력, 빅캠과 함께 UP! UP!

전체 데이터 현황 318 종

서울시민의 업종별 카드소비 패턴 🔍 검색하기

경제/금융 | TXT DB

서울시민의 업종별 카드소비 패턴 데이터

국내 전 지역 가맹점을 대상으로 서울시민의 개인카드(신용카드와 체크카드 포함, 법인카드 제외) 사용한 패턴을 집계한 데이터입니다.

서울시민 거주지 기준으로 가맹점소재지별/일별/업종대분류별, 업종대분류별/일별/시간대별, 업종소분류별/월별/성별/연령대별로 카드이용금액과 이용건수, 업종소분류별/월별/시간대별로 카드이용건수와 소액결제건수를 집계한 데이터입니다...

제공 기관	신한카드	적재주기	월
공간 범위	서울전역	시간 범위	2016.01 ~ 2022.08

01 서울시 빅데이터 캠퍼스의 '서울시민의 업종별 카드소비 패턴' 데이터를 이용

02 '카드소비 업종코드' 및 '성별 연령대별 카드소비패턴'에 해당하는 4만개의 관측치를 확보 및 병합하여 분석 진행

03 총9개의 칼럼으로 구성
(업종코드, 기준년월, 성별, 연령대별, 카드이용금액계, 카드이용건수계, 대분류, 중분류, 소분류)

데이터 분석

데이터 소개

서울시 소비데이터

(2021.01 - 2022.08)

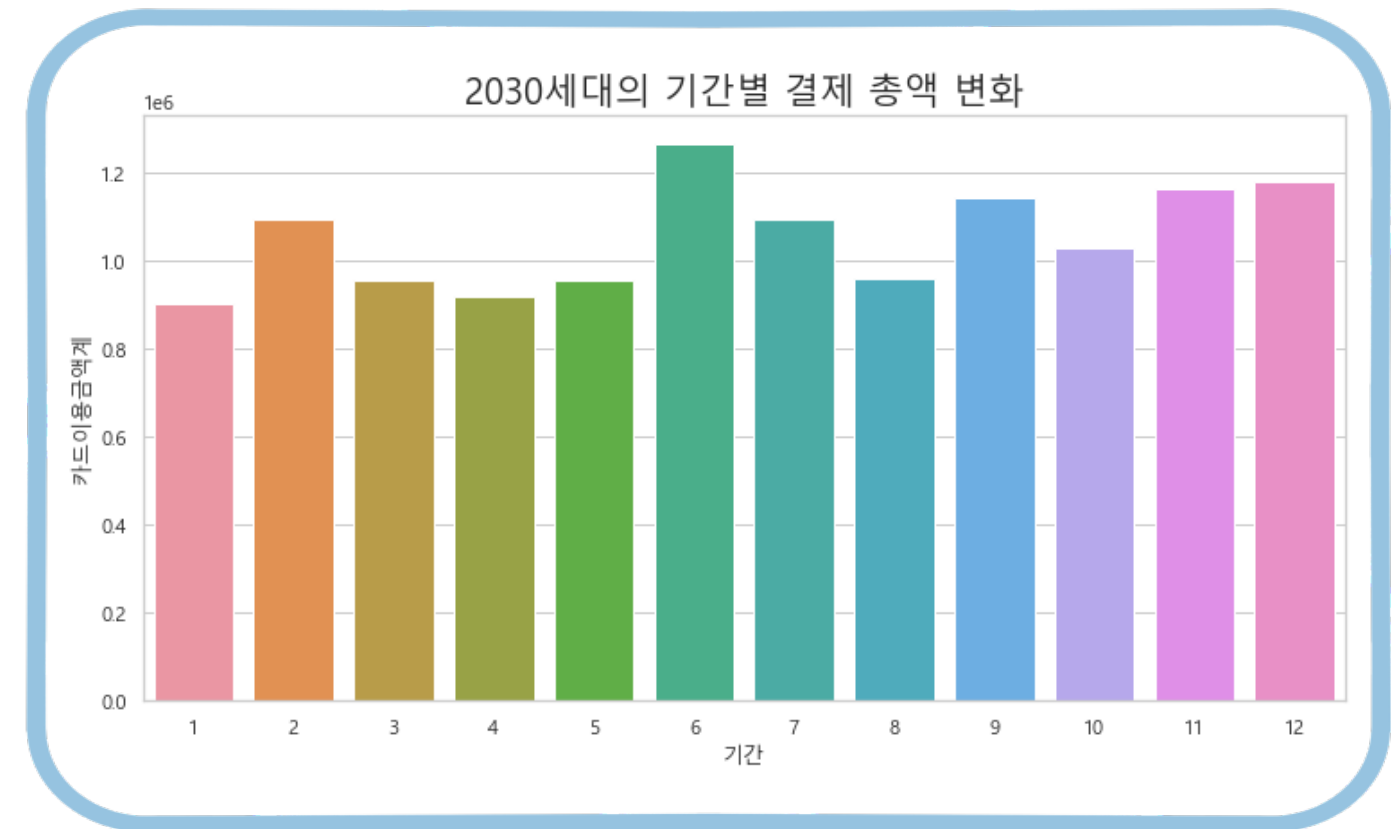
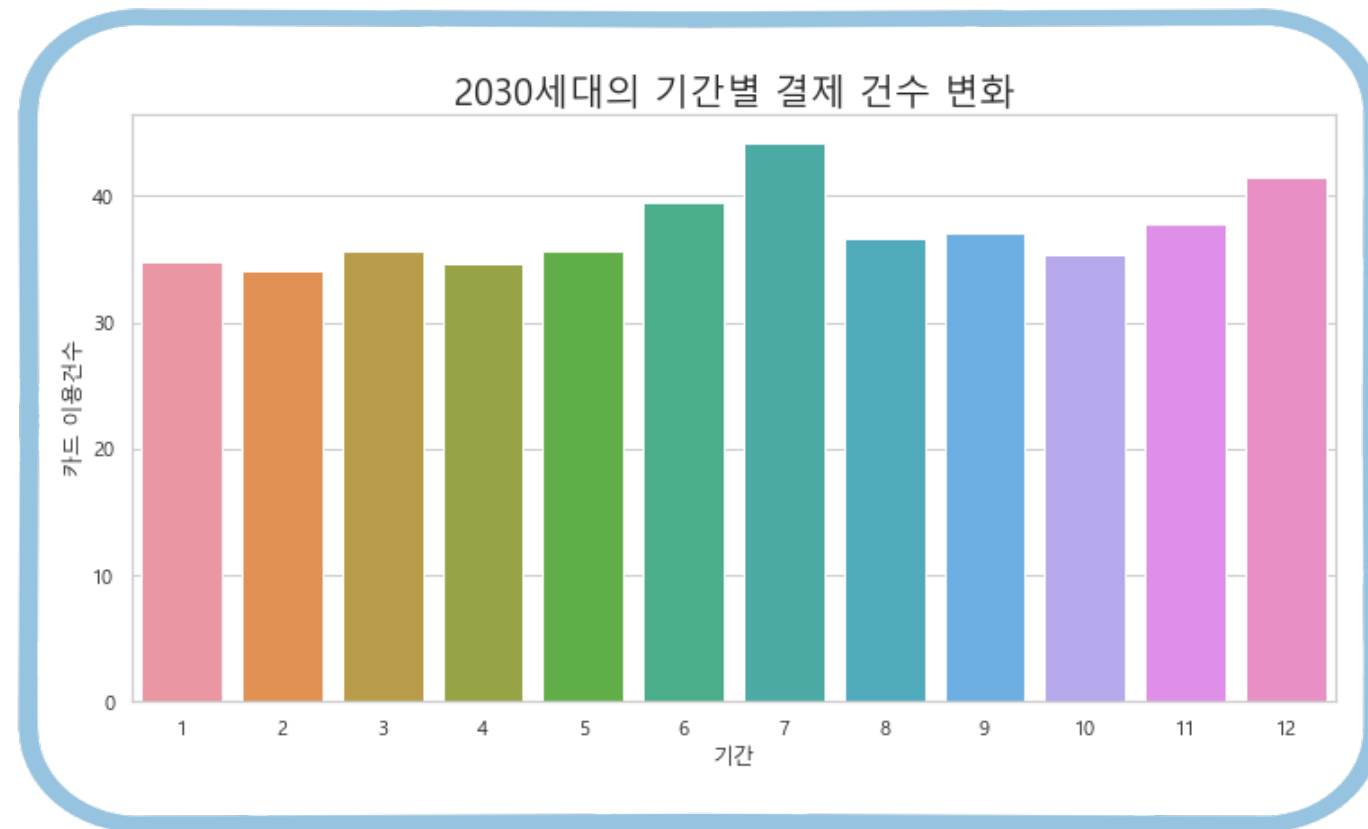
Columns	Non-Null Count	Dtype
서울시민업종코드	39999 non-null	object
대분류	39999 non-null	object
중분류	39999 non-null	object
소분류	39999 non-null	object
기준년월	39999 non-null	int64
고객주소블록코드	39999 non-null	int64
성별	39999 non-null	object
연령대별	39999 non-null	object
카드이용금액계	39999 non-null	int64
카드이용건수계	39999 non-null	int64

