

빅데이터 활용 미래 사회문제 해결 아이디어 해커톤

고립·은둔 청년 실태 분석과 예측 및 해결방안 제안

최인영 이승아 정은수 류승권



목차

1. 서론

- (1) 연구 배경
- (2) 주제 선정 이유

3. 데이터 분석

- (1) 데이터 분포
- (2) 분석 방법
- (3) 분석 결과 및 해석

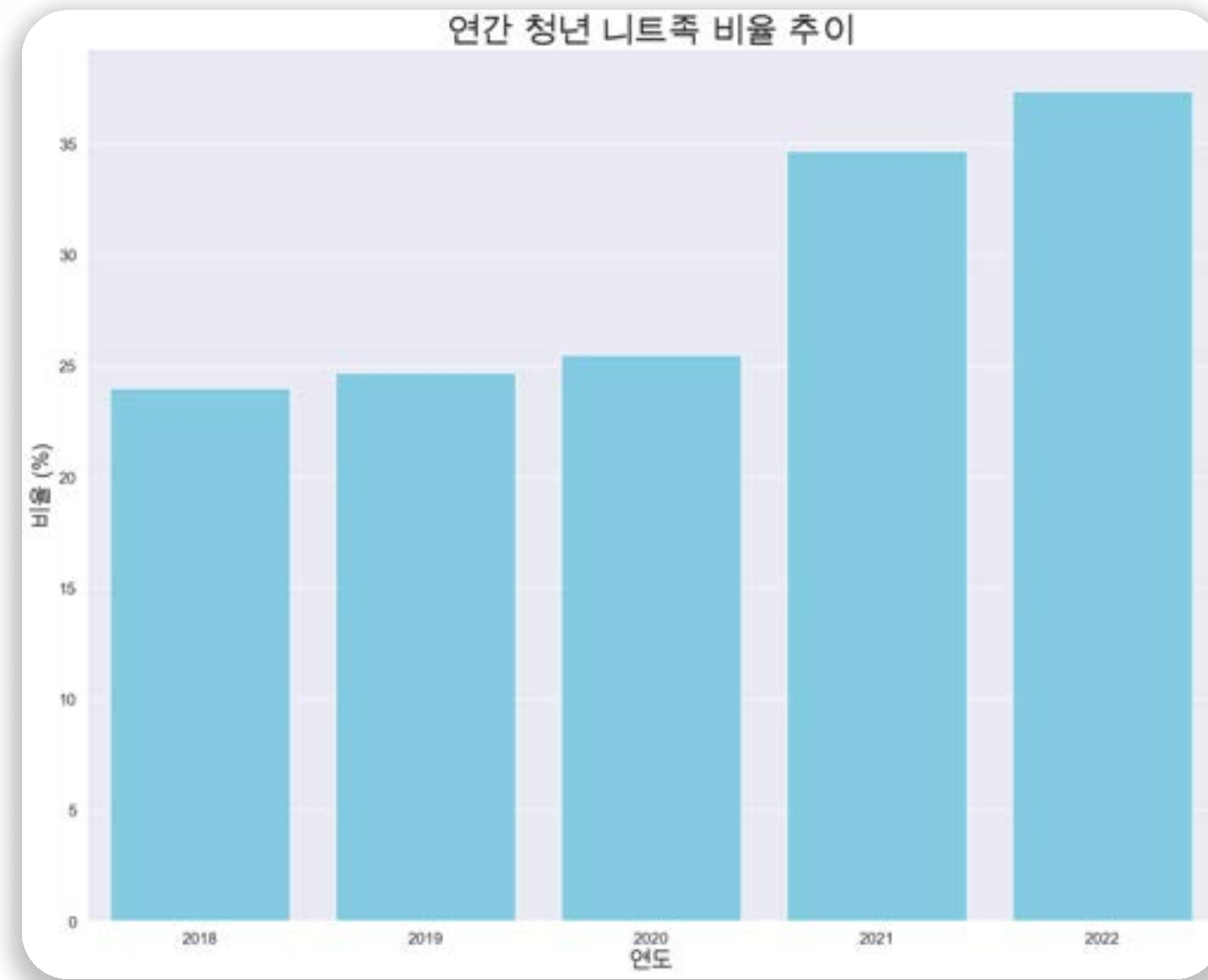
2. 데이터 전처리

- (1) 활용 데이터 소개
- (2) 결측치 현황 및 처리
- (3) 데이터 검정

4. 활용 방안

- (1) 해결방안 제시
- (2) 기대효과

01 | 연구 배경

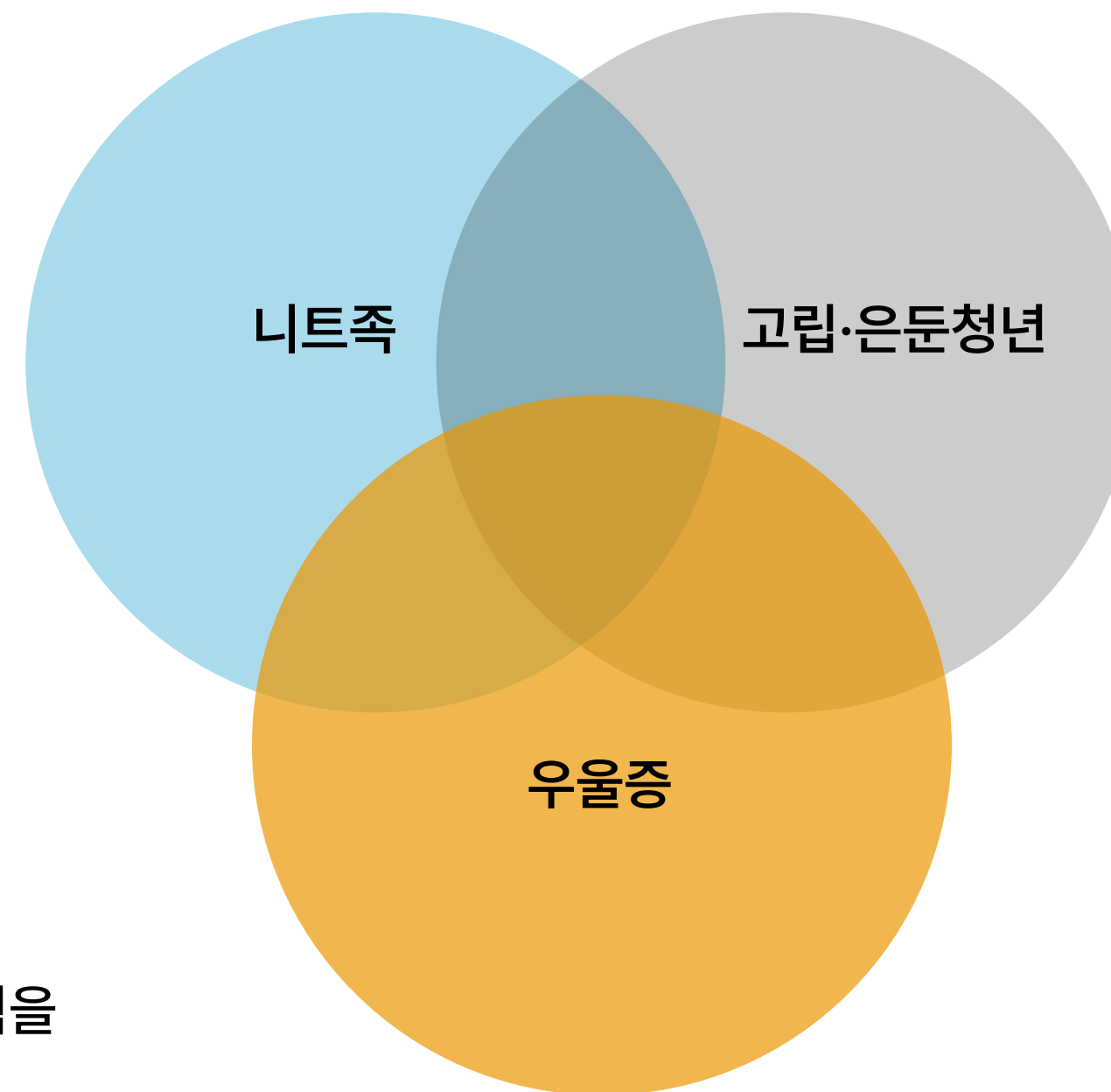


니트족 : 나라에서 정한 의무교육을 마친 뒤에도 진학이나 취직을 하지 않으면서 직업훈련도 받지 않는 사람

그냥 쉬는 '니트' 청년 증가세...지원책은 충분할까?

니트는 은둔생활의 전 단계로 적절한 조치가 취해지지 않는 경우 고립, 은둔 단계로 접어들 위험이 크다.

데일리팝

