



수원시 카드 소비데이터 분석 기반 정책 제안

소비 패턴 분석 및
맞춤형 정책 방향성 제시

2026. 01. 07.

연금술 팀 - 백승규, 서준혁, 임동현, 이예준

Contents

1. 프로젝트 배경 및 목적
2. 원본 데이터셋 소개
3. 데이터 전처리
4. EDA 및 최종 데이터셋
5. 분석 설계 및 모델링
6. 분석 결과
7. 결론 및 정책 제언
8. 분석의 한계 및 보완점

| 1. 프로젝트 배경 및 목적

프로젝트 배경 및 목적

실증 데이터 기반 정밀 정책 타겟팅의 필요성

! 현황 (As-Is)

기존 복지 정책은 소득·연령 등 정량적 지표 중심의 보편적 접근 방식을 채택하고 있습니다. 이는 실제 생활 양식과 소비 행태의 다양성을 반영하지 못해 정책 사각지 대를 발생시키는 한계가 있습니다.

◎ 목표 (To-Be)

실증적 소비 행태 분석을 통해 잠재적 위험군을 구체적으로 식별하고, 자원의 효율적 배분과 선제적 정책 개입을 위한 과학적 근거를 마련합니다.

"무엇을(What) 소비하는가"



"어떻게(How) 지원할 것인가"

| 2. 원본 데이터셋 소개

원본 데이터셋 소개

	A	B	C	D	E	F	G	H	I	J	K	L
1	ta_ymd	cty_rgn_no	admi_cty_no	card_tpbus_c	card_tpbus_n	card_tpbus_n	hour	sex	age	day	amt	cnt
2	20251001	41111	41111560	D01	소매/유통	가전제품	3	F	6	3	379702	2
3	20251001	41111	41111560	D01	소매/유통	가전제품	3	M	5	3	60662	2
4	20251001	41111	41111560	D01	소매/유통	가전제품	6	F	6	3	128676	2
5	20251001	41111	41111560	D02	소매/유통	건강/기호식품	3	F	7	3	98706	4
6	20251001	41111	41111560	D02	소매/유통	건강/기호식품	4	F	6	3	798682	2
7	20251001	41111	41111560	D02	소매/유통	건강/기호식품	4	F	8	3	172268	2
8	20251001	41111	41111560	D02	소매/유통	건강/기호식품	4	M	5	3	186800	6
9	20251001	41111	41111560	D02	소매/유통	건강/기호식품	4	M	7	3	256827	9
10	20251001	41111	41111560	D02	소매/유통	건강/기호식품	4	M	8	3	441118	2

첨부된 문서

[카드소비 데이터_202511.zip](#)

다운로드

01 : 00:00 ~ 06:59

01 : 0-9세

01 : 월요일

[카드소비 데이터_202510.zip](#)

다운로드

02 : 07:00 ~ 08:59

02 : 10-19세

02 : 화요일

[카드소비 데이터_202509.zip](#)

다운로드

03 : 09:00 ~ 10:59

03 : 20-29세

03 : 수요일

[카드소비 데이터_202508.zip](#)

다운로드

04 : 11:00 ~ 12:59

04 : 30-39세

03 : 목요일

[카드소비 데이터_202507.zip](#)

다운로드

05 : 13:00 ~ 14:59

05 : 40-49세

04 : 금요일

[카드소비 데이터_202506.zip](#)

