

여가 불평등 해소를 위한 AI 여가 제약 인식 솔루션 : K-Means로 발굴한 고제약 집단(C0) 중심

부제: 행동 장벽 완화를 위한 2-Track 제언

혼자 품DA / 이연지

목차

CONTENTS

01

서론

문제 배경 및 가설 설정

02

분석 설계

전체 분석 프레임워크, 사용 데이터 및 핵심 변수 정의

03

분석 결과

초기 가설 기각과 '행동 장벽'의 발견 과정

04

솔루션 제언

'행동'을 바꾸는 2-Track 솔루션 및 기대효과

05

결론

최종 결론 및 핵심 발견 내용 요약

06

부록

OLS 상세 결과, solo-gap 그래프

1-1. 문제의 심각성: 양이 아닌 '질'의 여가 불평등

여가 시간	개선 4.3시간 (2024)
1인당 여행 일수	악화 8.54 일/연 (2024)
여가생활만족도	개선 39.4% (2025)
문화여가지출률	개선 5.51% (2024)
여가시간 충분도	개선 65.1% (2024)
문화예술 및 스포츠 관람횟수	개선 7.0 회/연 (2023)

출처: 지표누리 – 국민 삶의 질 지표

표면적으로, 2024년 기준 '여가 시간'(4.3시간), '여가시간 충분도'(65.1%) 등 양적(Quantitative) 지표는 양호

하지만, '여가 생활 만족도'는 39.4%에 불과하며, 적극적인 여가 활동의 지표인 '1인당 여행일수'(8.54일)은 양적 지표로도 부족함.

Key Message

시간은 있지만, 질적으로 만족하지 못하는 '질적 여가 불평등'이 핵심 문제입니다.

1-2. 가설 설정

'질적 만족도'의 핵심은 '관계'이므로, '여가를 혼자 즐기는(solo)' 것이 만족도를 낮출 것이라는 직관적인 가설을 설정하였습니다.

1차 가설 설정

'여가를 혼자 즐기는(Solo) 사람'이 '함께 즐기는(social) 사람'보다 만족도가 낮을 것이다.

(본 보고서의 'solo'는 '1인 가구'가 아닌 '주요 여가 활동 동반자=혼자'인 사람을 의미합니다.)

=> 만약 '혼자'라는 현상(what)이 '질적인 여가 불평등'의 원인이 아니라면?

2차 가설 설정

질적 불평등을 초래하는 것은 '동반자 없음'이 아니라 '삶의 복합적인 제약'이다.

2-1. 전체 연구 프레임워크

Step 1. 1차 가설 검증

Solo vs social
(Mann-Whitney U)

Step 2: 핵심 집단 발굴

'제약' 기반 재분류
(K-means)

Step 3: 장벽 랭킹

'행동' 원인 규명
(OLS)

02 분석 설계

2025년 문화체육관광 통계데이터 활용대회

2-2. 사용 데이터 및 핵심 변수 정의

사용 데이터: 분석의 신뢰도를 위한 데이터

핵심 변수 정의: 1차 가설(X1)과 2차 가설(X2)을 '6대 요인(X3)'으로 통제하여 검증

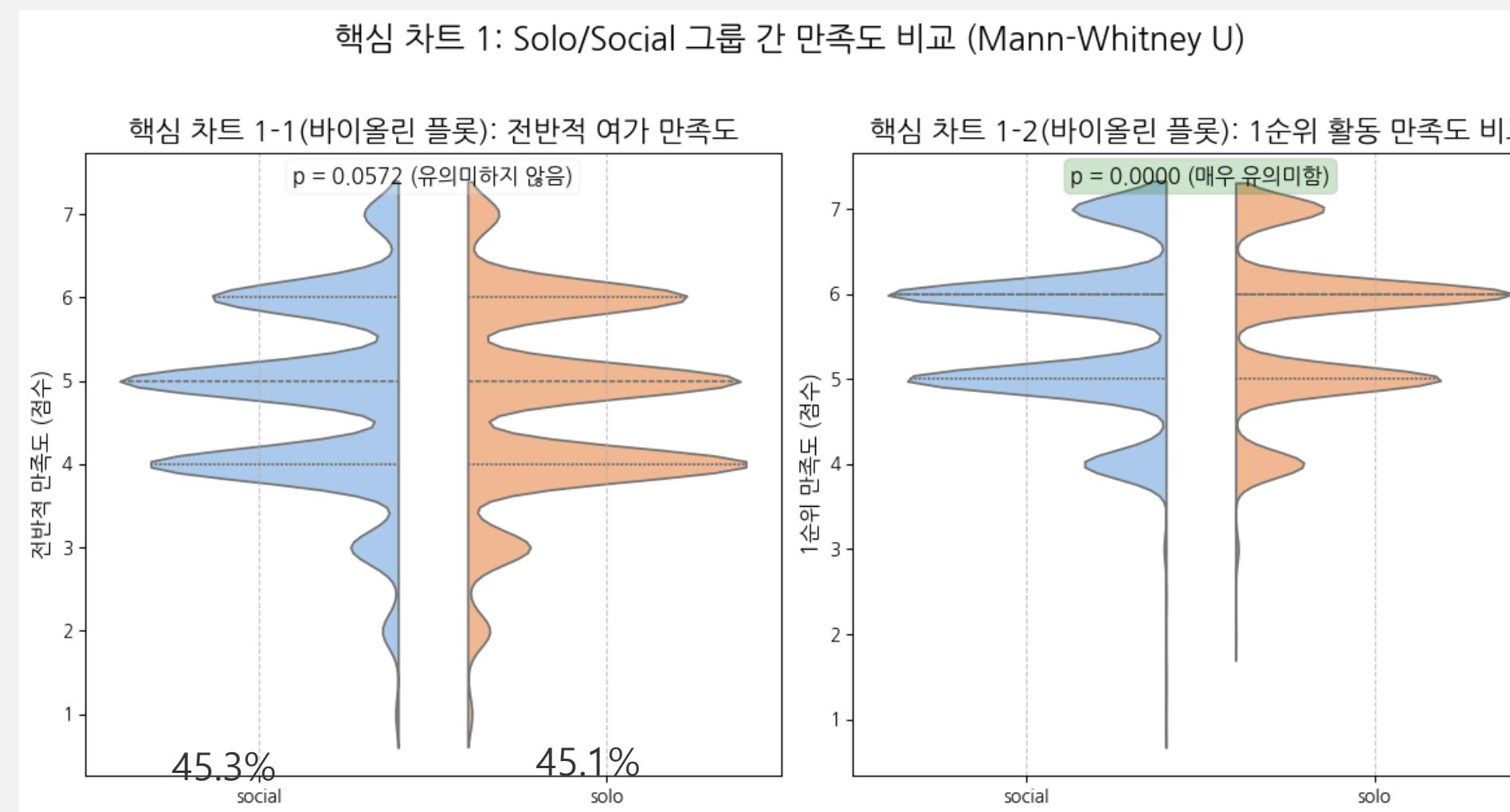
구분	데이터명	역할
핵심	국민여가활동조사	모델링 및 클러스터 정의
보조	국민문화예술활동조사	메인 데이터 분석 내용 교차검증

변수 그룹	역할	핵심 변수
Y (목표)	(행동 대리 지표)	만족도_참여1순위
X1 (1차 가설)	가설 검증 대상	Solo_여부 (동반자)
X2 (2차 가설)	핵심 발견 변수	- '제약_질병장애', '제약_고립' 등 '제약_' 9종 - '인식_사회적 관계' 등 '인식_' 3종
X3 (통제 변수)	신뢰도 확보 (6대 요인)	'소득_가구', '나이', '성별', '주당_평균_근무시간' 등

03 분석 결과

2025년 문화체육관광 통계데이터 활용대회

3-1. 1차 가설 기각 - '혼자'라는 현상은 원인이 아니었다.