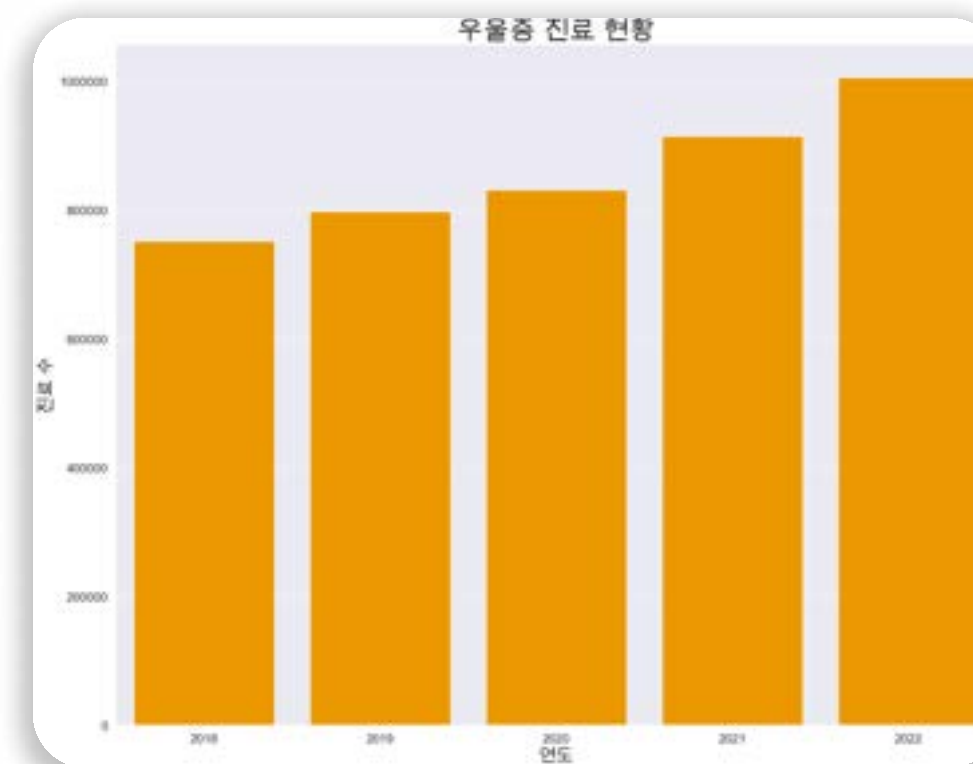


**[취재N팩트] '취업실패·대인관계' 때문에...
고립·은둔청년 54만 명**

한국보건사회연구소 "19~39세 청년 가운데 고립·은둔 청년 비율이 5%에 달한다"

통계청 "중대한 질병이나 장애 없이 구직활동을 하지 않는 청년이 7년 전 24만9천 명에서 올해 7월 40만2천 명으로 크게 늘어.."

YTN 뉴스



"절반이 중증 우울증" 어둠속에 방치된 61만명

응답자의 18.5%가 정신건강 관련 약물을 복용한다고 답해 일 반청년 8.6%보다 2배 이상 높았다. 또한 10명 중 8명은 '가벼운 수준 이상의 우울'(중증수준 이상은 57.6%)을 겪고 있다고 답했다.

파이낸셜 뉴스

서론

02 | 주제 선정 이유

외로움이 불러올 미래는...코로나 기간 우울증 환자 10년 새 최대폭 '증가'

권준수 서울대 정신건강의학과 교수는 “외로움은 결국 우울증을 비롯해 치매, 동맥경화 등 모든 질병의 요인이 된다”면서 “극단적인 경우 자살로 이어진다”고 경고했다.