다운로드

06 : 15:00 ~ 16:59

06 : 50-59세

04 : 토요일

07 : 17:00 ~ 18:59

07 : 60-69세

05 : 일요일

08 : 19:00 ~ 20:59

08 : 70-79세

06 : 토요일

09 : 21:00 ~ 22:59

09 : 80-89세

07 : 일요일

10 : 23:00 ~ 23:59

10 : 90-99세

08 : 토요일

11 : 100세 이상

11 : 100세 이상

09 : 일요일



카드 소비 데이터

민간데이터 카드 : 카드 소비 데이터

총평점 ★★★★★ OK



-민간 카드사의 소비 데이터를 바탕으로 시군구 및 행정동 단위에서 업종별 소비 패턴을 분석할 수 있도록 구성된 자료

-주요 항목: '기준년월일', '시군구코드', '행정동코드', '카드사 업종분류코드',
'카드사 업종대분류명', '카드사 업종중분류명', '시간대', '성별', '연령별', '요일', '매출금액', '매출건수'

| 3. 데이터 전처리

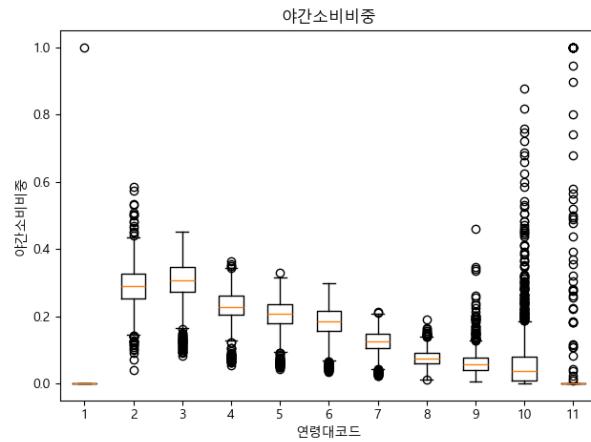
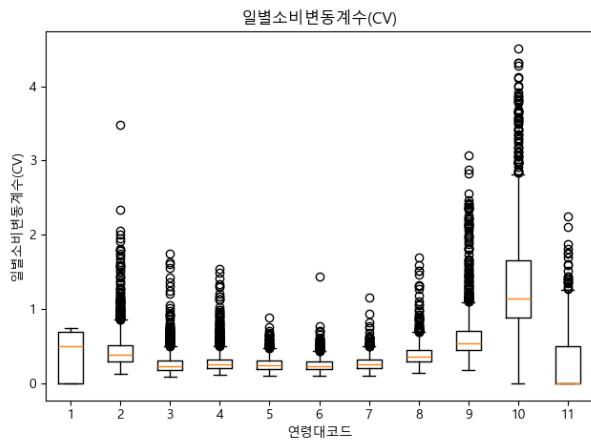
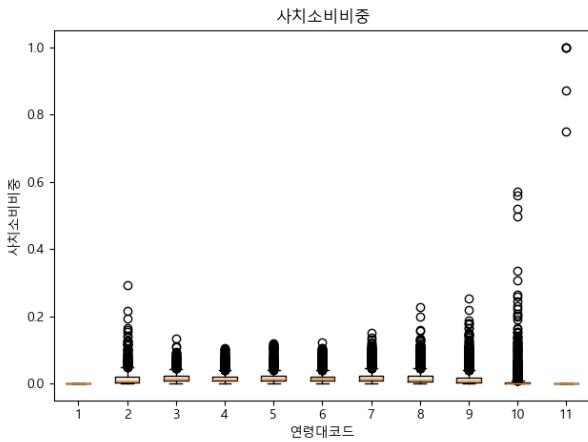
데이터 전처리

소비 카테고리 재분류	월 단위 집계(Grain 통일)	시간/맥락 파생변수 생성	소비의 질 • 안정성 파생변수 생성
<ul style="list-style-type: none">• 80여 업종 중분류를 성격별로 통합• 카테고리별 소비 비중 계산<ul style="list-style-type: none">- 필수 지출 비중(Essential share)- 선택 지출 비중(Optional share)- 사치 지출 통합비중(Luxury_combined)	<ul style="list-style-type: none">• 기준 단위로 잡계• 연월 X 행정동 X 연령대• amt / cnt 합계 및 카테고리별 합계	<ul style="list-style-type: none">• 야간소비비율(night_ratio)<ul style="list-style-type: none">◦ 야간 시간대 소비액 / 전체 소비액• 주말소비비율(weekend_ratio)<ul style="list-style-type: none">◦ 주말(토·일) 소비액 / 전체 소비액	<ul style="list-style-type: none">• 업종 집중도 HHI<ul style="list-style-type: none">◦ $HHI = \sum p_i^2$ (p_i = 업종별 소비 비중)• 업종 다양성 엔트로피<ul style="list-style-type: none">◦ $-\sum p_i \log(p_i)$• 일별 소비 변동계수 (daily_cv)<ul style="list-style-type: none">◦ 월 내 일별 소비액의 표준편차 / 평균