개인 소비데이터

(2021.01 - 2022.01)

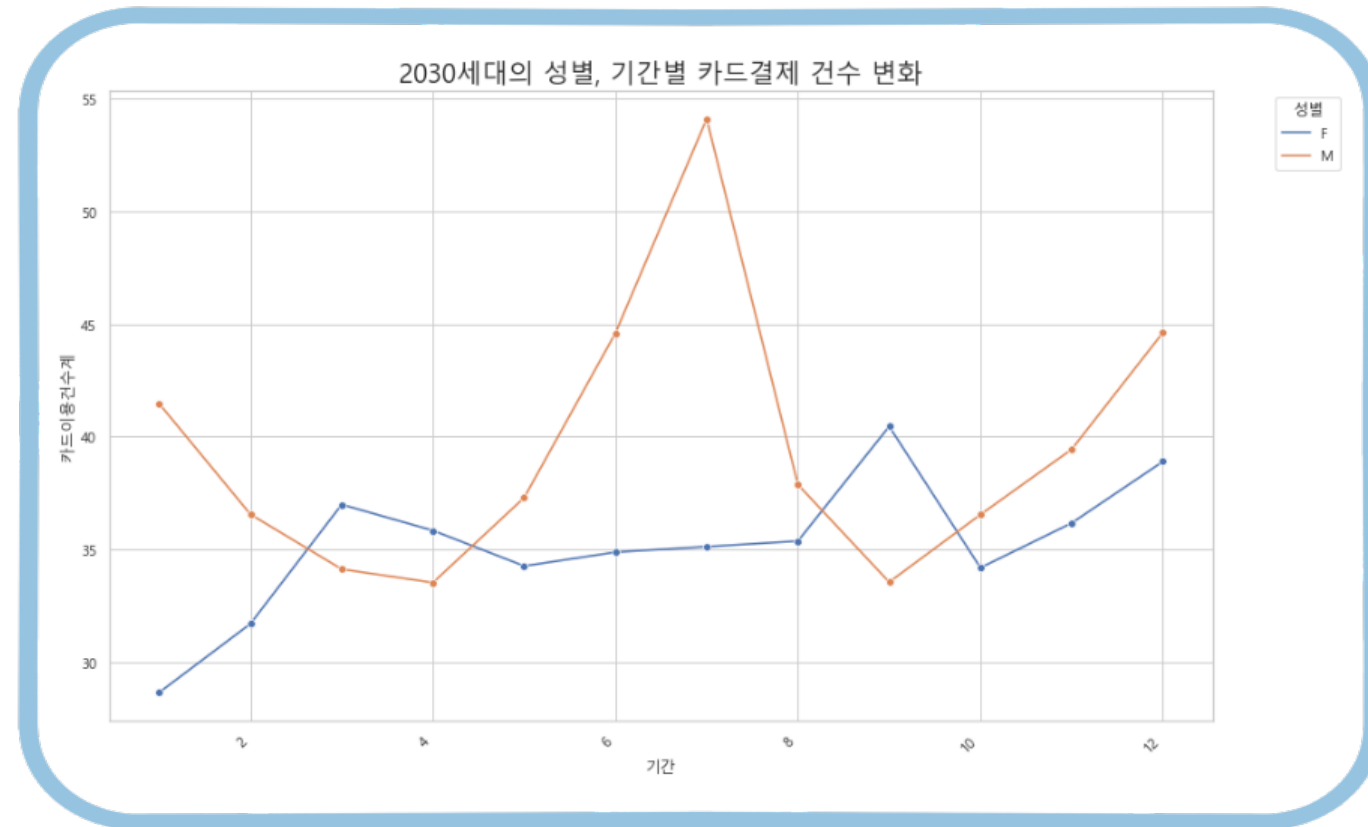
Columns	Non-Null Count	Dtype
날짜	3741 non-null	object
금액	3741 non-null	int64
소분류	3741 non-null	object
업종코드	3741 non-null	object
대분류	3741 non-null	object
중분류	3741 non-null	object