서울신문

“외로움은 사회의 질병”...뉴욕, 외로움 명예대사 임명

사회적 고립은 인지 기능 저하, 불안 및 면역력 악화와 관련 있는 것으로 나타났다. 웨스트하이머의 대사 임명을 두고 “외로움과 사회적 고립은 개인의 문제가 아니라 사회적 질병이다”라고 설명했다.

코메디닷컴

[뉴스속 용어]장애 인정 추진되는 '사회적 장애'

보건복지부가 9일 발표한 '제6차 장애인정책종합계획(2023~2027)'에 의학적 장애로 규정된 '장애'의 개념을 사회적 장애로 확대한다는 내용이 포함됐다. 현재 국회에 관련 법이 계류중이며 법안이 통과되면, '장애 개념이 사회적 장애까지 확장돼 히키코모리(은둔형 외톨이)'와 같은 사회적 장애도 장애로 인정돼 필요한 지원을 받을 수 있다.

아시아경제

외로움과 사회적 고립은 점차 질병과 장애로 여겨지고 있다.
따라서 외로움과 고립을 겪는 고립은둔청년의 증가는 속히 해결해야 할 사회적 문제라고 볼 수 있다.

사회적 문제 해결

고립이 일상화되면 저출산을 가속화시키는 등의 사회적 문제가 발생한다.
고립 기간이 10년이 넘은 경우에는 10명 가운데 9명이 자살을 생각한다

경제적 문제 해결

청년재단은 이들에게 드는 사회적 비용이 연간 7조 5000억원이라고 추산했다.
청년 니트를 모두 일하게 할 수 있다면 GDP가 최소 1.5% 성장한다.

사회·경제적 문제 및 사회 전반의 문제를 해결하기 위해 고립은둔청년의 사회진출 방안 마련이 필요하다.

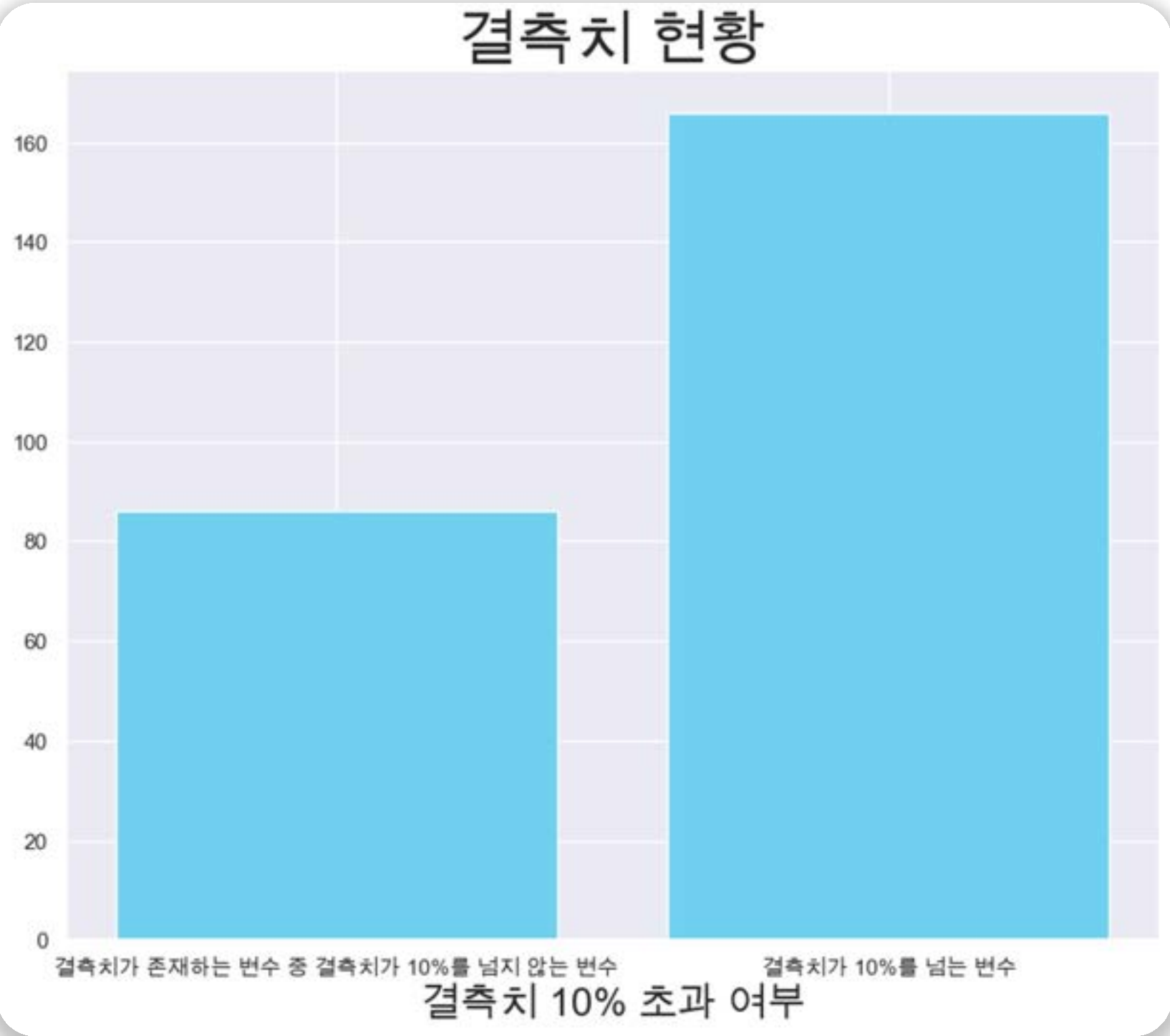
서울특별시_고립은둔 청년 실태조사 결과
주요 변수 소개

고립은둔청년 데이터프레임 행 : 448

고립은둔 청년에 해당하지 않는 데이터프레임의 행 : 4985

설문자의 기본적인 정보부터 고립사유, 외로움과 우울 사유 등이 있다.