전체 경기도 데이터 수가 매우 많아 '수원시' 데이터만을 사용함.

| 4. EDA 및 최종 데이터셋

EDA



EDA 단계에서 연령대별 주요 지표(일별변동계수(CV), 야간비중, 사치소비비중 등)의 박스플롯을 확인한 결과, 연령대 1(10세 미만)과 연령대 9~11(80세 이상) 구간에서 지표 분포의 분산이 과도하게 커지고 극단치(outlier)가 집중되는 현상이 관찰되었다.
→ 해당 구간은 거래량/총액이 작아 지표가 쉽게 튀는 희소성 영향이 클 수 있어, 모델링에서는 연령대 2~8(10~79세)만 사용했습니다.

최종 데이터셋

“행정동×연령대별 소비 패턴을 특징화한 최종 Feature Table”

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	기준연월	행정동코드	연령대	총지출액	총지출액로그	필수지출비중	선택지출비중	사치지출통합비중	야간소비비율	주말소비비율	업종집중도HHI	업종다양성엔트로피	일별변동계수CV	이상치여부
2	Jan-22	41111560	10-19세	21959154	16.90469469	0.607155267	0.350851449	0.032924629	0.251238367	0.390960462	0.154279077	2.567577959	0.489772687	0
3	Jan-22	41111560	20-29세	1145335679	20.8589636	0.773230542	0.138968608	0.021488105	0.258945356	0.339673486	0.1669695	2.476974262	0.246674646	0
4	Jan-22	41111560	30-39세	3011619013	21.82574365	0.784655351	0.131809662	0.011153323	0.195526455	0.342103377	0.183668271	2.383529002	0.133104052	0
5	Jan-22	41111560	40-49세	4995663699	22.33183611	0.798995103	0.123077133	0.015914262	0.165763771	0.323877812	0.181515047	2.40142007	0.148595304	0
6	Jan-22	41111560	50-59세	6221947770	22.55134884	0.812957056	0.093312823	0.015873106	0.144811351	0.33729724	0.171481059	2.397009092	0.174826497	0
7	Jan-22	41111560	60-69세	3722122585	22.03755993	0.799485873	0.089599836	0.013371459	0.098128375	0.304662855	0.164490332	2.382455157	0.215480745	0
8	Jan-22	41111560	70-79세	896466389	20.61397136	0.794369144	0.101471307	0.014364833	0.052680062	0.279050359	0.14805514	2.413990698	0.293988049	0
9	Jan-22	41111560	80-89세	119038894	18.59496085	0.755198591	0.136929355	0.013853691	0.034082054	0.248896499	0.124374196	2.49703539	0.464653961	0
10	Jan-22	41111566	10-19세	61791079	17.93926958	0.370723483	0.506221748	0.09788939	0.394459029	0.334193436	0.078307713	2.88276257	0.330190562	0
11	Jan-22	41111566	20-29세	2585203843	21.6730702	0.489759796	0.431457294	0.041988313	0.385528166	0.304500482	0.085579944	2.996773672	0.146822155	0
12	Jan-22	41111566	30-39세	2459436173	21.62319796	0.544393821	0.376445465	0.02558741	0.287574181	0.286599774	0.087124834	3.053225902	0.233743284	0
13	Jan-22	41111566	40-49세	2842598306	21.76798437	0.560249109	0.354767547	0.014965748	0.223635794	0.287437518	0.086366836	3.110580995	0.200515262	0
14	Jan-22	41111566	50-59세	2818187886	21.75935992	0.616658257	0.292472097	0.016462947	0.213215037	0.295633213	0.099351763	2.976745116	0.171950065	0
15	Jan-22	41111566	60-69세	1309030382	20.99255253	0.659880634	0.249794187	0.016139257	0.131617984	0.272358846	0.113251309	2.821182414	0.228698165	0

| 5. 분석 설계 및 모델링

분석 설계 (Methodology)