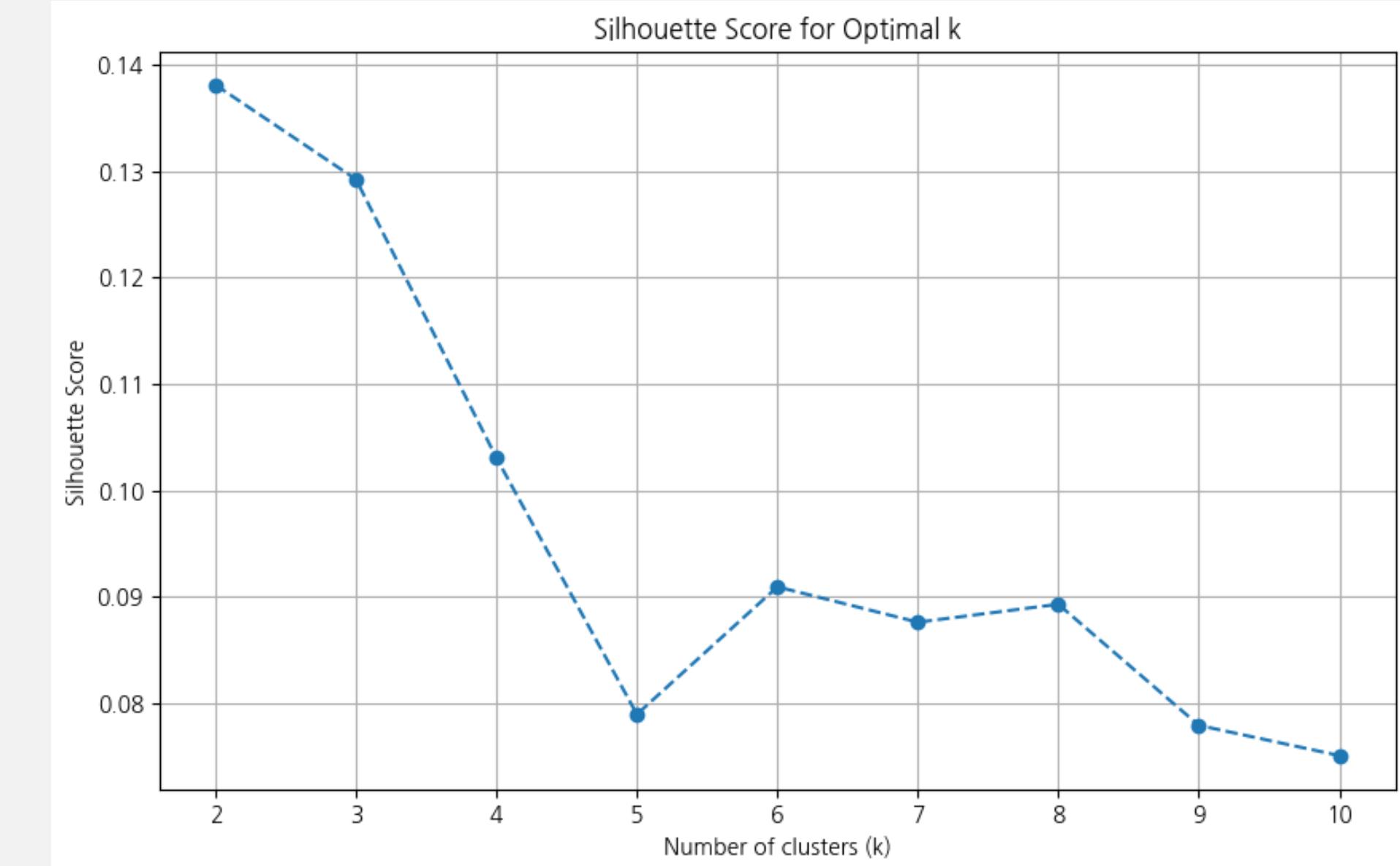
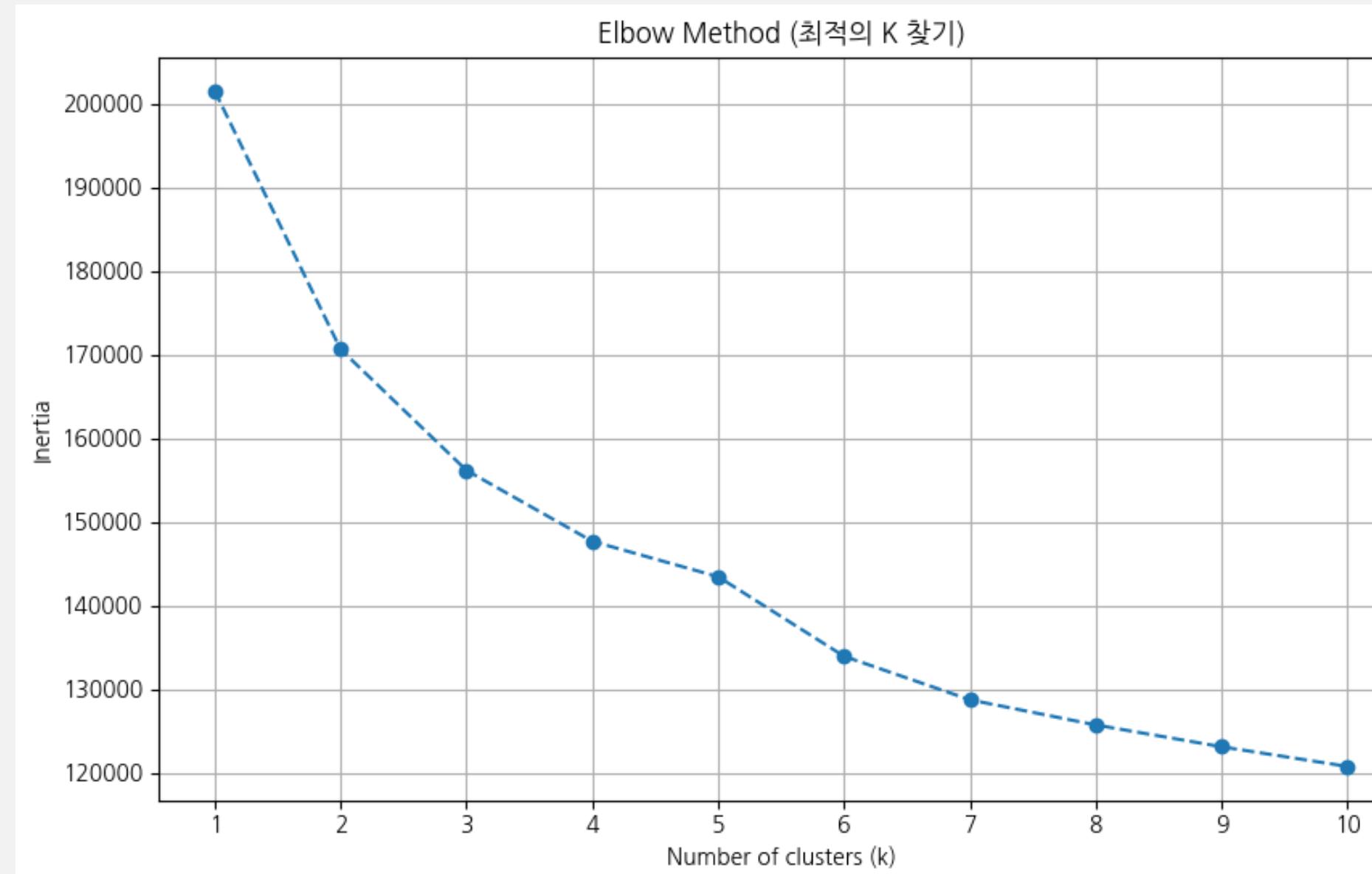
- '전반적 만족도' ($p=0.057$): Mann-Whitney U 검정 결과, 두 집단의 만족도 차이는 통계적으로 유의미하지 않았습니다.
- '1순위 활동 만족도' ($p < 0.001$)는 통계적으로 유의미했으나, 두 집단의 평균/중앙값 차이(효과 크기)가 실질적으로 매우 작아, 정책적/실무적 의미는 제한적임을 확인했습니다.

Key Message

1차 가설 기각 : 두 핵심 지표 모두에서 'Solo' 여부는 **만족도의 핵심 구동 요인(Key Driver)**이 아님을 확인했습니다. => 2차 가설: '근본 원인(제약)'을 확인해야 합니다.

3-2. K-Means (1): 어떻게 집단을 구분하였는가?

2차 가설 검증을 위해 '제약 변수' 9종을 투입하여 K-Means를 실행하였습니다.