변수명	설명	변수명	설명	변수명	설명
SQ1	성별	A1	노동 여부	A13_(1~4)	물리적 고립 사유
SQ2_X	연령대	A2_1_(1~9)	미취업 이유	A18_(1~4)	외로움 척도 사유
SQ4	거주자치구	A4_(1~8)	생활비 출처	B1_(1~5)	평소 생활 패턴
SQ6_R	동거인수	A7	외출 빈도	B12_(1~9)	우울 척도
SQ7_(1~12)	동거 가족	A8	은둔 기간		
SQ8_(1~13)	비동거 가족	A12_(1~4)_(1~4)	정서적 고립 사유		



1. 결측치 10%를 넘어가는 변수가 결측치 10%를 넘지 않는 변수보다 훨씬 많이 분포하고 있음. 결측치 10%가 넘는 변수는 제거

2. 결측치가 존재하는 변수 중 결측치가 10%를 넘지 않는 변수

변수명	결측치 (%)	변수명	결측치 (%)	변수명	결측치 (%)
SQ5	1.451116	B2	2.902231	DQ2	1.451116
A5	1.451116	B2_R	2.902231	DQ3	2.884092
A6	1.451116	B3	2.884092	B4	5.514239
A10	2.884092	B5_2	2.176673		
A11	2.884092	B5_3	4.280791		

3. 최빈값으로 결측치를 대체하여 활용

선정된 변수

'SQ1','SQ2','SQ2_X', 'SQ2_R', 'SQ3', 'SQ4','SQ6_R','A1','A7','A9','A13_1','A13_2','A13_3','A13_4','A15','A17','A18_1',
'A18_2','A18_3','A18_4','A19_1','A19_2','A19_3','A19_4','A19_5','A19_6','A20_1',
'A20_2','A20_3','A20_4','A20_5','A20_6','A20_7','A20_8','B1_1','B1_2','B7','B8','B9_1','B9_2','B9_3','B9_4',
'B10','B11','B12_1','B12_2','B12_3','B12_4','B12_5','B12_6','B12_7','B12_8','B12_9','DQ1','DQ5','D Q6', 'SQ5', 'A5', 'A6', 'A10', 'A11', 'B2', 'B2_R', 'B3', 'B4',
'B5_2', 'B5_3', 'DQ2', 'DQ3'

연속형 변수 SQ2(나이)

1. 정규성 검정 (Shapiro-Wilk Test)

고립은둔청년 집단과 나이, 고립은둔청년이 아닌 집단과 나이 두 집단에 대한 정규성 검정 실시

2. T – 검정 or 맨-휘트니 검정 (Mann-Whitney Test)

정규성 여부에 따라 T – 검정 또는 맨-휘트니 검정을 실시

범주형 변수

1. 명목 변수

단순 범주로 분류된 명목형 변수의 경우에는 카이제곱 독립성 검정 (Chi-Squared Test of Independence)을 실시

2. 순위 변수

순서 관계가 존재하는 순위형 변수의 경우에는 맨- 휘트니 검정 (Mann-Whitney Test) 을 실시

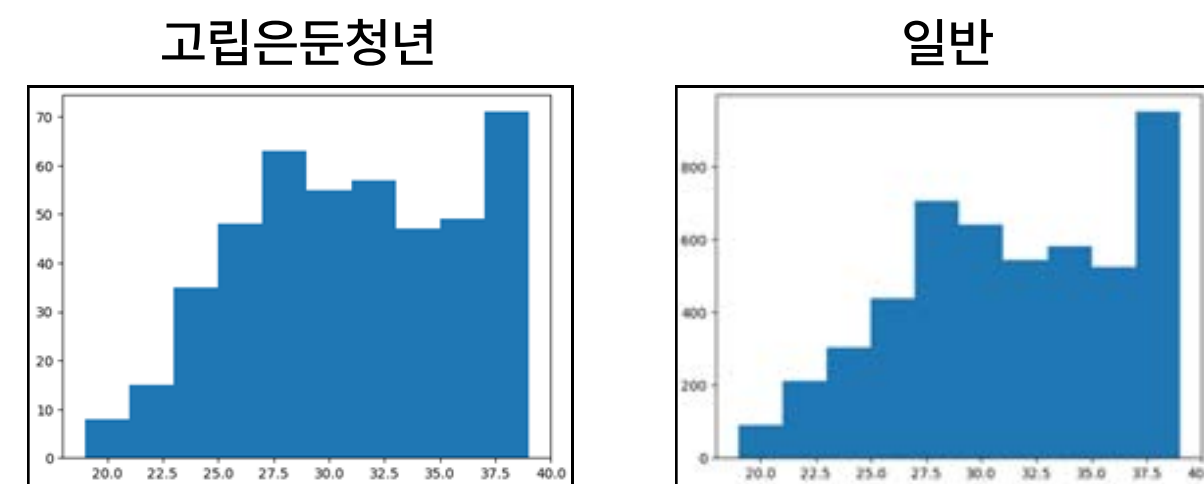
연속형 변수 SQ2(나이)

정규성 검정

1. 고립은둔 청년과 고립은둔청년이 아닌 집단 분할 후 정규성 검정 (Shapiro - Wilk Test)

고립은둔청년 집단 pvalue 약 7.65e-08

고립은둔청년이 아닌 집단 pvalue 약 2.51e-33



→ 정규성을 띄지 않으므로 맨-휘트니 검정 실시

2. 두 집단을 대상으로 맨-휘트니 검정을 실시

Pvalue 약 0.15017 ...

두 표본의 중앙값은 동일하다고 할 수 있다.
따라서 연령과 고립은둔여부는 관계가 없다.

범주형 변수

1. 명목 변수

카이제곱 검정 실시

SQ1(성별), SQ3(지역[시,도])

위의 두 변수가 pvalue가 0.05보다 크므로 고립은둔여부와 관계가 없는 것으로 나타났다.