소비 행태 기반 군집 분석 프레임워크

INPUT

- 분석데이터

전처리된 수원시 카드 소비 데이터

- 업종별 지출구조

- 결제시간대 분포

- 카드 사용 지역 분포

PROCESS

- 핵심방법론

Isolation Forest & 군집 분석

- 소비 구조 변수 활용

- 시공간 특성 반영

- 다차원 패턴 인식

OUTPUT

- 분석결과

3개 핵심 군집 도출

- 이질적 특성 규명

- 정책 대상 재정의

모델링

2단계의 모델과 정확성 검증

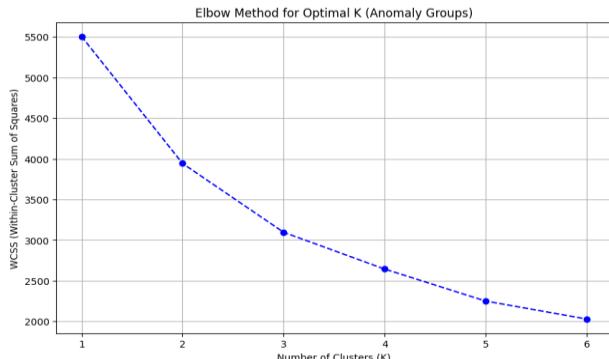
Isolation Forest

- 소비 데이터에서 상위 5%의 이상 징후를 포착

- 금융당국의 관리 지표인 한국은행 취약차주 비중(약 5%)을 모델 임계치로 설정하여, 거시 경제 지표와 정합성을 갖춘 실질적 위기 집단을 도출

K-means clustering

- 지표 간의 유사성을 계산하여 그룹화



Silhouette Score

- 군집화가 얼마나 잘되었는지 측정

- silhouette score = 0.3296
- 군집화가 잘 이루어졌다고 판단 가능

모델링 결과

- 3개의 그룹별 특징 도출

- 그룹 A : 생계 압박형 - 필수재 비중 79%인 5~60대
- 그룹 B : 상권 노출형 - 선택제 비중 44%와 높은 사치인 7~80대
- 그룹 C : 야간 활동형 - 야간소비가 19%로 가장 높은 2~40대

| 6. 분석 결과

군집 A

생계 압박형

소비가 생존 유지 수준에 집중된 집단입니다.



79%

필수재(식료품/생필품) 소비 비중

50-60대 연령층 집중

팔달구 우만동 일대 높은 비중

군집 특징 및 해석

군집 특징

수입의 대부분을 먹고사는 데(식비, 생필품 등)
만 써야하는 집단

해석

비합리적 과소비보다는 경제적 여력 부족으로
인한 소비 다양성 결핍이 위험 신호로 나타난
케이스

정책 방향

긴급 생계 지원

- 필수재 물가 안정 정책
- 구매 바우처 제공

고객 보호 프로그램

- 유통사의 유통기한 임박 상품 할인 정보 제공
- 금융사의 소액 연체 채무 조정 프로그램 연계

군집 B: 상권 노출형

장소 효과(Place Effect)에 의한 소비 취약 고령층

70-80대 연령층 집중

광교, 행궁동 일대 높은 비중

Definition

상업 인프라 밀집 지역 거주·체류로 인해
비필수재 소비에 빈번하게 노출된 고령층 집단

소비 구조 동일 연령대 타 지역군 대비 쇼핑, 유흥 등 지출 비중 상대적 우위

인구 통계 고령층 집중

공간 특성 인계동, 광교 등 상업·쇼핑 지구 집중

Core Interpretation

개인의 과소비 성향이 아닌, 소비 접근성이 높은 환경 요인,
즉 '장소 효과(Place Effect)'가 소비 규모를 결정

정책 방향

정부 (Government)