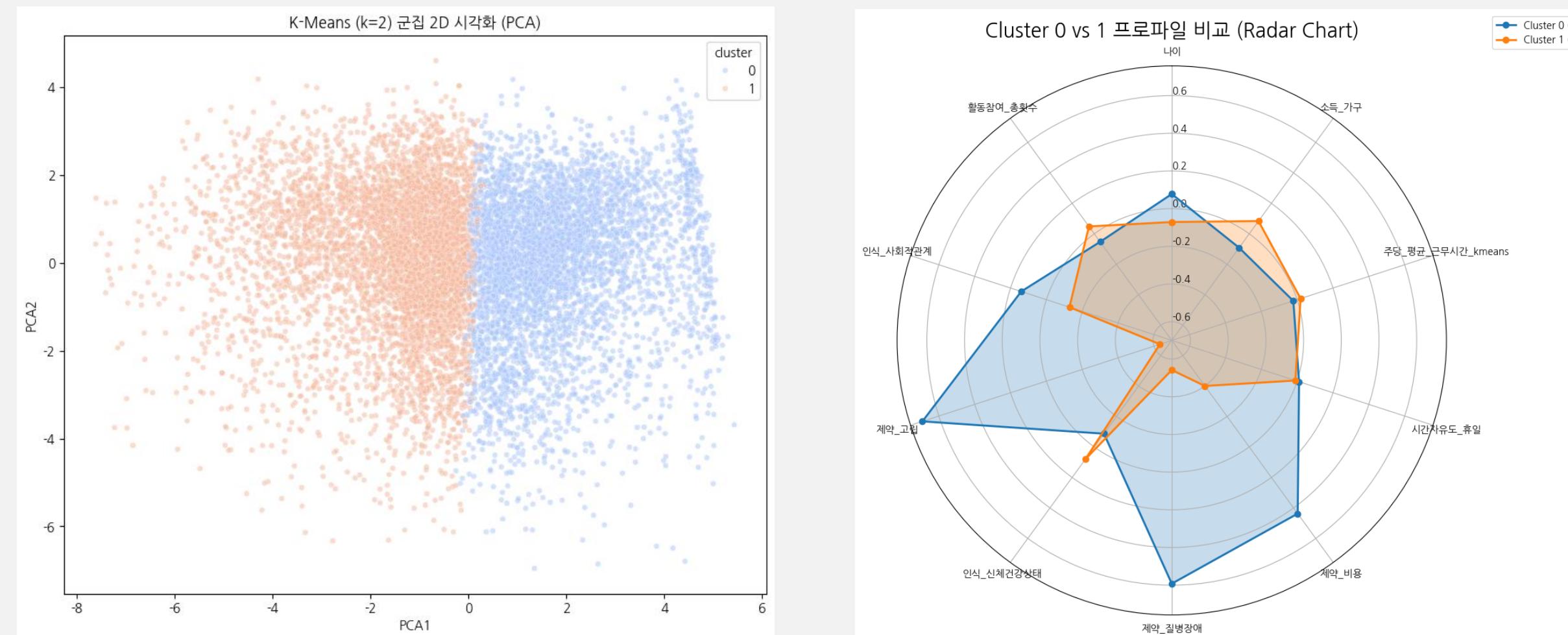


Key Message

- 최적의 k 선정 과정: Elbow Method에서는 잘 안보이는 이슈 -> 실루엣 점수 최고점(0.138)인 K=2를 채택
- 선정 이유: 점수가 낮은 이유로 완벽한 분리는 어렵지만, "정책적으로 유의미한 집단"을 식별하기는 것이 목적

3-3. K-Means (2): 'Cluster 0'과 'Cluster 1' 집단 발굴

K=2 클러스터링 완료 : cluster 1 (5260개), cluster 0 (4815) 라는 결과를 도출하였습니다.



PCA: K-means가 나눈 Cluster 그룹이 실제로 데이터 공간에서 잘 분리되어 있는지 시각적으로 확인할 수 있음

Radar Chart: 두 집단은 '소득', '나이', '질병 제약' 등 모든 핵심 제약 변수에서 명확하고 극적인 차이를 보이고 있음.

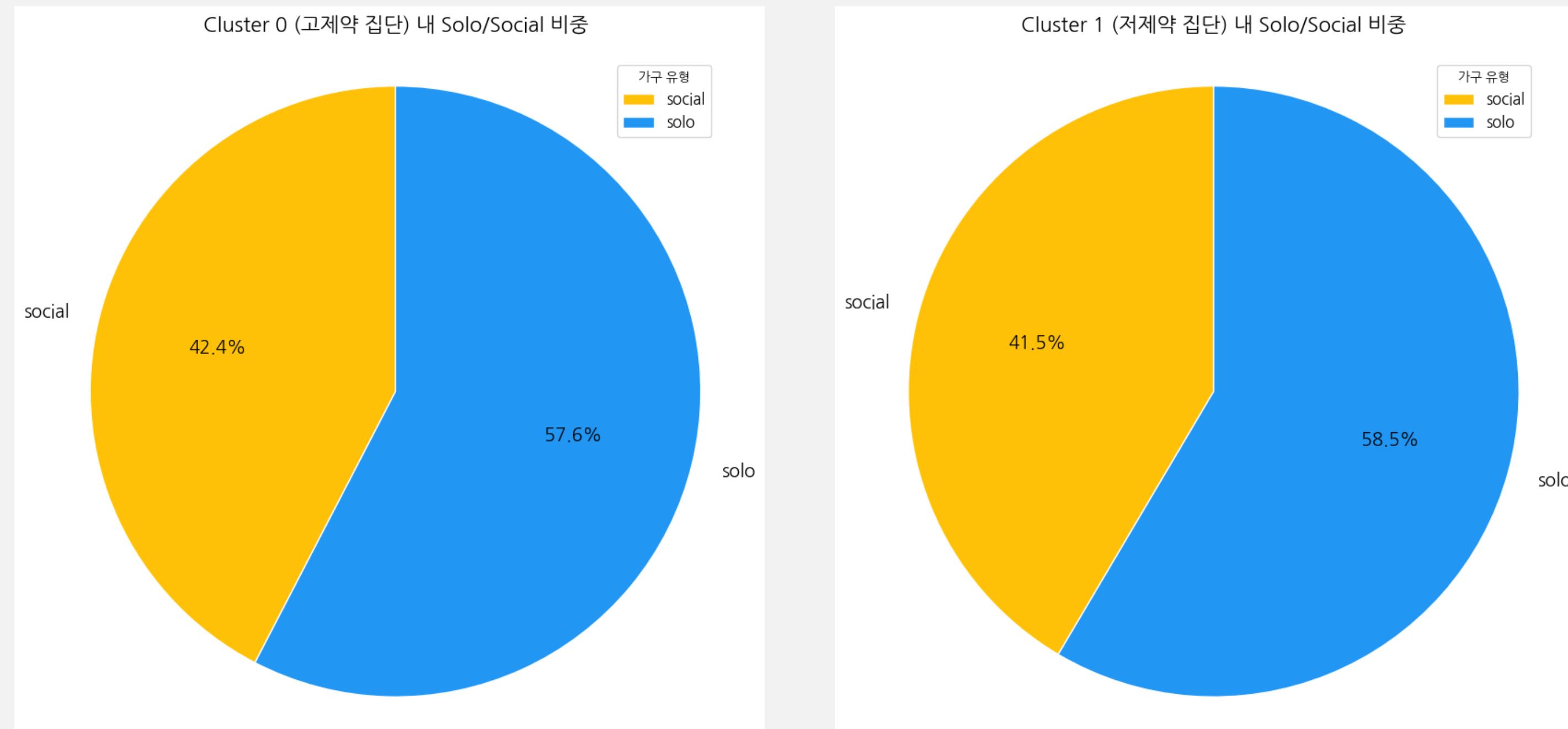
⇒ Cluster 0 (고제약) vs Cluster 1 (저제약)

⇒ 이는 K-means 분류가 정책적 타겟팅의 목적을 성공적으로 달성했음을 보여줌

Key Message

'프로파일(Radar)'과 '2차원 공간(PCA)'을 통해 K-Means가 Cluster 0과 Cluster 1을 성공적으로 분리했음을 알 수 있습니다.

3-4. K-Means (3): 'Cluster 0'는 'solo' 가 아니다.