2. 순위 변수

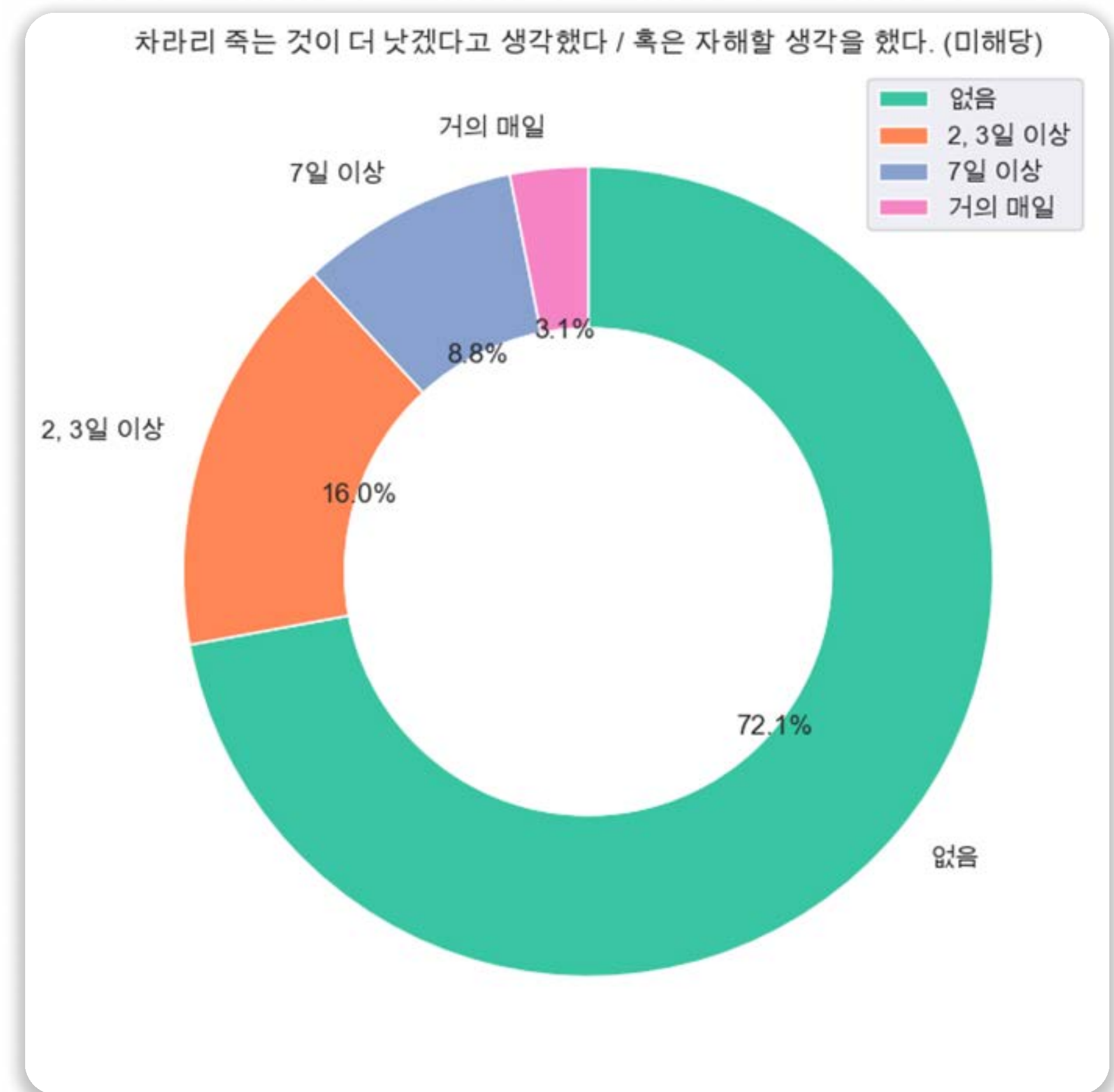
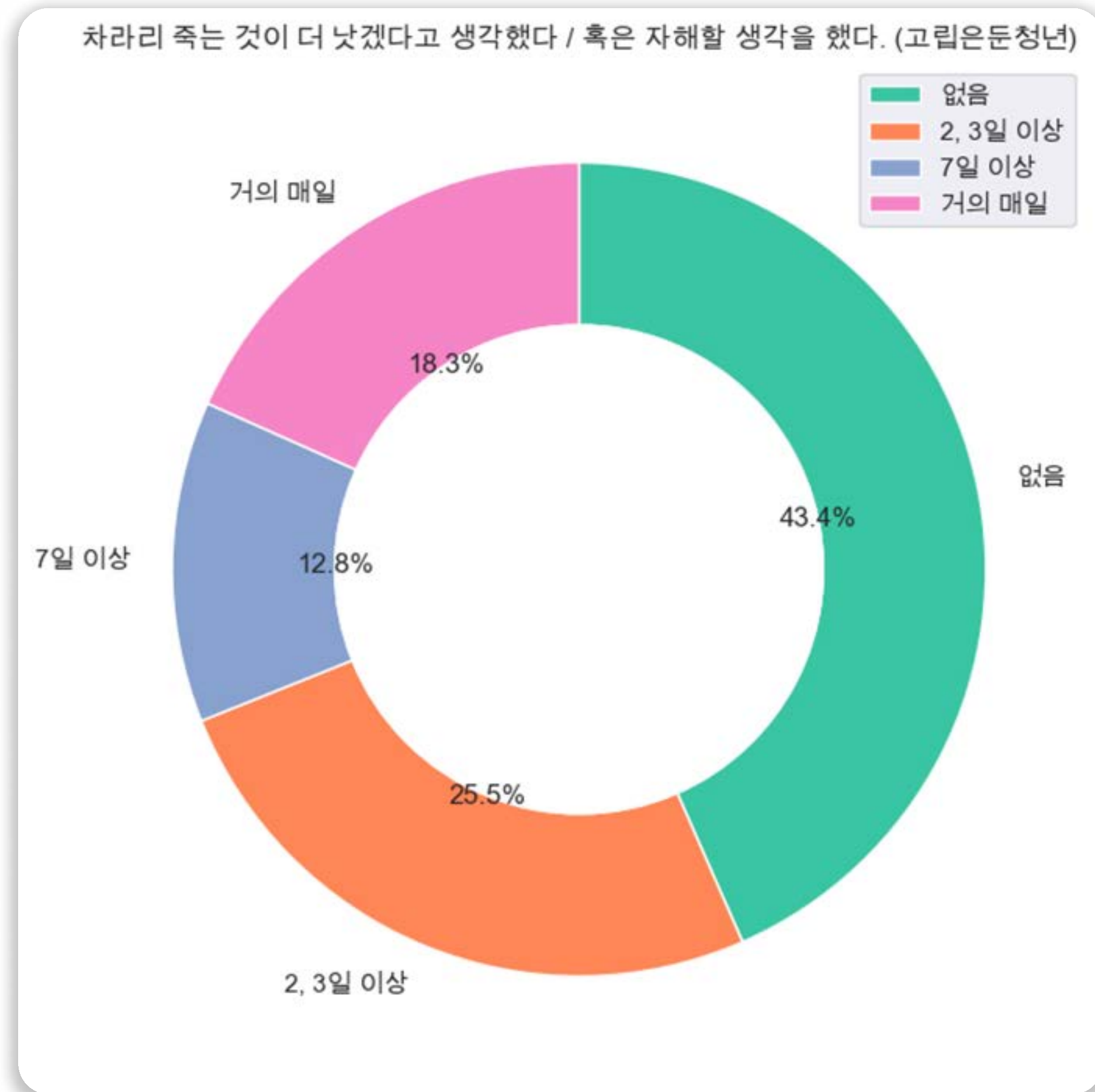
맨-휘트니 검정 실시