2030 기간별 결제 건수/총액 변화

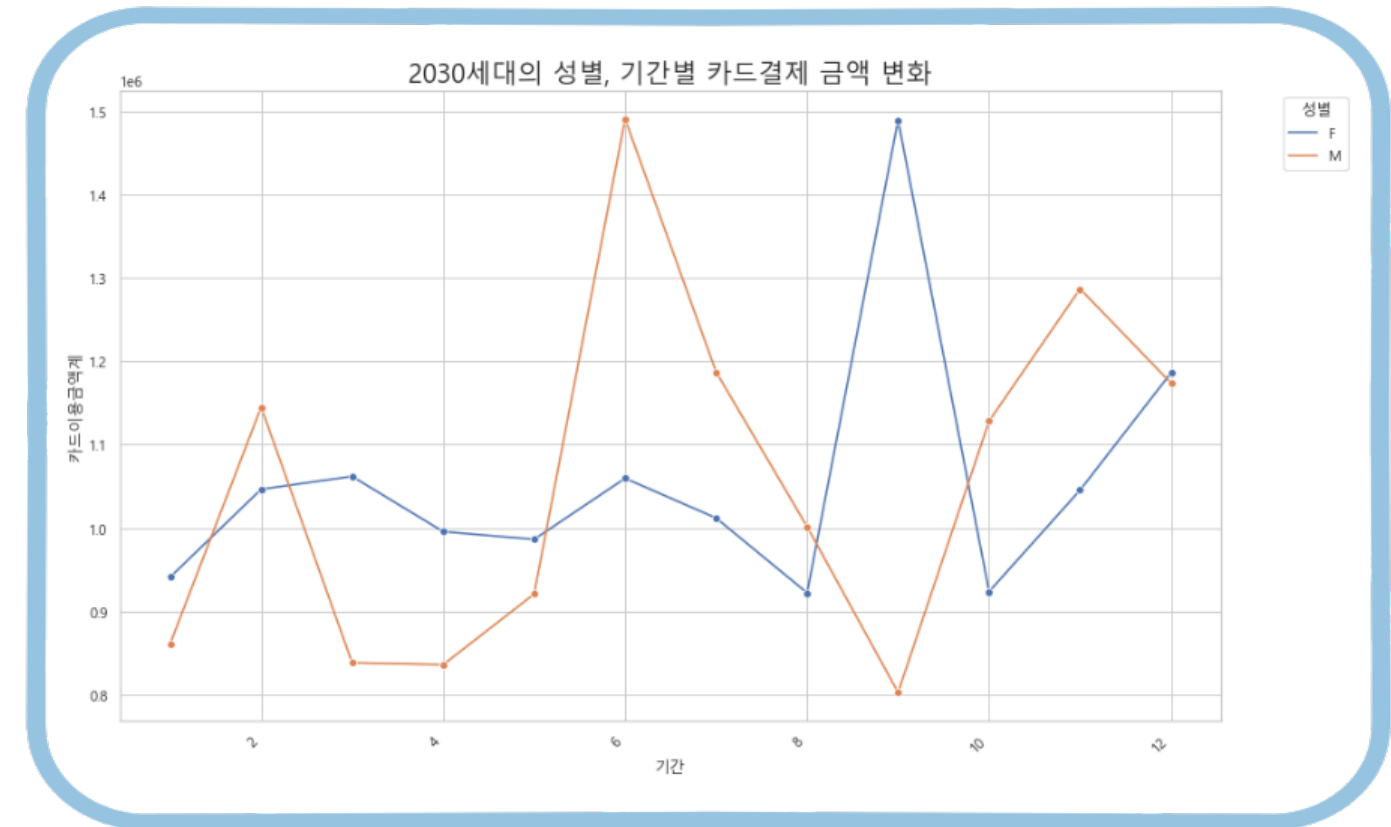


➡ 대체로 휴가철이나 연말 기간에 결제 건수와 금액 모두 높아지는 양상을 보임

2030 성별 · 기간별 결제 건수/총액 변화



남성 → 1월, 7월, 12월에 급격하게 카드 결제건수 증가
 여성 → 9월과 12월에 비교적 완만하게 카드 결제건수 증가

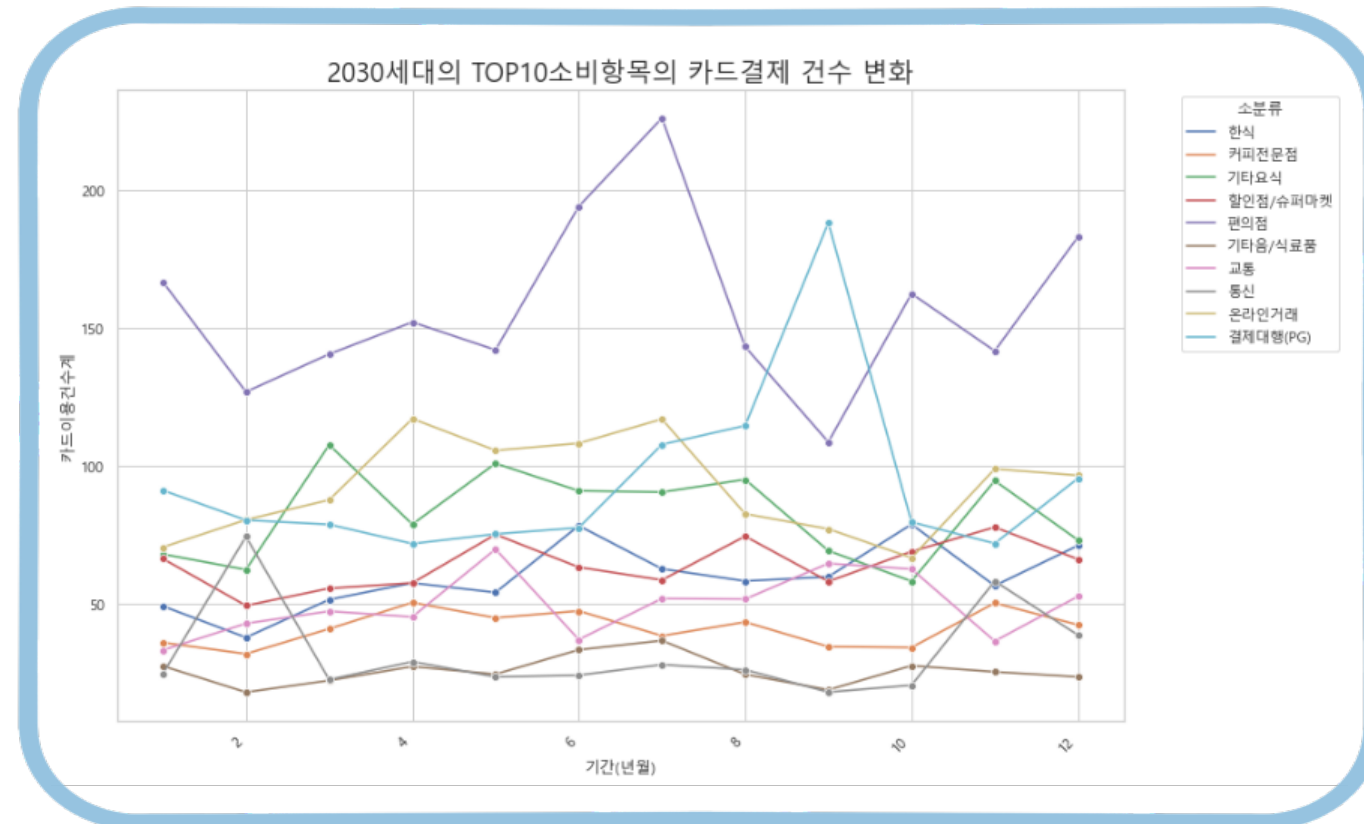


남성 → 6월에 지출이 150만원으로 급격하게 증가
 여성 → 9월에 지출이 150만원으로 급격하게 증가

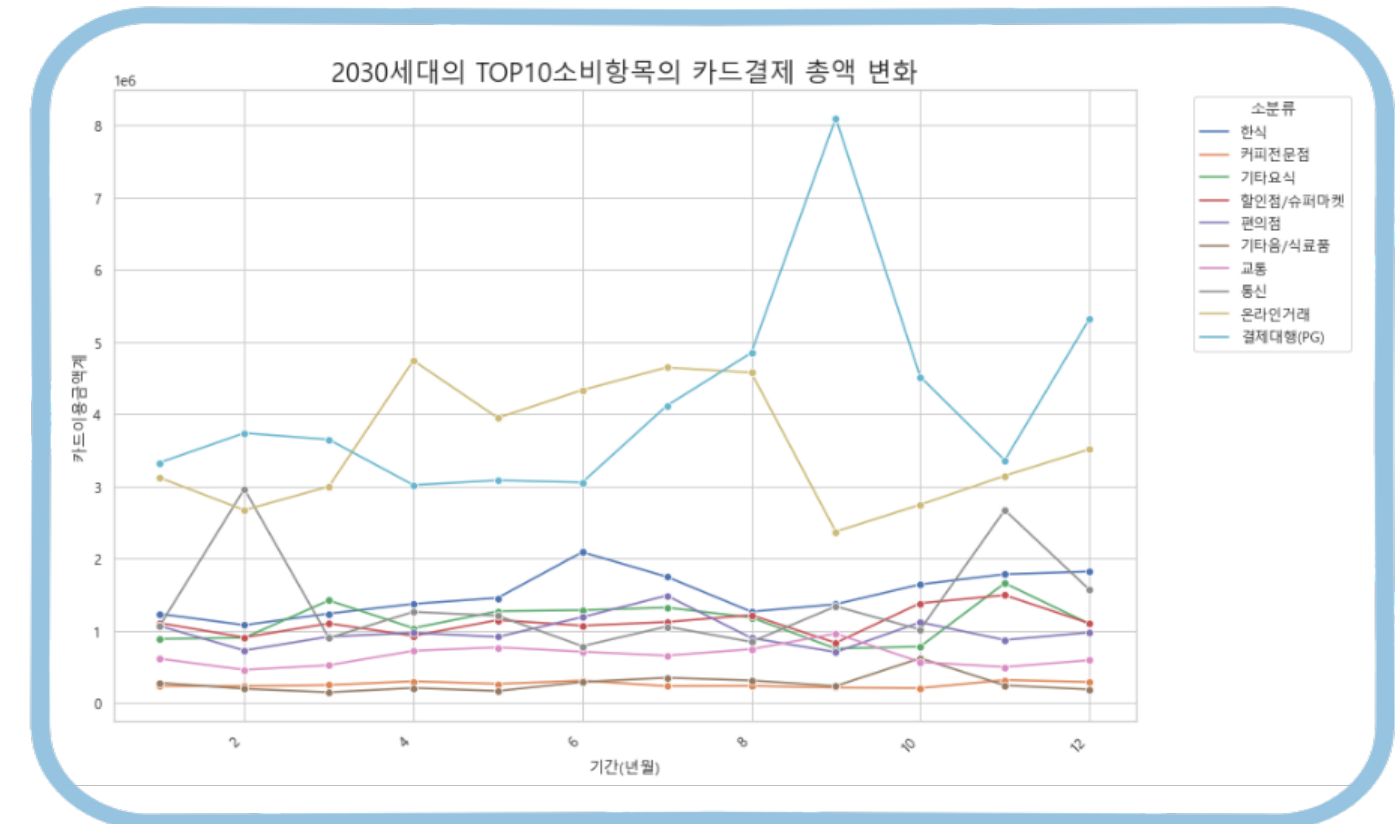


남성은 대체로 특정 기간에 집중된 소비 지향
 여성은 비교적 기간에 영향을 받지 않은 고른 소비 지향