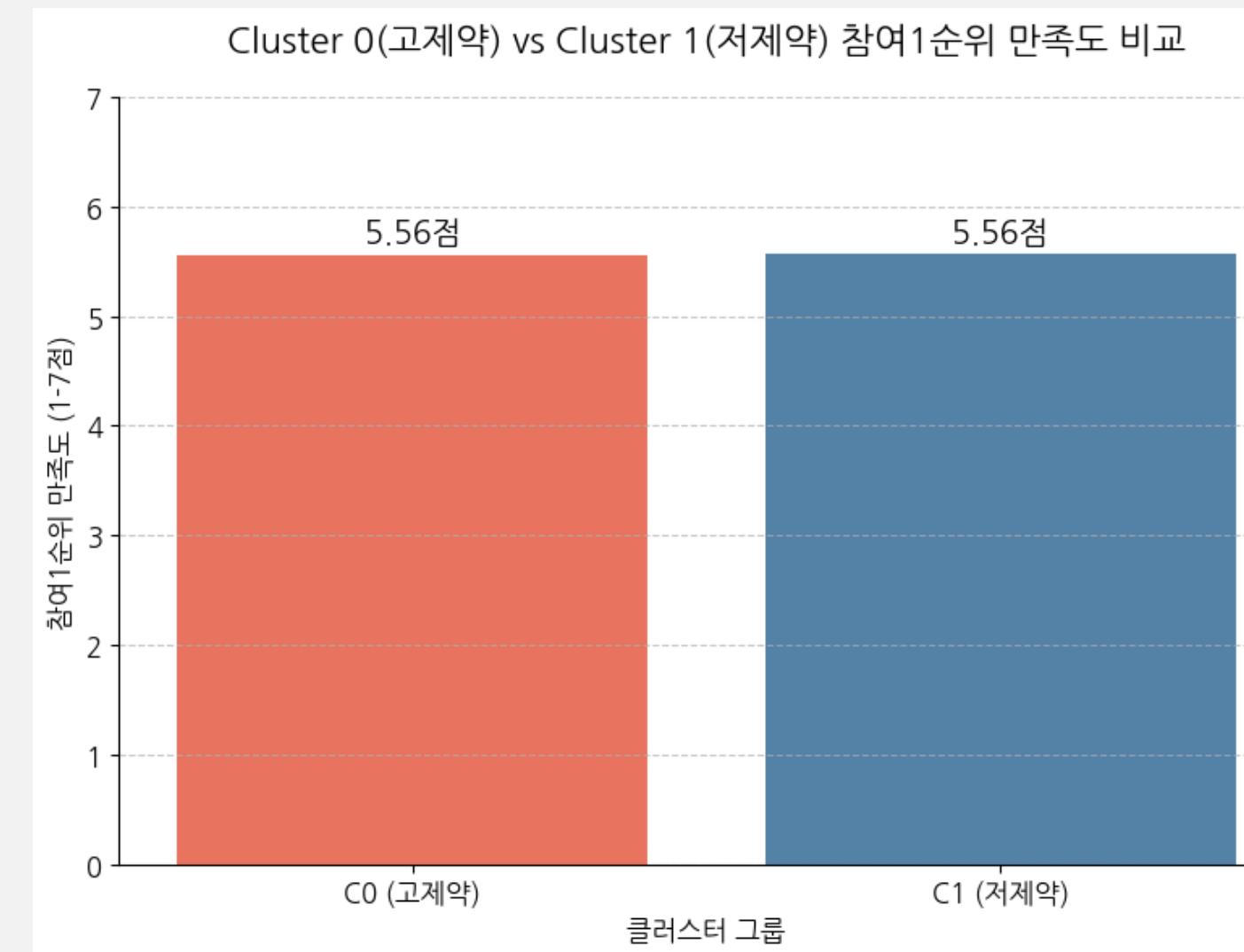
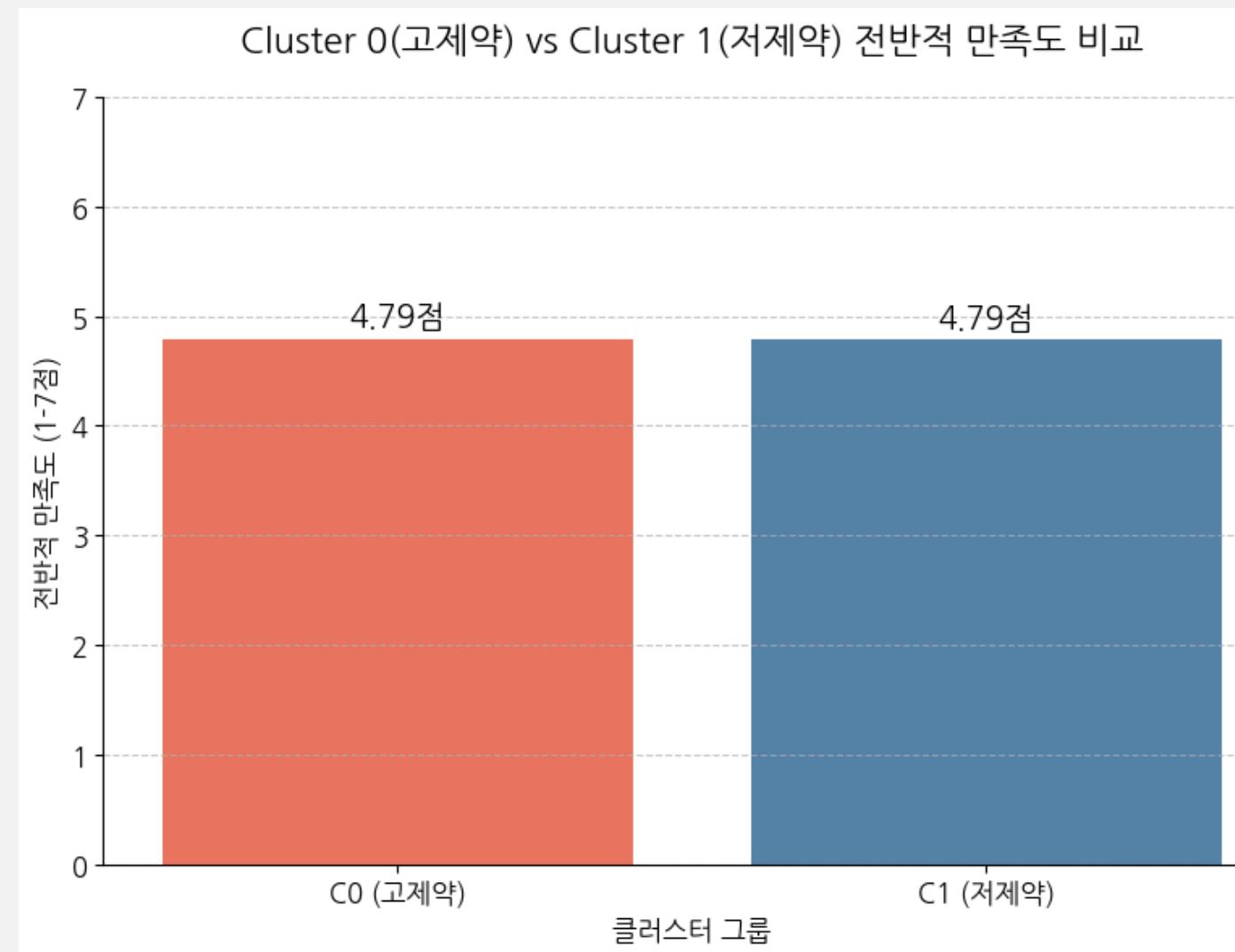
Cluster 0과 1 모두 'solo / social' 비율이 비슷함

=> K-means가 solo와는 '무관한' 새로운 '제약' (들)을 발견했다는 주장이 성립됨.

Key Message

Cluster 0는 'solo'만의 문제가 아닌, solo와 social이 혼재된 '제약 집단임을 다시 한 번 확인했습니다.

3-5. 분석의 전환점: '만족도' 지표의 함정



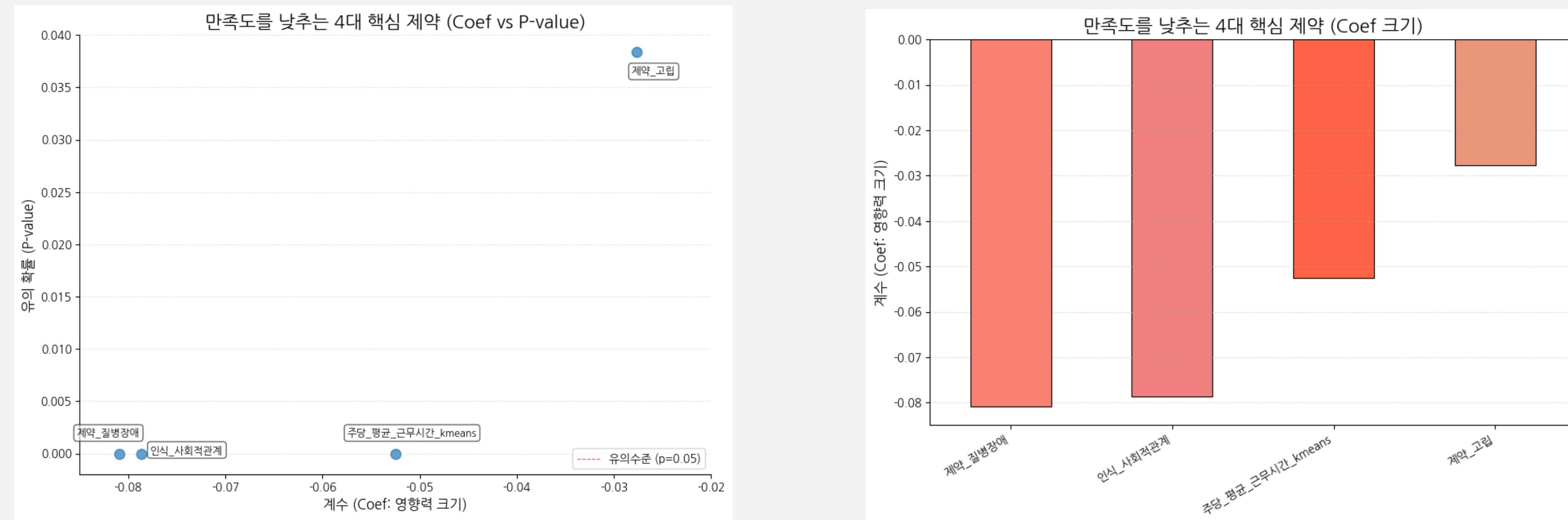
- 만족도가 고제약 집단 문제를 푸는 열쇠(Y)가 아님을 발견
- '만족도'가 아닌 '행동'을 막는 장벽에 집중해야 함
⇒ OLS 분석 실행 결정

Key Message

Cluster 0 (고제약) 집단과 Cluster 1 (저제약) 집단의 만족도 차이는 거의 없었습니다.
이는 '만족도' 자체가 잘못된 목표(Y)임을 시사합니다. 문제는 '행동의 장벽'에 있다고 추정합니다.