SQ2_X(5세단위 나이), SQ2_R(10세단위 나이), B9_3(세수/양치 빈도)

위의 세 변수가 pvalue가 0.05보다 크므로 고립은둔여부와 관계가 없는 것으로 나타났다.

SQ1, SQ2, SQ2_X, SQ2_R, SQ3, B9_3 6개의 변수를 분석과 예측모델 생성에서 제외한다.

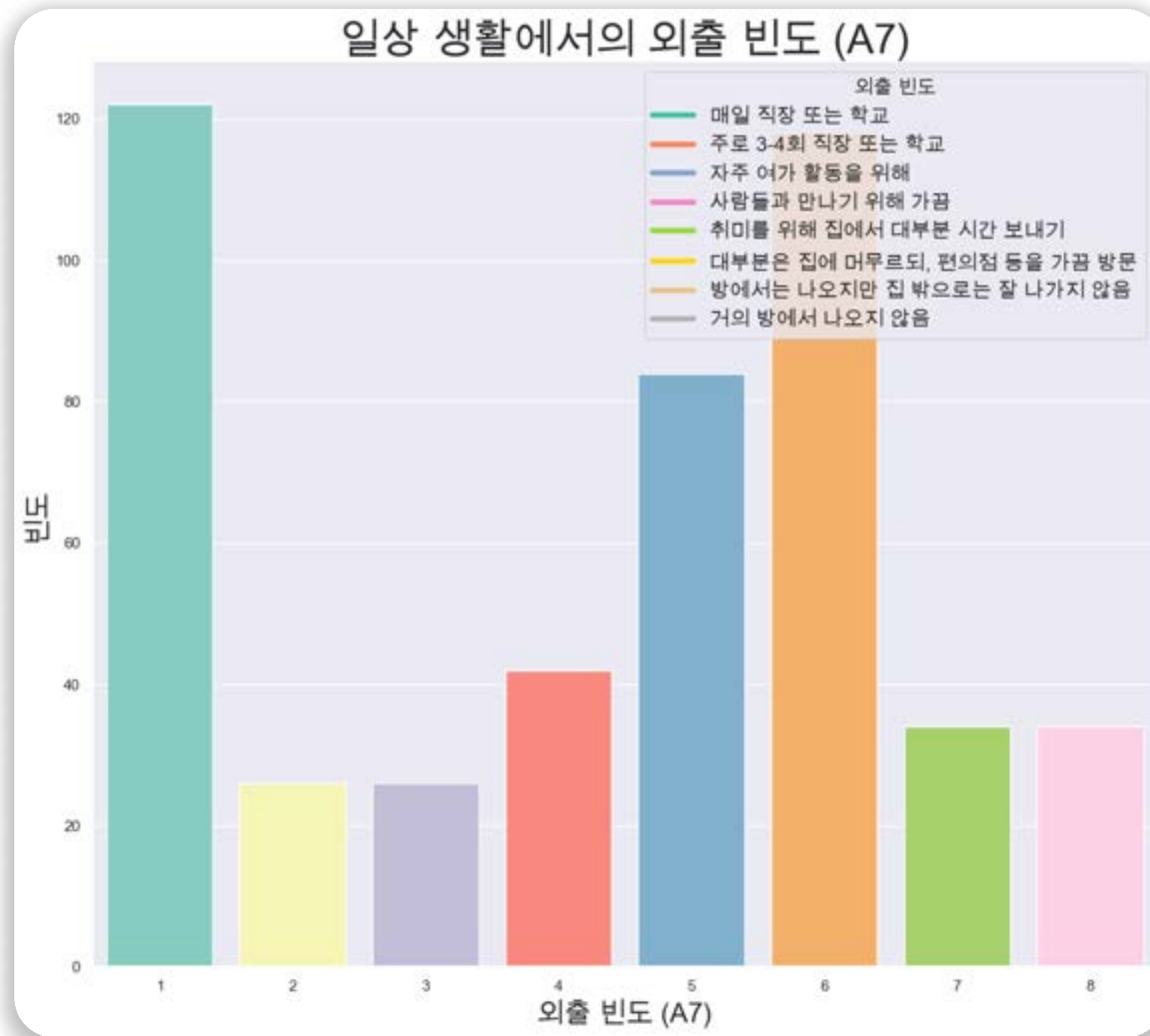
01 | 데이터 분포



고립은둔청년이 일반 청년들에 비해 자해 또는 자살 생각을 하는 비율이 높다는 것을 확인할 수 있다.