 지출면에서는 남녀 구분없이 변동 폭이 큼

2030 소비항목Top10 결제 건수/총액 변화



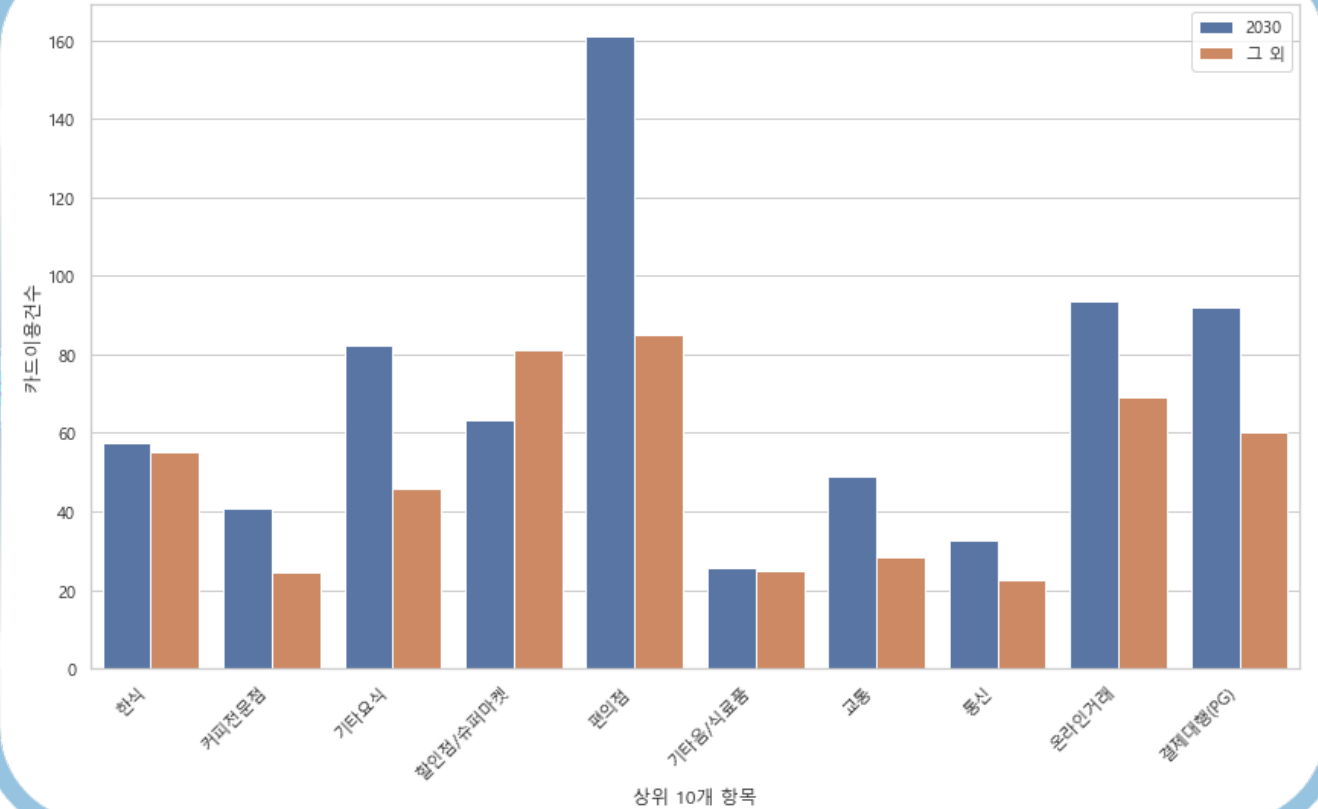
- ① 편의점, 온라인거래, 결제대행에 이용되는 카드결제건수
→ 평균 월 100회 이상, 가장 높음
- ② 기타음/식료품 및 통신에 이용되는 카드결제 건수
→ 평균 월 30회 미만, 가장 낮음
- ③ 편의점, 결제대행이 7월, 9월에 급격히 늘어나는 양상을 띠
- ④ 다른 항목은 대체로 고른 분포를 보여줌



- ① 결제대행, 온라인거래에 지출되는 카드결제총액
→ 평균 월 300만원 이상, 가장 많음
- ② 커피전문점, 기타음/식료품에 지출되는 카드결제 총액
→ 평균 월 30만원 미만, 가장 적음
- ③ 결제대행이 9월에 급격히 늘어나는 양상을 띠
- ④ 다른 항목은 대체로 고른 분포를 보여줌