3-6. OLS: '행동'을 막는 진짜 장벽 순위

'행동'과 가장 밀접한 지표($Y = \text{만족도}_\text{참여1순위}$)를 기반으로 OLS를 실행하여 "여가 행동을 막는 진짜 장벽"의 순위를 매겼습니다. OLS 분석 결과 중 coef가 음수인 것을 만족도를 낮추는 요인으로 간주하여 "여가 만족도"를 낮추는 4대 핵심 제약을 선정하였습니다.



Key Message

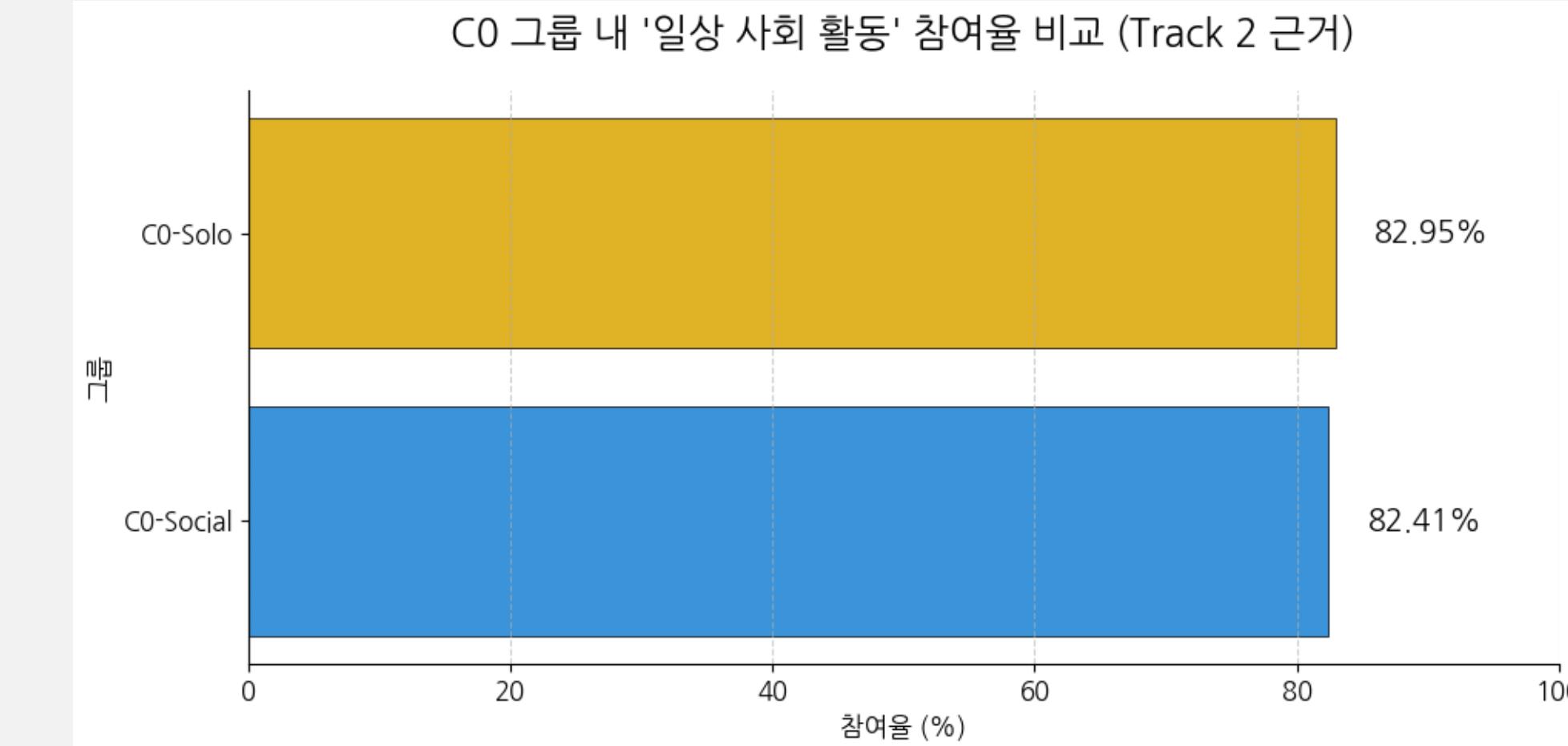
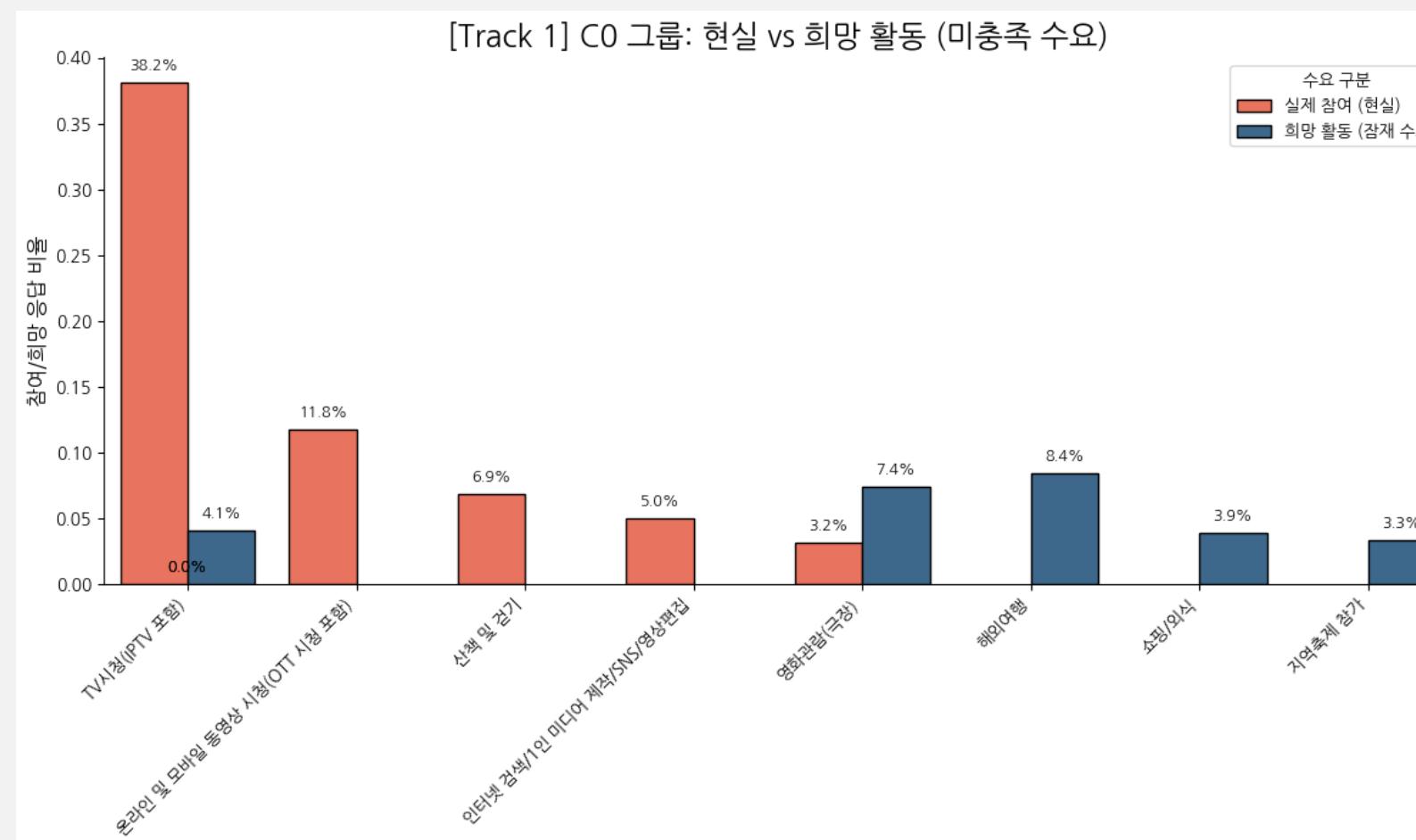
진짜 장벽 1순위는 "제약_질병장애"였으며,
이후 "인식_사회적관계(고립, 외로움)", "주당 평균 근무시간"이 그 뒤를 이었습니다.