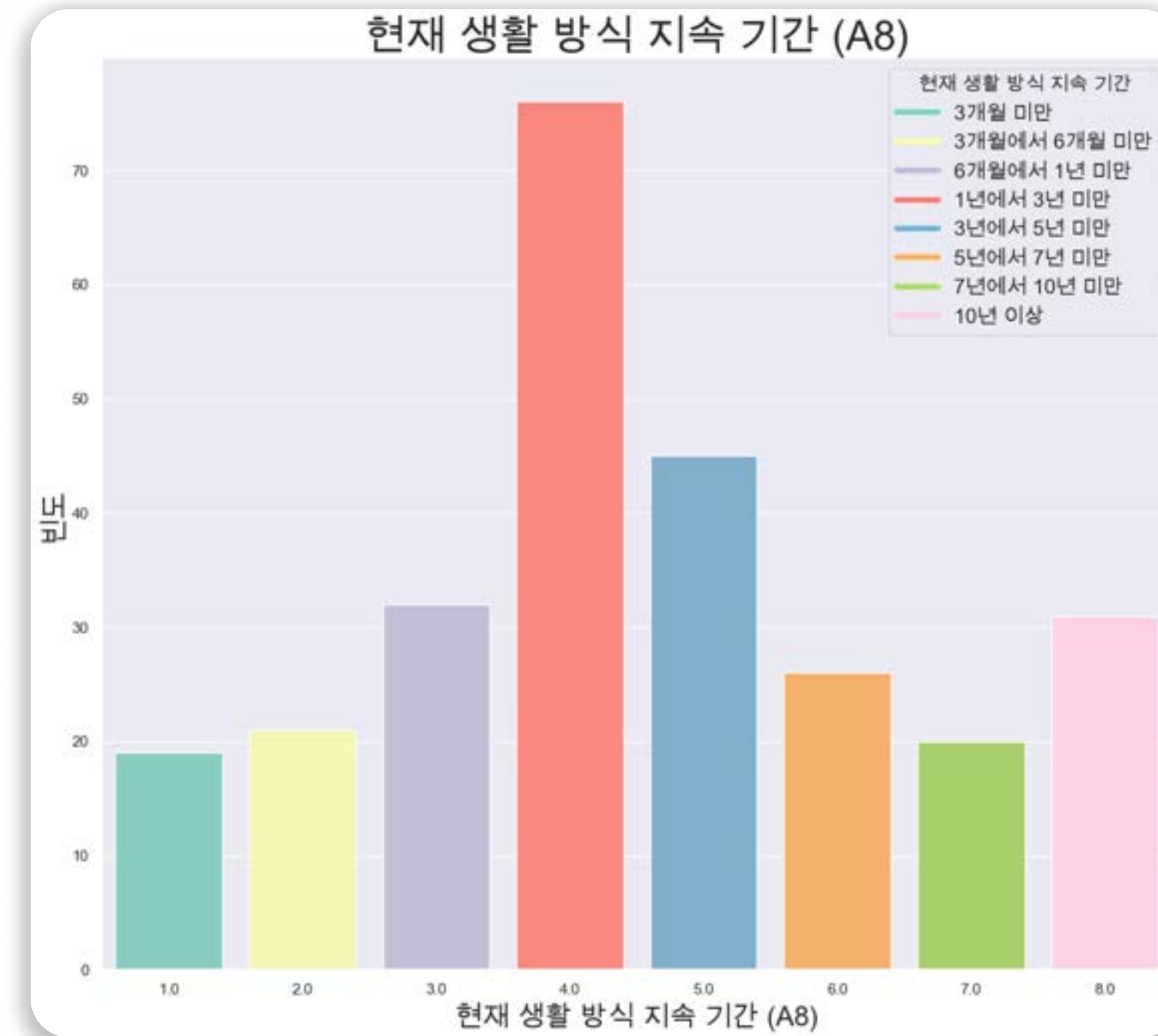
데이터 분석

01 | 데이터 분포



- 1위 : 매일 직장 또는 학교
- 2위 : 방에서는 나오지만 집 밖으로는 잘 나가지 않음
- 3위 : 자주 여가 활동을 위해

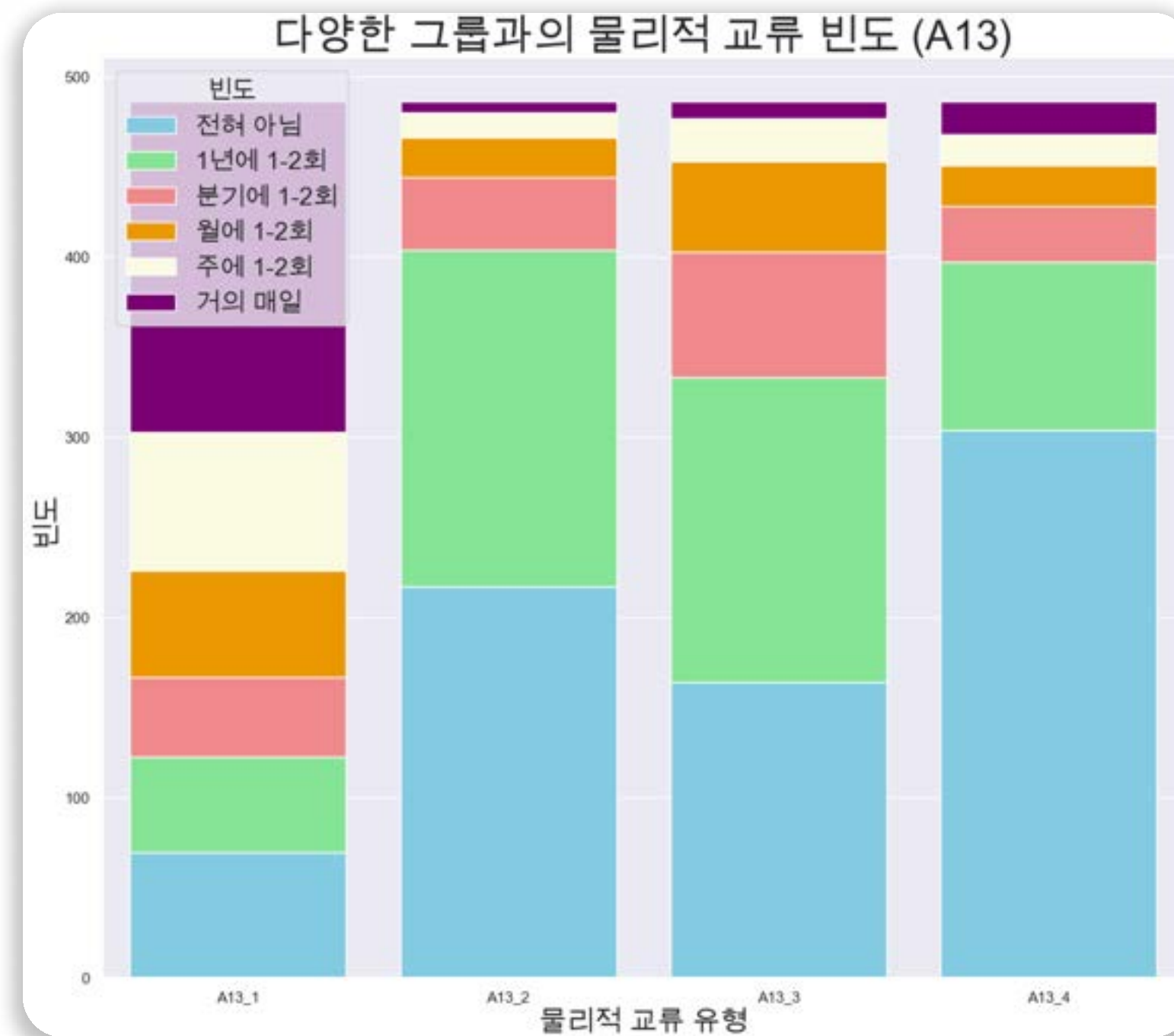
일이나 학교 외에는 대부분 집에 있는 시간이 많다는 것을 확인할 수 있다.