2030 vs 타 연령대 결제 건수/총액 변화

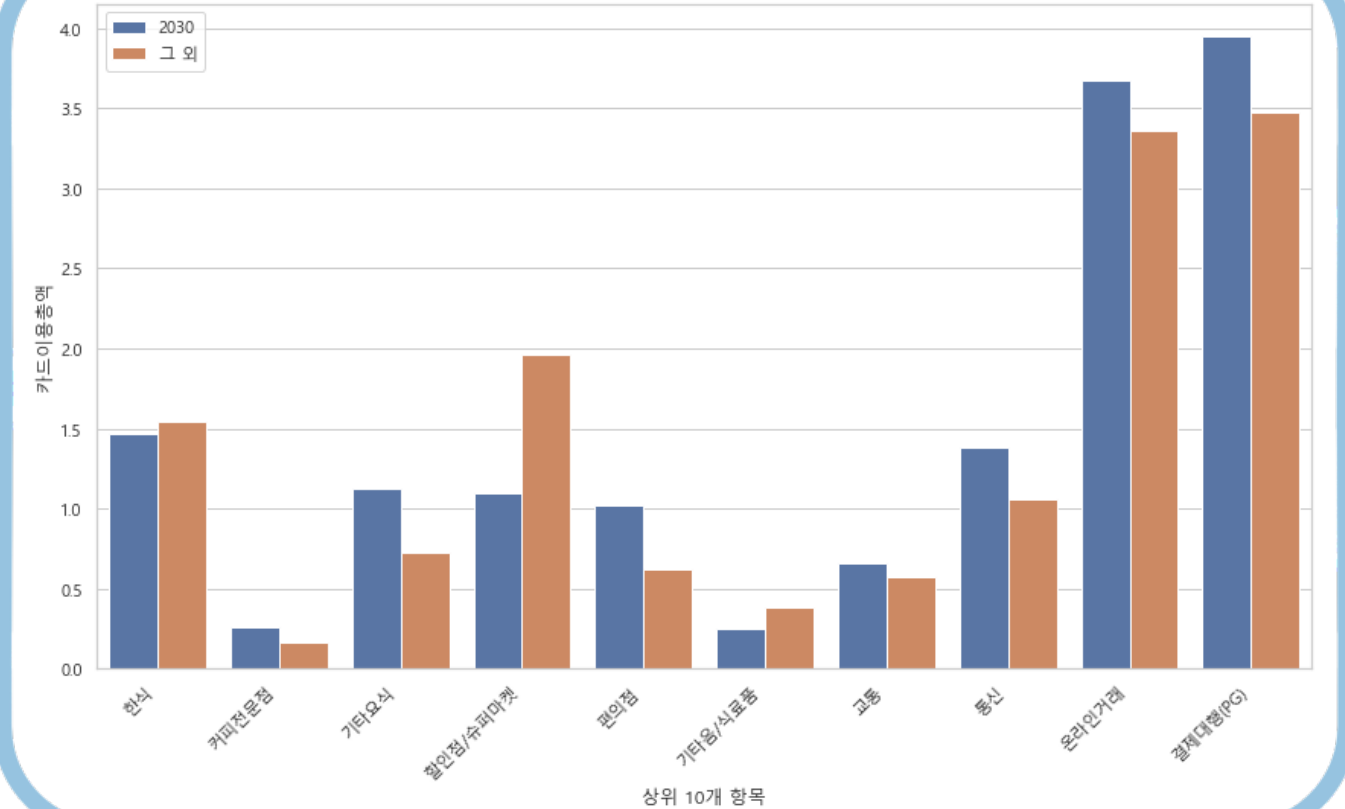
2030 vs 남은 연령대별 상위 10개 항목의 카드이용건수 비교



2030세대 → 편의점에서 카드사용이 가장 많음
반면 기타 음/식료품을 가장 적게 이용

타 연령대 → 편의점, 할인점/슈퍼마켓에서 카드사용이 가장 많음
반면 통신에 가장 적게 이용

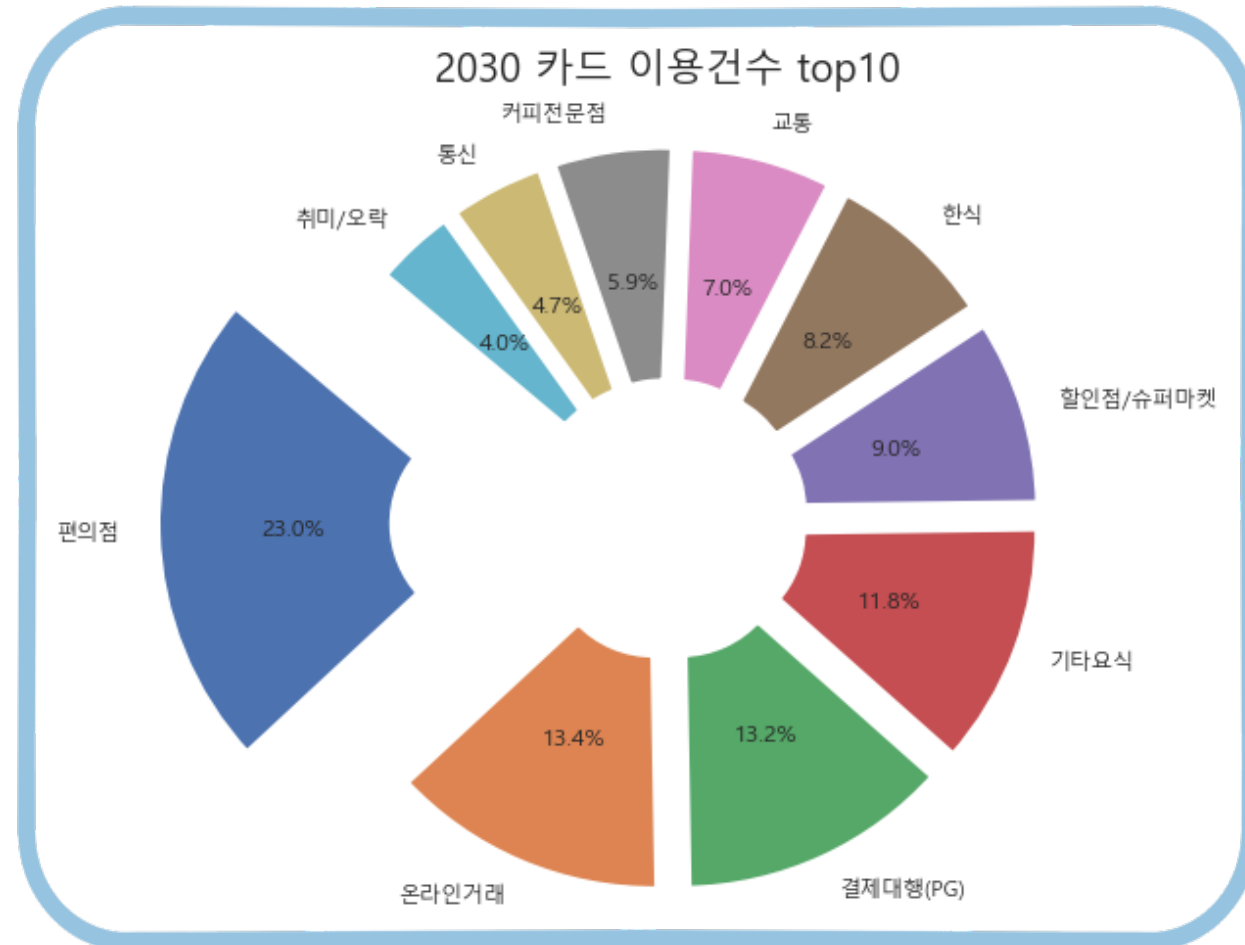
2030 vs 남은 연령대별 상위 10개 항목의 카드이용총액 비교



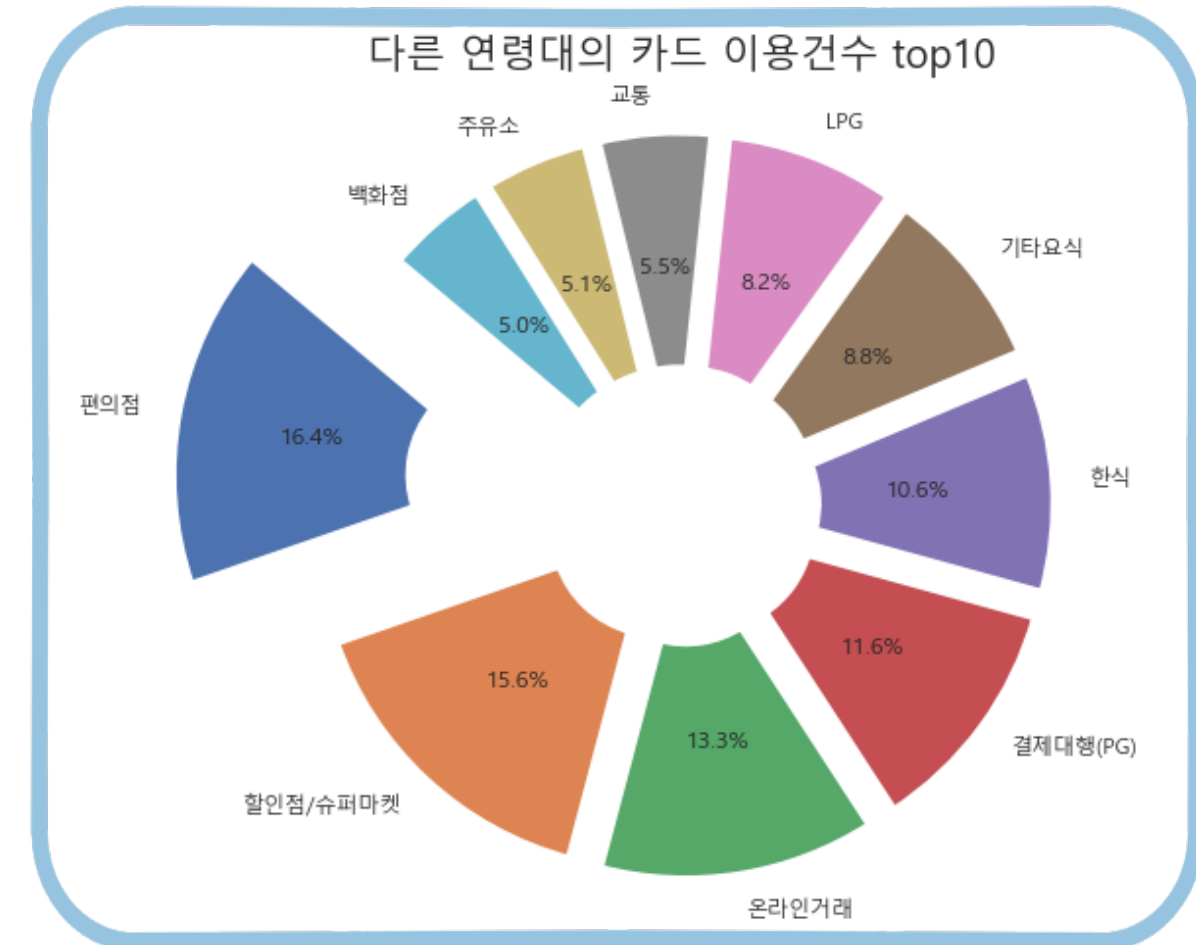
2030세대, 타연령대

→ 온라인거래, 결제대행에 대한 카드지출이 가장 많음
→ 반면 커피전문점, 기타음/식료품에 대한 카드지출이 가장 적음

2030 vs 타 연령대 결제 건수 비율



VS



이용건수에 대한 비율이 달라 순서가 다르긴 해도 TOP6내 소비항목은 두세대 모두 동일
 2030세대 → 취미/오락, 통신, 커피전문점의 상대적 이용비율 높음
 타 연령대(10대 제외) → 백화점, 주유소, LPG의 상대적 이용비율 높음



3. 소비예측 모델



예측모델 장단점

	RandomForest Regressor	LightGBM	ARIMA
장점	<ul style="list-style-type: none"> ▪ 다수의 의사결정 트리를 결합하여 과적합에 강함 ▪ 비선형 관계를 잘 처리함 ▪ 특성 스케일링이 필요 없음 ▪ 결측값을 다룰 수 있음 	<ul style="list-style-type: none"> ▪ 효율적이고 빠른 훈련 속도, 특히 대규모 데이터에 유용 ▪ 낮은 메모리 사용량 ▪ 범주형 특성에 대해 좋은 성능 ▪ 결측값을 자체적으로 처리함 	<ul style="list-style-type: none"> ▪ 단순하고 수학적인 이해가 쉬움 ▪ 시계열 데이터의 추세와 계절성을 고려하여 모델링 가능 ▪ 파라미터의 개수가 많지 않아서 해석이 비교적 간단 ▪ 보통 파라미터들이 자동적으로 최적화됨
단점	<ul style="list-style-type: none"> ▪ 대규모 데이터셋에서 계산 비용이 상당할 수 있음 ▪ 고차원이고 희소한 데이터에서 성능이 좋지 않을 수 있음 	<ul style="list-style-type: none"> ▪ 하이퍼파라미터를 제대로 튜닝하지 않으면 과적합 발생 가능 ▪ 전통적인 의사결정 트리에 비해 해석 가능성이 제한적임 	<ul style="list-style-type: none"> ▪ 계절성이 일정하게 반복되어야 한다는 가정이 있어야 함 ▪ 이상치와 극단값에 민감함 ▪ 긴 시간을 예측할 때는 성능이 떨어짐 ▪ 가정이 제한적 (정상성, 등분산 등) ▪ 비선형 트렌드나 복잡한 패턴, 다변수의 시계열 데이터 등에는 적합하지 않음

모델 적용 과정

하이퍼 파라미터 적용 전

	model score (train set)	model score (val set)
RF	0.95	0.67
LGBM	0.85	0.72

하이퍼 파라미터 적용 후

	model score (mse)	model score (train set)	model score (val set)
RF	0.99	0.95	0.66
LGBM	0.81	0.76	0.73