03 분석 결과

2025년 문화체육관광 통계데이터 활용대회

3-7. 행동 분석 (정책 근거)

2-Track으로 나누어서 행동 중심으로 여가 불평등을 해소하는 정책을 세우고자 합니다.



[Track 1 근거] 미충족 수요: 현실(TV 시청) vs 희망(여행/관광)
⇒ 정책 방향: '갇힌 행동'(OLS 1순위 장벽)의 해방

[Track 2 근거] C0 – solo와 C0-social의 일상 사회 활동에는 차이가 없음
=> 정책 방향: '정서적 장벽' (OLS 2순위 장벽) 해소

Key Message

- OLS 1순위 장벽(제약_질병장애)은 C0가 '여행'을 포기하고 'TV 시청'에 갇혀있는 '물리적 행동' 문제를 설명합니다.
- OLS 2/3순위 장벽(제약_고립/인식_사회적관계)은 C0-Solo가 행동(참여율 83%)은 하지만, '정서적'으로 고립된 문제임을 설명합니다.

03 분석 결과

2025년 문화체육관광 통계데이터 활용대회

3-8. 교차 검증

'국민문화예술활동조사' 데이터(Culture)로 K-Means(K=2) 및 장벽 분석을 교차 실행하였습니다.

Culture C0 그룹의 Top 5 관람 걸림돌 확인

검증 항목	국민여가활동조사 (Main)	국민문화예술활동조사 (Validation)	교차검증 결과
C0 1순위 장벽	질병/장애, 고립	공급/정보, 시설/접근성, 시간, 비용, 고립	"고립" 일치
C0 – solo 행동	참여율 높음 (83%)	참여율 높음	일치

관람 걸림돌	응답 비율 (%)
공급/정보	32
시설/접근성	25
시간	19
비용	15
고립	8

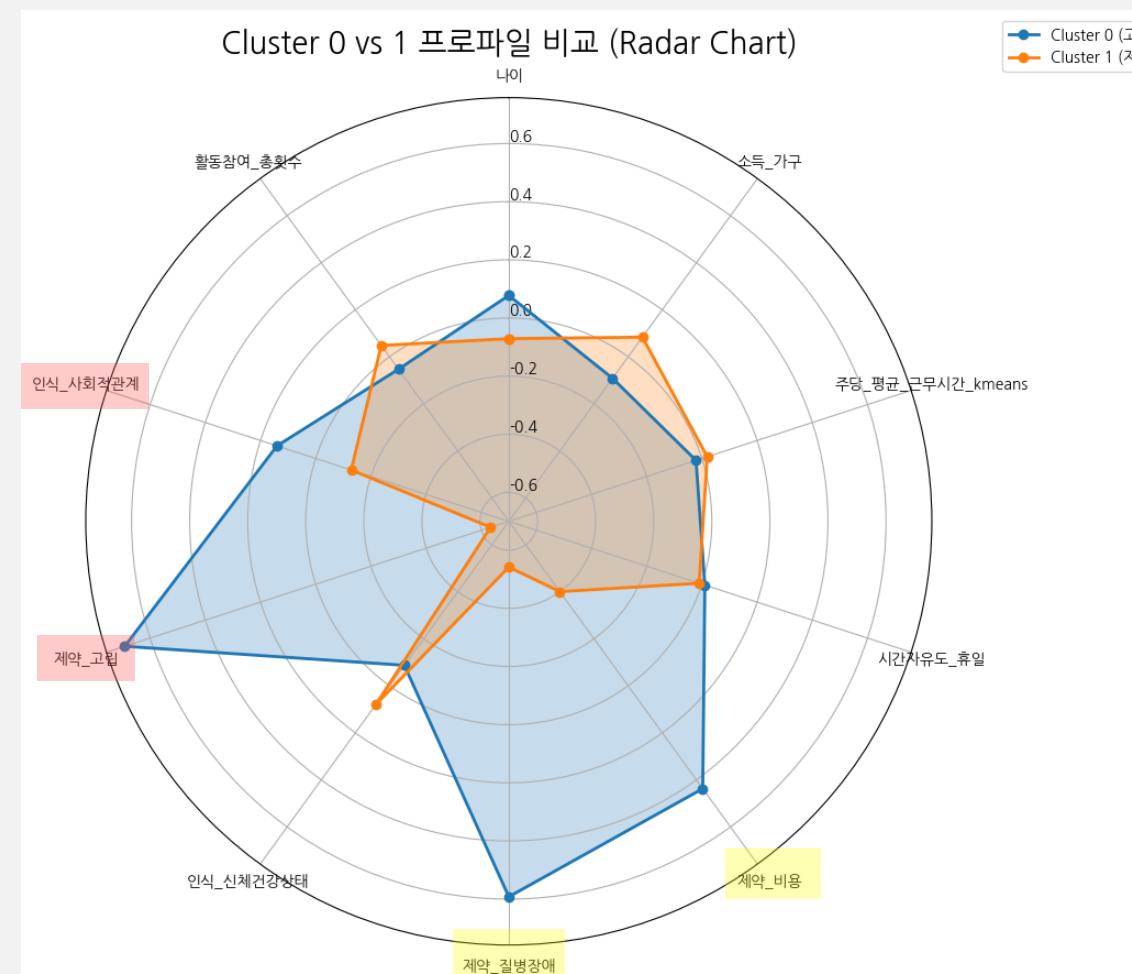
Key Message

다른 데이터셋에서도 "핵심 장벽(비용/고립)"과 "solo의 행동 특성"이 유사하게 재현되었습니다.

4-1. The Opportunity: '이중 제약 집단'의 등장

K-Means Radar Chart:

'물리적 제약'과 '사회적 제약' 확인 가능



	물리적 제약	사회적 제약
현실 : 국가 통계가 말하는 문제	국민여가활동조사(MCST, 2024) 결과, 1순위 여가 장벽은 ' 경제적 부담 (51.6%) '입니다.	사회조사(통계청, 2023) 결과, "도움받을 곳 없음" 응답이 41.9% (1인 가구)에 달해 ' 사회적 고립 '이 심각합니다.
분석: 본 연구의 발견	이는 저희 OLS 1순위 장벽인 ' 제약_질병장애 ' 및 C0의 '저소득/고비용' 특성과 일치하는 ' 물리적 제약 '입니다	이는 저희 OLS 2/3순위 장벽인 ' 제약_고립 ' / ' 인식_사회적관계 '와 정확히 일치하는 ' 정서적/사회적 제약 '입니다.

Key Message

C0는 '물리적 제약(비용/건강)'과 '사회적 제약(고립)'을 동시에 겪는 '이중 제약 집단'이며, 기존 문화 복지 정책들은 이들에게 도달하지 못하는 경우가 많습니다.

4-2. The Solution: AI '제약 인식' 내비게이터

AS-IS: '선호도' 기반 여가 활동 추천

"무엇을 원하세요?" (Preference)

"선호 테마" (예: 럭셔리, 힐링)

"희망 지역" (예: 제주, 강릉)

"높은 평점/인기"

유저 선호도에 맞춰서 '프리미엄 호텔' 제안
-> 현실적인 문제에 부담을 가지고 여가 생활 포기하는 경우가 많음

TO-BE: '제약' 기반 여가 활동 추천

"무엇을 원하시고, 무엇이 여가 생활을 막고 있나요?" (Constraint)

"신체/건강 제약이 있으신가요?" (OLS 1순위)

"비용 제약이 있으신가요?" (MCST 1순위)