소비자 보호 강화

- 상권 내 고령층 대상 금융·소비 상담 지원
- 디지털 정보 격차 해소를 위한 가격 비교 교육

기업 (Corporate)

고객 보호 프로그램

- 금융/카드사의 고액 결제 시 보호자 알림 서비스
- 유통사의 합리적 소비 안내 서비스 개발

군집 C: 야간 활동형

Night-time Active Group

야간·교대 근무 등 특정 노동 구조로 인해
심야 시간대 편의 중심의 소비 패턴을 보이는 집단

주요 특징

- 📍 영통일대 높은 비중
- 👤 인구 통계
20-40대 생산가능인구 집중
- 🕒 시간 특성
저녁 ~ 심야 시간대 집중

핵심 해석

배달, 편의점, 심야 외식, 교통비 등에
집중된 소비구조 보임

정책 방향

정부 (Government)

시간대 맞춤형 복지
심야 교통비 지원, 야간 근무자 대상 건강검진 확대 및 식생
활 개선 프로그램

기업 (Corporate)

노동 환경 연계 서비스
플랫폼/유통사의 야간 할인 프로모션

| 7. 결론 및 정책 제언

결론 및 정책 제언

데이터 기반 선제적 예측 시스템으로의 전환

소비 데이터 분석을 통해 기존 정책 지표로 포착하기 어려운 **3개의 군집**을 성공적으로 식별하였습니다.

이는 획일적 지원을 넘어, 각 군집의 핵심 원인(생계, 환경, 노동)에 기반한 정밀 개입의 필요성을 증명합니다.

01

정책 패러다임 전환

사후 대응 중심의 복지 체계에서
데이터 기반의 선제적 위기 예측 및 예방 시스템으로
근본적인 전환이 필요합니다.

02

민관 협력 모델 구축

공공 데이터와 민간 카드 데이터를 결합한
'위기 예측 거버넌스'를 구축하여
데이터의 정합성과 정책의 실효성을 제고해야 합니다.

03

파일럿 프로젝트 추진

특정 군집이 집중된 지역(예: 팔달구 우만동)을 대상
으로 정책 효과성 검증을 위한
시범 사업(Pilot Project) 착수를 제안합니다.

| 8. 분석의 한계 및 보완점

분석의 한계 및 보완점

데이터 정밀도 제고를 위한 향후 과제

01

세대 간 카드 혼용에 따른 식별 한계

10, 20대들의 부모님 카드 사용 패턴이 존재함에 따라, 데이터상의 카드 소유주와 실제 소비 주체가 일치하지 않는 경우가 발생합니다. 이는 특정 연령대의 소비 행태를 정밀하게 분리하여 분석하는데 제약 요인으로 작용합니다.

02

데이터 노이즈 및 이상치 관리

일시적인 대액 결제나 특수 업종에서의 결제 등 **이상치(Outlier)**로 인한 데이터 튜닝 현상이 관찰됩니다. 향후 보다 정교한 이상치 탐지 및 보정 알고리즘을 적용하여 통계적 왜곡을 최소화할 필요가 있습니다.

03

시계열 데이터의 확장성 확보

현재 분석은 **2022년부터 축적된 데이터**를 기반으로 하고 있어, 장기적인 소비 트렌드 변화를 포착하는데 한계가 있습니다. 향후 더 긴 기간의 데이터를 확보한다면 계절성 및 거시경제 변화를 반영한 더욱 정확한 분석이 가능할 것으로 기대됩니다.

질의응답

경청해 주셔서 감사합니다.

백승규, 서준혁, 임동현, 이예준