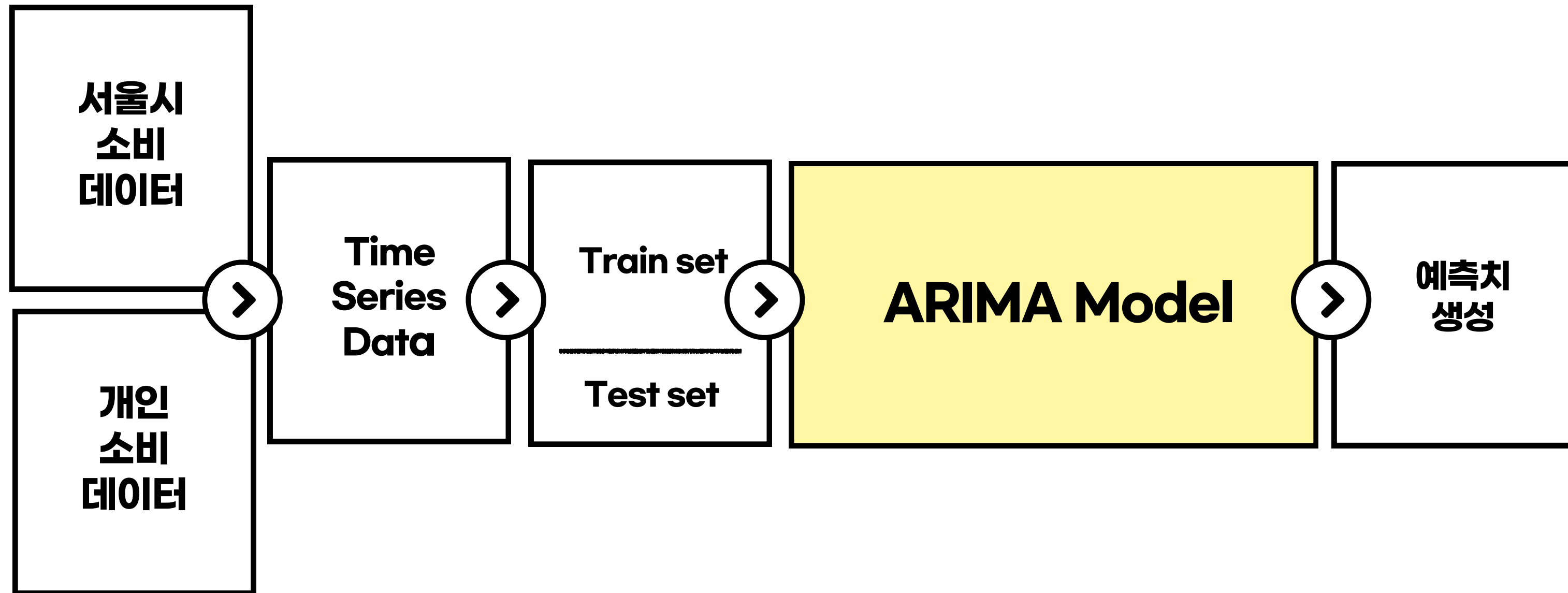
하이퍼 파라미터를 적용하였으나 모델 성능 지표가 개선되지 않음



단기간 예측력이 뛰어난 **ARIMA 모델**을 사용하기로 함

모델 적용 과정

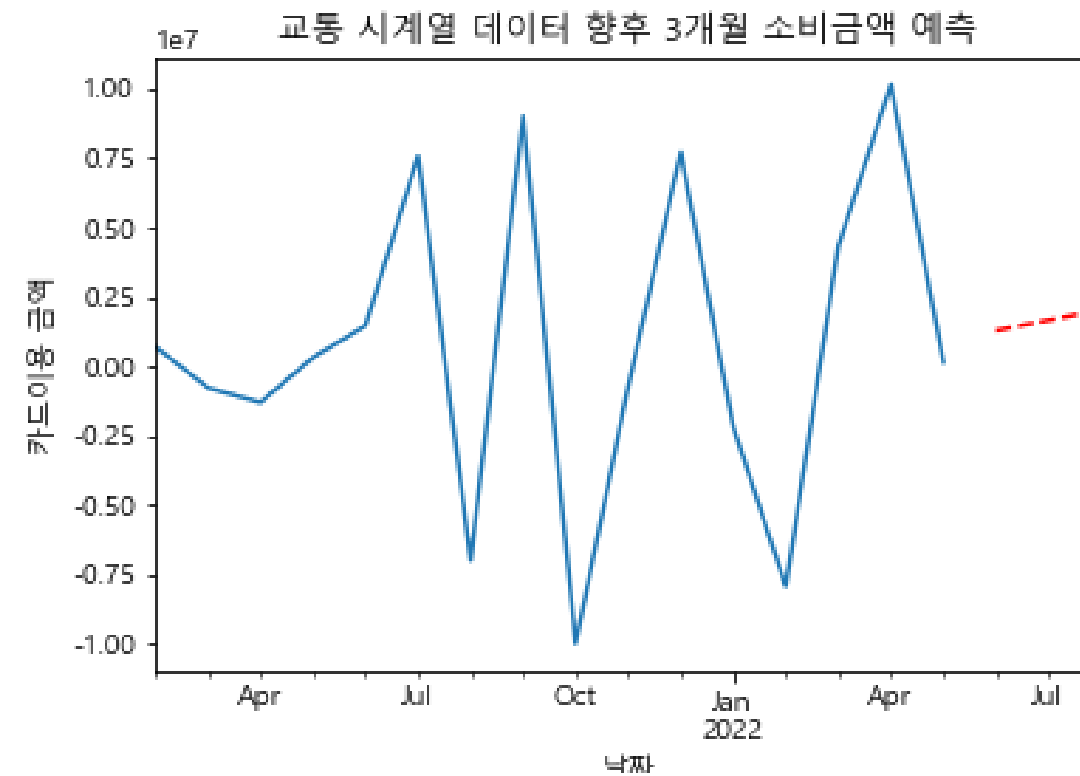
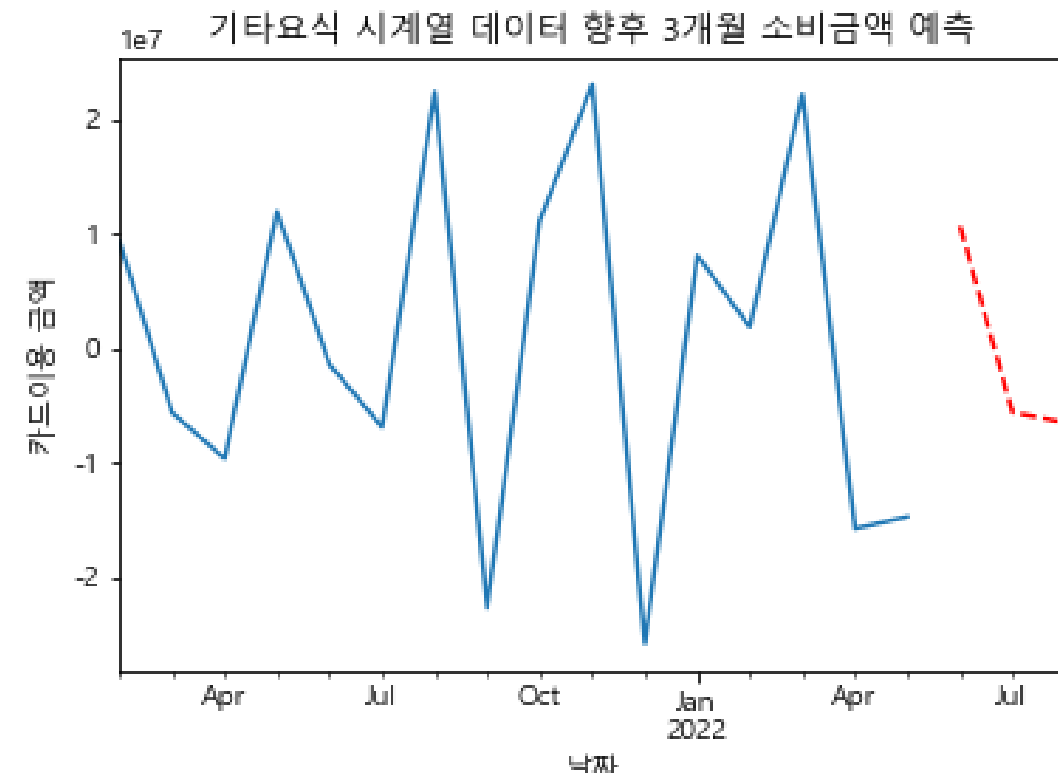
소비예측 모델 학습 과정