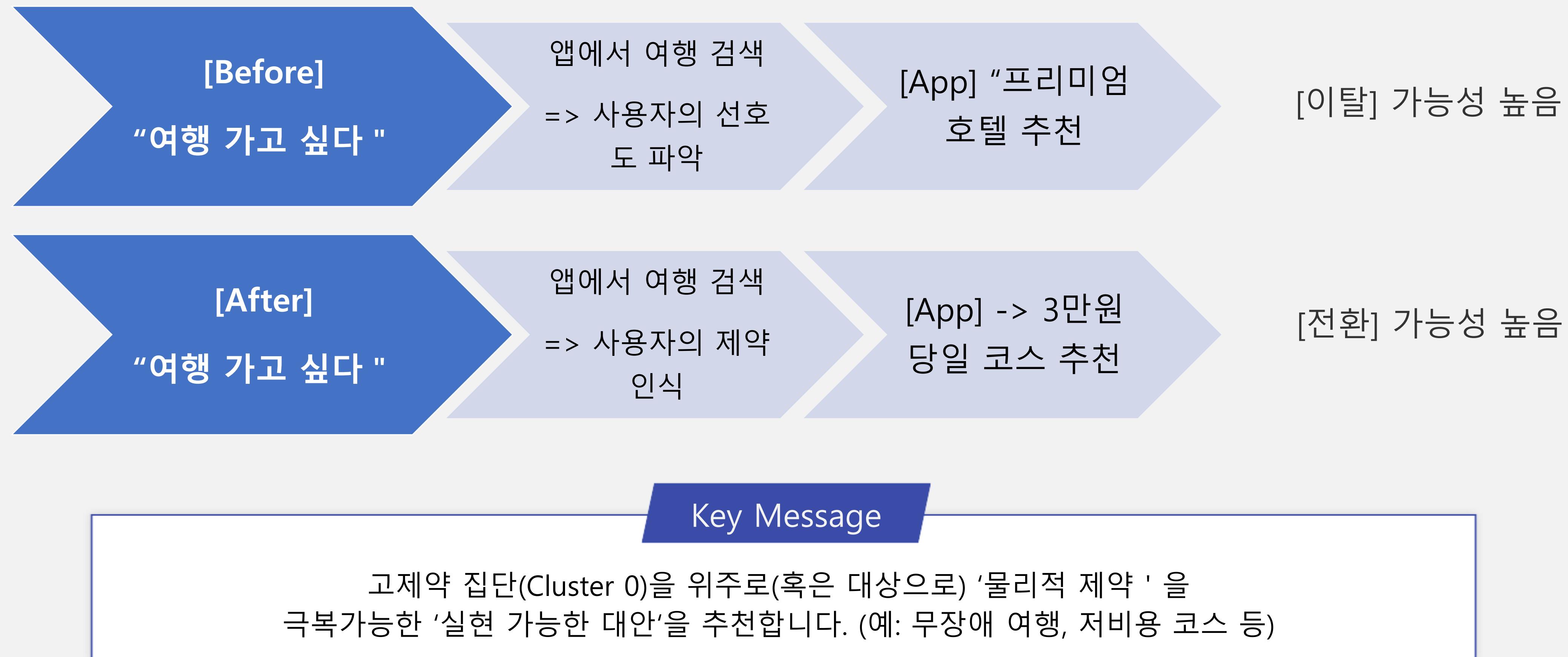
"사회/정서적 제약이 있으신가요?" (OLS 2순위)



유저의 제약을 선제적으로 파악한 후, 상황에 맞춰서
"이번 주말 3만원 당일치기 코스"를 추천

4-3. How it works?: Track 1 (물리적 제약 극복)

Cluster 0 전체 대상 : AI가 사용자의 제약을 파악 (OLS 1순위 (질병/장애))와 미충족 수요(여행 등)을 인식



4-4. How it works?: Track 2 (사회적 제약 극복)

Cluster 0 – solo 맞춤 : AI가 사용자의 제약을 파악 (OLS 2순위 (고립/외로움 등)) 등 인식



고제약 집단(Cluster 0) 중 solo 집단을 맞춤화하여 '만남 주선' 방식 대신
'관심사 기반 1인 참여' 커뮤니티를 선제적으로 매칭하도록 합니다.

4-5. 기대 효과 및 핵심 KPI 설정

구분	목표 KPI (행동 지표)
C0 전체 (물리적 제약 부문 중심 해결)	추천 수용률(Click-Through) 10% 상승, 여행 전환율 5% 상승
C0 – solo (사회적 제약 부문 중심 해결)	커뮤니티 연결율 30% 상승, 동호회 참여율 20% 상승
정책/플랫폼	'고제약 집단' 유저 확보 및 충성도 증가

Key Message

본 솔루션의 기대 효과는 측정 가능한 '행동' KPI로 증명됩니다.
이 KPI는 제안하는 정책의 구체적인 성공 척도입니다

5-1. 결론 및 핵심 발견 내용 요약

1차 가설 기각

- '만족도_전반적' 통계적 유의성 만족 안함 ($p > 0.05$)
- '만족도_참여1순위' solo & social 차이 없음

=> Solo (혼자 여가 생활)는 핵심 원인이 아니었다.

핵심 발견

여가 불평등 '6대 고전 요인'을 통제한 OLS 분석 결과
=> '신체적/정서적 장벽' 이 행동을 막는 1순위 원인임을 규명

솔루션

고제약 집단 Cluster 0 (이중 제약 집단)을 위해 AI '제약 인식' 내비게이터라는 구체적이고 실현 가능한 솔루션 제안

최종 결론

'이중 제약 집단(CO)'을 위해 기존의 '선호도' 기반이 아닌 AI '제약 인식' 내비게이터라는 구체적이고 실현 가능한 솔루션을 제안했습니다.
이는 '만족도'라는 모호한 목표가 아닌, '참여율', '전환율' 등 측정 가능한 '행동 KPI'를 통해 여가 불평등을 실질적으로 해소하는 데 기여할 거라 기대합니다.

OLS 상세 결과

OLS Regression Results									
Dep. Variable:	만족도_참여1순위	R-squared:	0.119						
Model:	OLS	Adj. R-squared:	0.116						
Method:	Least Squares	F-statistic:	41.01						
Date:	Sun, 16 Nov 2025	Prob (F-statistic):	4.56e-246						
Time:	20:11:35	Log-Likelihood:	-12322.						
No. Observations:	10075	AIC:	2.471e+04						
Df Residuals:	10041	BIC:	2.496e+04						
Df Model:	33								
Covariance Type:	nonrobust								
	coef	std err	t	P> t	[0.025	0.975]			
const	5.5025	0.026	215.556	0.000	5.452	5.553			
solo_여부	0.0218	0.009	2.532	0.011	0.005	0.039			
제약_시간	0.0470	0.011	4.387	0.000	0.026	0.068			
제약_비용	0.0067	0.010	0.666	0.506	-0.013	0.027			
제약_질병장애	-0.0809	0.011	-7.166	0.000	-0.103	-0.059			
제약_고립	-0.0276	0.013	-2.071	0.038	-0.054	-0.001			
제약_시설접근	0.0297	0.015	2.023	0.043	0.001	0.058			
제약_정보부족	0.0008	0.014	0.060	0.952	-0.027	0.028			
제약_경험부족	0.0343	0.014	2.494	0.013	0.007	0.061			
제약_프로그램	-0.0087	0.015	-0.585	0.559	-0.038	0.020			
제약_기후여건	0.0182	0.011	1.678	0.093	-0.003	0.039			
나이	0.0045	0.015	0.292	0.770	-0.026	0.034			
소득_가구	0.0028	0.010	0.269	0.788	-0.017	0.023			
주단_평균_근무시간_kmeans	-0.0525	0.009	-5.539	0.000	-0.071	-0.034			
시간자유도_평일	0.0987	0.012	8.509	0.000	0.076	0.121			
시간자유도_휴일	0.1046	0.012	9.019	0.000	0.082	0.127			
이용_충분도	0.0168	0.008	1.982	0.047	0.000	0.033			
인식_사회적관계	-0.0786	0.009	-8.360	0.000	-0.097	-0.060			
인식_정신건강	0.0420	0.010	4.102	0.000	0.022	0.062			
인식_신체건강상태	0.1019	0.012	8.769	0.000	0.079	0.125			
활동참여_총횟수	0.0360	0.009	4.004	0.000	0.018	0.054			
장애여부	0.2512	0.067	3.763	0.000	0.120	0.382			
유연근무제_시행_여부	-0.0011	0.045	-0.024	0.981	-0.090	0.088			
성별_여성	0.0361	0.017	2.064	0.039	0.002	0.070			
총인상태_배우자_있음	0.0455	0.028	1.617	0.106	-0.010	0.101			
총인상태_사별	0.0401	0.046	0.865	0.387	-0.051	0.131			
총인상태_이혼	0.0597	0.048	1.254	0.210	-0.034	0.153			
최종학력_대학(4년제_미만)	0.0189	0.027	0.706	0.480	-0.034	0.072			
최종학력_대학(4년제_이상)	0.0184	0.022	0.849	0.396	-0.024	0.061			
최종학력_대학원_박사_과정	0.1533	0.195	0.785	0.432	-0.229	0.536			
최종학력_대학원_석사_과정	0.1047	0.076	1.369	0.171	-0.045	0.254			
최종학력_무학	0.0170	0.092	0.186	0.853	-0.163	0.197			
최종학력_중학교	-0.0329	0.034	-0.970	0.332	-0.099	0.034			
최종학력_초등학교	-0.0435	0.040	-1.099	0.272	-0.121	0.034			
Omnibus:	19.682	Durbin-Watson:		1.564					
Prob(Omnibus):	0.000	Jarque-Bera (JB):		19.736					
Skew:	-0.108	Prob(JB):		5.18e-05					
Kurtosis:	3.022	Cond. No.		52.7					