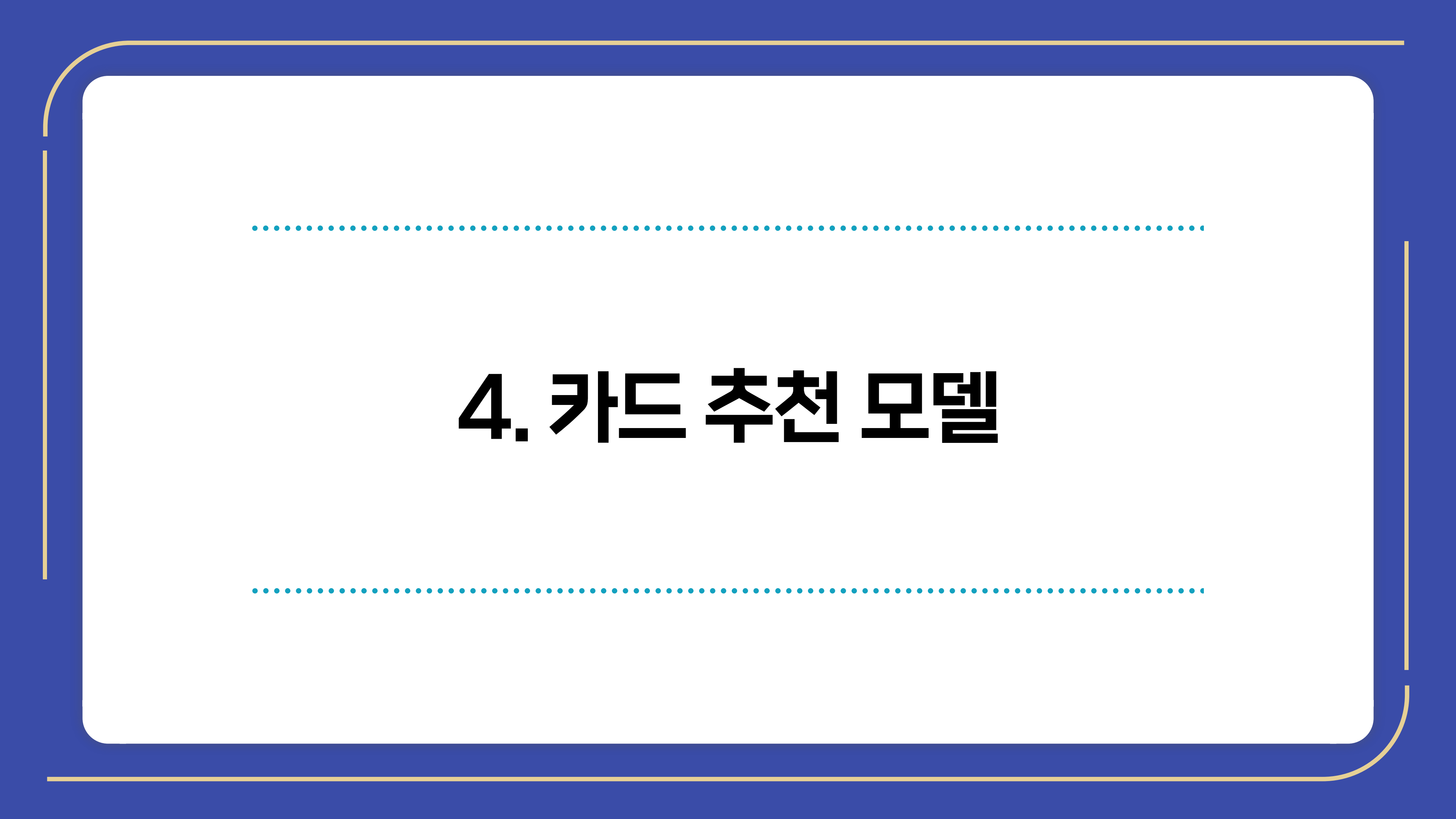


예측치 그래프


예측데이터 생성 (2022.06 - 2022.08, 3개월)

과거데이터와 예측데이터 비교 시, 시간 흐름에 따른 추세가 잘 반영됨.
학습데이터 부족으로 성능은 좋지 않은 편임.



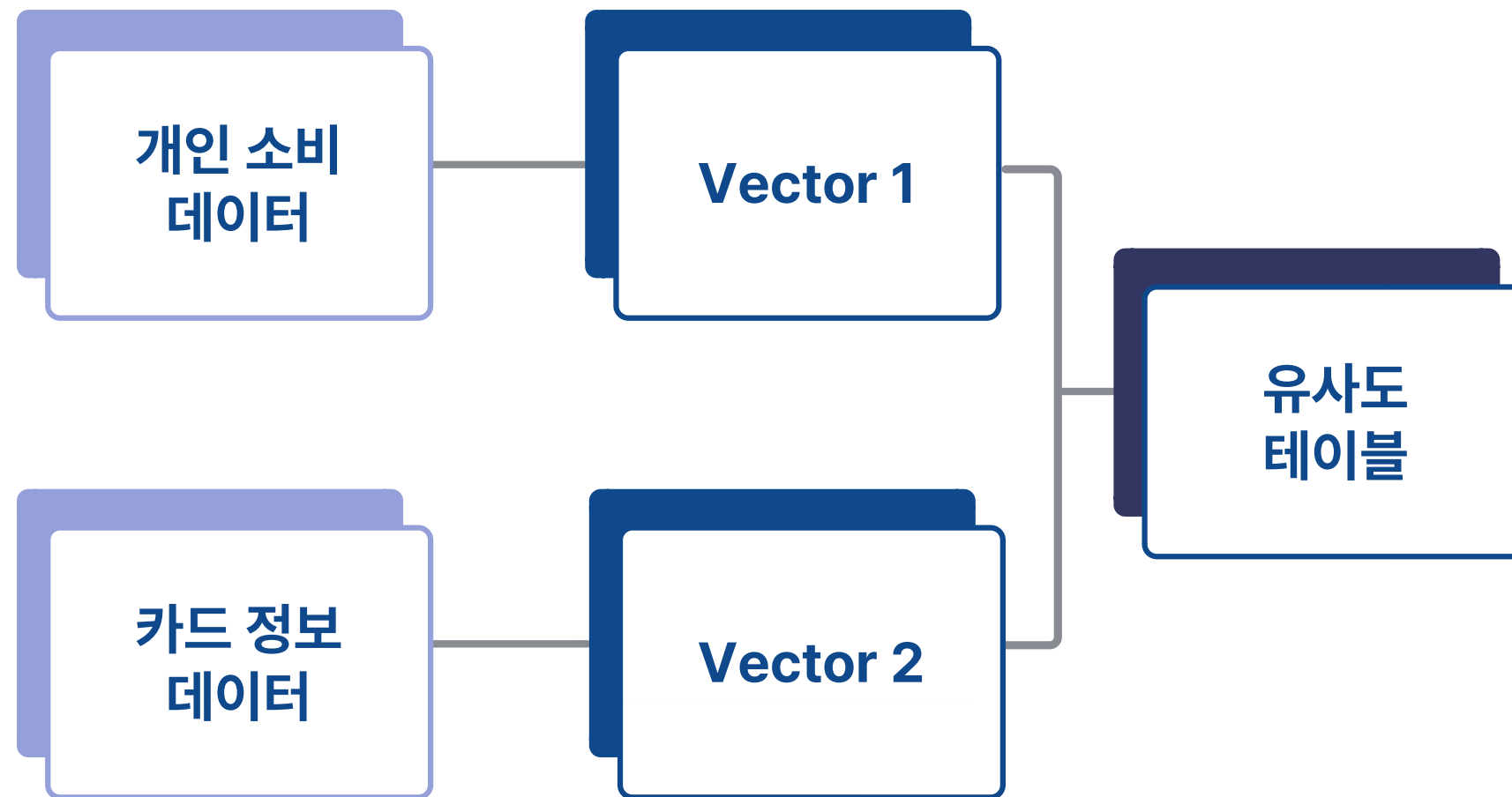


4. 카드 추천 모델



카드 추천 알고리즘

카드 추천 알고리즘



출처 : 비씨카드 특허 [카드 상품 추천 방법 및 카드사 서버]

- 업종 빈도 가중치(CF weight)를 사용하여 업종 간 빈도 차이 조절
- 업종별 카드 사용량에 따라 가중치 값이 증가하는 것을 방지
- 지역, 연령, 시간 등 복합적인 요소를 이용하여 카드 상품 추천 정확도 향상
- 고객 소비 형태를 벡터화/군집화하여 다양한 고객 간의 소비 유사도 파악