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

1. 모델 전체 요약 (표의 윗부분)

- Dep. Variable (종속 변수): 만족도_참여1순위 => '예측하고 싶은 목표(Y)'가 무엇인지 보여줌
- .R-squared (R제곱, 결정계수): 0.119 => "이 모델이 '만족도'가 왜 변하는지를 11.9% 설명한다"는 뜻 (사회과학 데이터에서는 이 수치가 낮아도 괜찮음)
- Prob (F-statistic) (F-통계량의 p-value): 4.56e-246 => 4.56 곱하기 \$10^{-246}\$이라는 뜻으로, 거의 0에 가까움
(해석) 이 값이 0.05보다 작으므로, "이 모델은 통계적으로 매우 유의미하며, 적어도 변수 중 하나는 만족도에 영향을 미친다"는 것을 증명

2. 핵심 결과: 계수표 (표의 중간 부분) - '핵심 발견'**이 담긴 본체

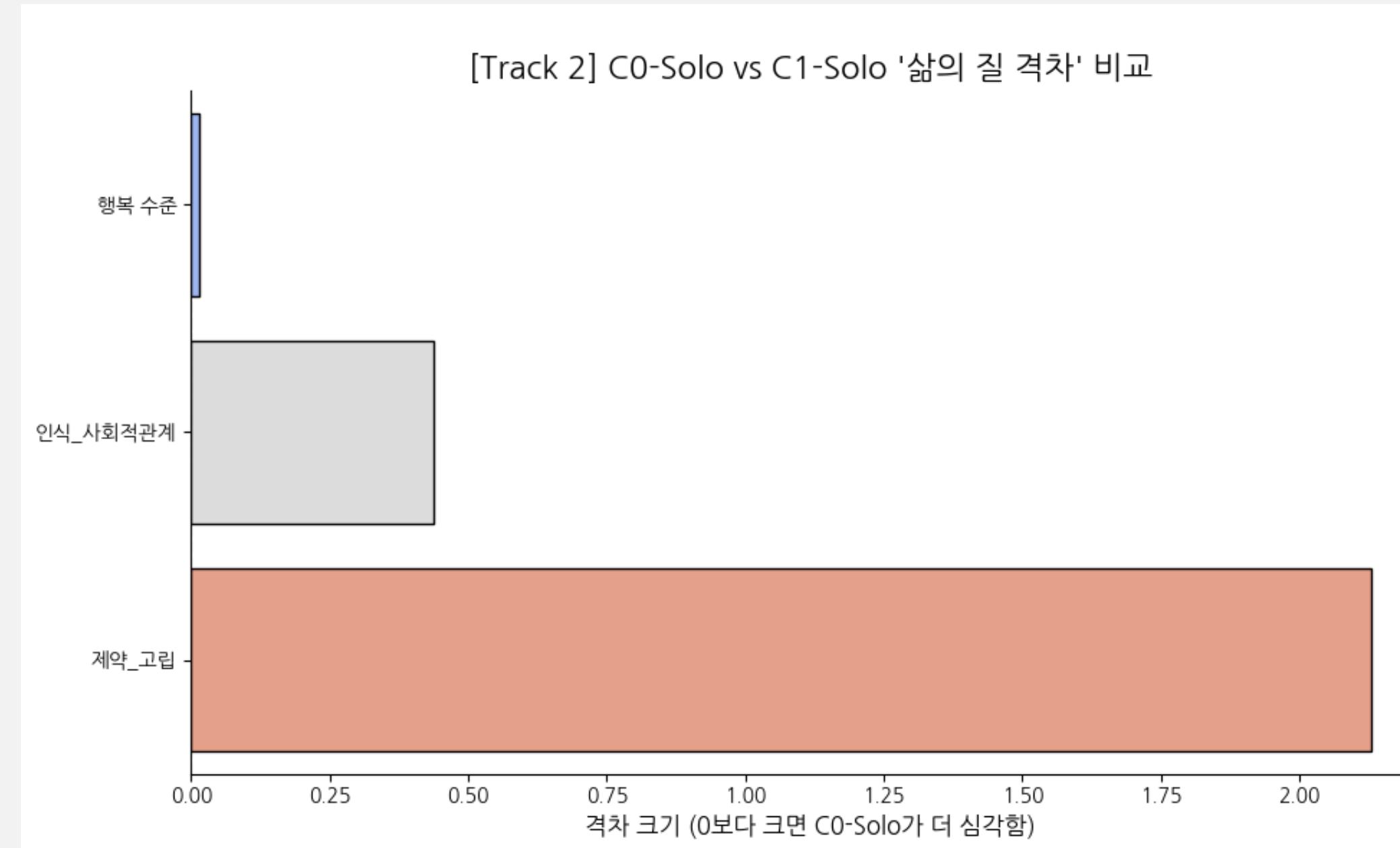
[제약_질병장애 변수를 예시로 설명]. 제약_질병장애 | coef -0.0809 | std err 0.011 | t -7.166 | P>|t| 0.000

- coef (계수): -0.0809:
 - [의미] 가장 중요한 값. "다른 모든 조건이 같다면, '질병/장애' 제약이 1점 높아질 때마다, '만족도'는 -0.0809점 낮아진다"는 뜻
 - [해석] 마이너스(-) 값이므로 **부정적인 영향(장벽)**을 의미하며, 숫자의 절댓값이 클수록 강력한 원인
- P>|t| (p-value): 0.000
 - [의미] 이 coef 값이 우연히 나왔을 확률
 - [해석] 이 값이 0.05보다 작으므로(0.000 < 0.05), **이 변수(-0.0809)는 우연이 아니라 '진짜' 영향을 미친다**는 것을 증명
- t (t-값): -7.166
 - [의미] coef 값을 std err로 나눈 값으로, p-value를 계산하기 위한 통계치. (보통 절댓값이 2보다 크면 유의미함)
- [0.025 / 0.975] (95% 신뢰구간): [-0.103 / -0.059]
 - [의미] "95% 확률로, 이 변수의 실제 영향력(coef)은 -0.103과 -0.059 사이에 있을 것이다"라는 뜻
 - [해석] 이 구간에 0이 포함되지 않으므로(중요), 이 변수의 영향력은 0이 아니라는 것, 즉 유의미하다는 것을 다시 한번 증명

3. 모델 진단 (표의 아랫부분)

- Durbin-Watson: 1.564(의미) 잔차(오차)의 독립성을 검증. (보통 1.5 ~ 2.5 사이면 양호함)
- Prob(Omnibus) / Prob(JB): 0.000 / 5.18e-05: 잔차(오차)가 정규분포를 따르는지 검증
(0.05보다 크면 정규분포를 따름)
- (해석) 이 모델은 잔차가 정규분포를 따르지 않지만, No. Observations(관측치)가 10,075개로 매우 크므로 중심극한정리에 의해 OLS 결과를 신뢰할 수 있음

solo-gap (클러스터별 삶의 질 격차 비교) 분석



1. 이 차트가 보여주는 것

- 똑같이 '혼자(Solo)' 여가를 즐기는 사람이라도, C0(고제약)에 속한 Solo와 C1(저제약)에 속한 Solo는 삶의 질이 얼마나 다른가?"를 보여줌
- 막대의 길이는 두 그룹 간의 '격차(Gap)'입니다. (C0-Solo 점수 빼기 C1-Solo 점수)
- X축(0보다 큰 값)은 C0-Solo가 C1-Solo보다 얼마나 더 심각한지를 의미

2. 핵심 해석

- 제약_고립 (약 +2.1): '고립으로 인한 제약'은 C0-Solo가 C1-Solo보다 압도적으로 심각
- 인식_사회적관계 (약 +0.45): C0-Solo가 자신의 '사회적 관계'를 훨씬 더 부정적으로 인식
- 행복 수준 (매우 작음): '행복 수준' 역시 C0-Solo가 더 낮음

3. 이 차트가 왜 중요한가?

- "여가 불평등의 진짜 문제는 'Solo' 그 자체가 아니라, 'Solo'이면서 동시에 C0(고제약)인 상태"임을 증명
- C0-Solo는 '물리적 제약'(C0)과 '사회적 제약'(Solo)의 '이중제약'을 겪는 핵심 정책 대상입니다.

=> Track 2 정책이 왜 다른 Solo가 아닌 C0-Solo 그룹을 특별히 타겟팅해야 하는지에 대한 완벽한 근거 제공

감사합니다.

혼자 품DA / 이연지