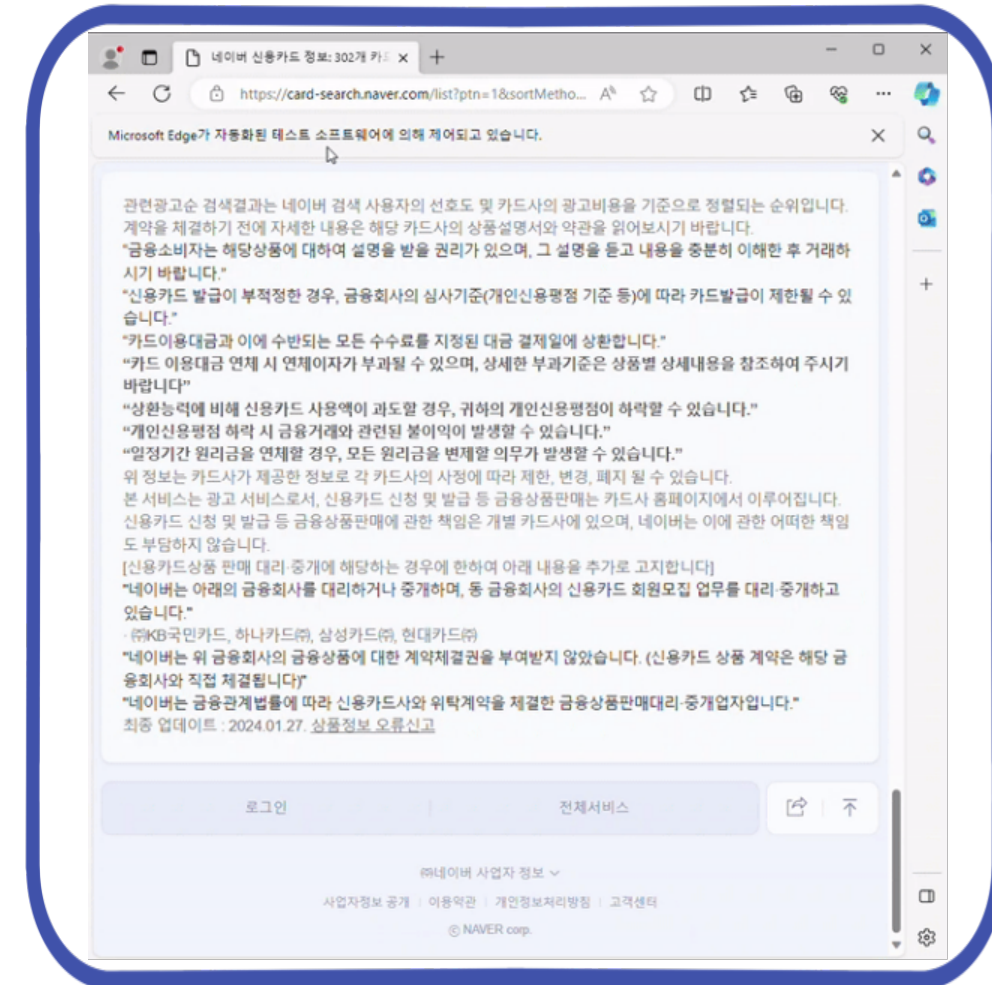
카드 혜택 크롤링

카드별 상세 페이지 url 추출

html 파싱

카드이름, 카테고리, 상세혜택 추출

csv 파일로 저장



[카드혜택.csv] shape 312 X 3

name

category

benefits

카드 추천 알고리즘

알고리즘 구성 요소

업종 빈도 가중치

$$CFweight = 1 + \log(CF)$$

역상품 빈도

$$IGF = \log \frac{N}{GF}$$

가중치

$$Weight = CFweight \times IGF$$

쿼리 길이

$$QueryLength = \sqrt{(\text{제1}Weight)^2 + (\text{제2}Weight)^2 + \dots + (\text{제}mWeight)^2}$$

정규화

$$Normalization = \frac{Weight}{QueryLength}$$

카드 혜택 크롤링

name	category	benefits
신한카드 The Classic+	['바우처', '항공마일리지', '주유', '카페/베이커리'...]	['쇼핑, 요식, 호텔 등 기프트옵션', '대한항공, 아시아나...]
NH농협 LCC UniMile 카드	[포인트/캐시백, '통신', '대중교통'..]	[저비용 항공사 LCC UniMile 적립 국내외가맹점 이용...]
LOCA Professional	['포인트/캐시백', '납부 혜택', '바우처',...]	['국내 모든 가맹점에서 1% 캐시백 즉시 자동결제 ..]
⋮		

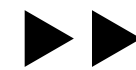
카드혜택.shape : 312 X 3

312개 카드의 카드명과 카드혜택 추출

알고리즘 결과

결제건수 기준 TOP7 소분류 추출

외식, 대중교통, 간편결제, 카페/베이커리, 편의점, 대형마트, 통신



2030 소비패턴 맞춤 카드 추천



KB국민 청춘대로 특특카드

카페/베이커리, 외식, 통신, 대중교통

스타벅스 최대 60% 할인
버거/패스트푸드 최대 30% 할인!
간편결제 이용시, 외식업종 10%할인
3대 통신사 통신요금 10% 할인
버스, 지하철, 택시 10% 할인



KB국민 특특Pay카드

간편결제, 대중교통, 편의점, 카페/베이커리

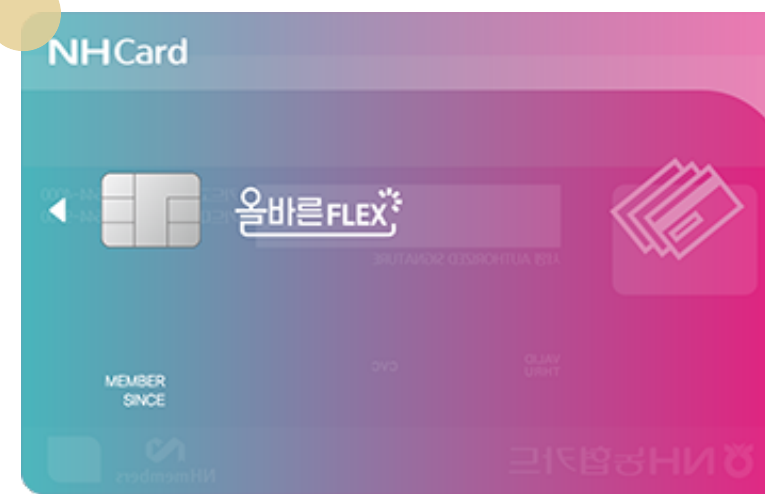
온/오프라인 간편결제 최대40%할인
버스, 지하철 이용시 10% 청구할인
GS25, CU 이용시10% 청구할인
간편결제 사용시 카페 최대 40%할인



스타벅스 현대카드

카페/베이커리, 쇼핑, 간편결제, 대중교통

카페 업종 이용 시 스타벅스 별 적립
쇼핑 업종 이용 시 스타벅스 별 적립
간편결제 이용 시 스타벅스 별 적립
대중교통 이용 시 스타벅스 별 적립



NH농협 올바른FLEX 카드

카페/베이커리, 외식, 문화, 쇼핑, 대중교통

스타벅스 50% 청구할인
스트리밍 서비스 20% 청구할인
배민, 요기요 10% 청구할인
온라인쇼핑 5% 청구할인
지하철, 버스 7% 청구할인



5. 결론

결론

결론 및 활용방안

- 빅데이터를 활용한 소비 예측 및 추천 서비스 도입
-> 기존 과거 데이터만 사용한 알고리즘에 비해 추천 성능이 향상될 것으로 기대됨
- 사용자 중심 : 소비패턴에 맞는 카드 추천을 통해 효율적인 소비 가능
- 마케팅 중심 : 현재 소비의 중심이 되는 2030대를 겨냥한 신상품 및 마케팅 기획
-> 신규 고객 유치
- 신용카드가 단순한 결제 수단을 넘어 소비자 라이프스타일과 조화를 이룬 편리한 서비스 도입

결론

한계점

- 소비데이터가 집단별(연령, 카테고리)로 이미 grouping 되어 2030 소비 특성을 분석하는데 한계가 존재함
- 개인정보가 많이 포함되어 있는 카드 데이터의 특성 상 사용가능한 변수가 제한적임
- 카드혜택 금액 한도, 실적금액에 따른 혜택 차등 적용 등의 세부사항 까지 고려하지 못함
- 분석 환경의 한계로 다양한 모델 구현 및 파라미터 조정이 어려움

개선방안

- 소비자 행동 변화를 수치화 한 소비자 물가지수, 뉴스기사 등의 경제심리보조지수와 같은 지표를 컬럼으로 추가
- 수집기간 늘리기

이상 발표를 마치겠습니다

감사합니다