- 1위 : 1년에서 3년 미만
- 2위 : 3년에서 5년 미만

고립은둔 생활을 한지 1년 이상이 된 청년들이 많다는 것을 확인할 수 있다.

01 | 데이터 분포



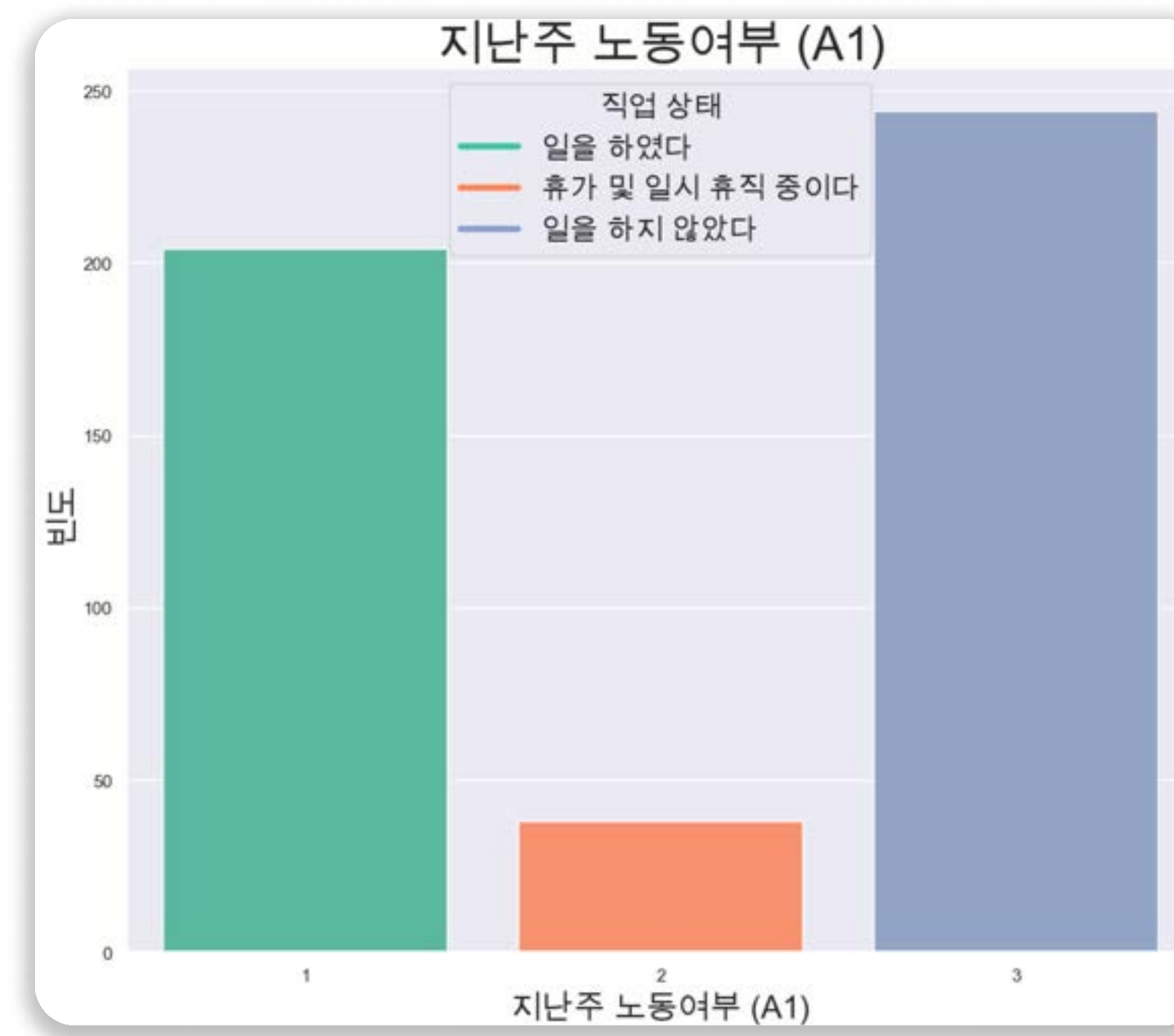
[A13_1] : 가족들과의 대면교류

[A13_2]: 친인척들과의 대면 교류

[A13_3] : 친한 친구나 친한 사람과의 대면 교류

[A13_4] : 직장, 학교, 동네에서 알고 지내는 사람들과 대면 교류
(업무 상 교류 제외)

가족을 제외하고 다른 사람들과 대면 교류를 하지 않는
비율이 높다는 것을 확인 할 수 있다.



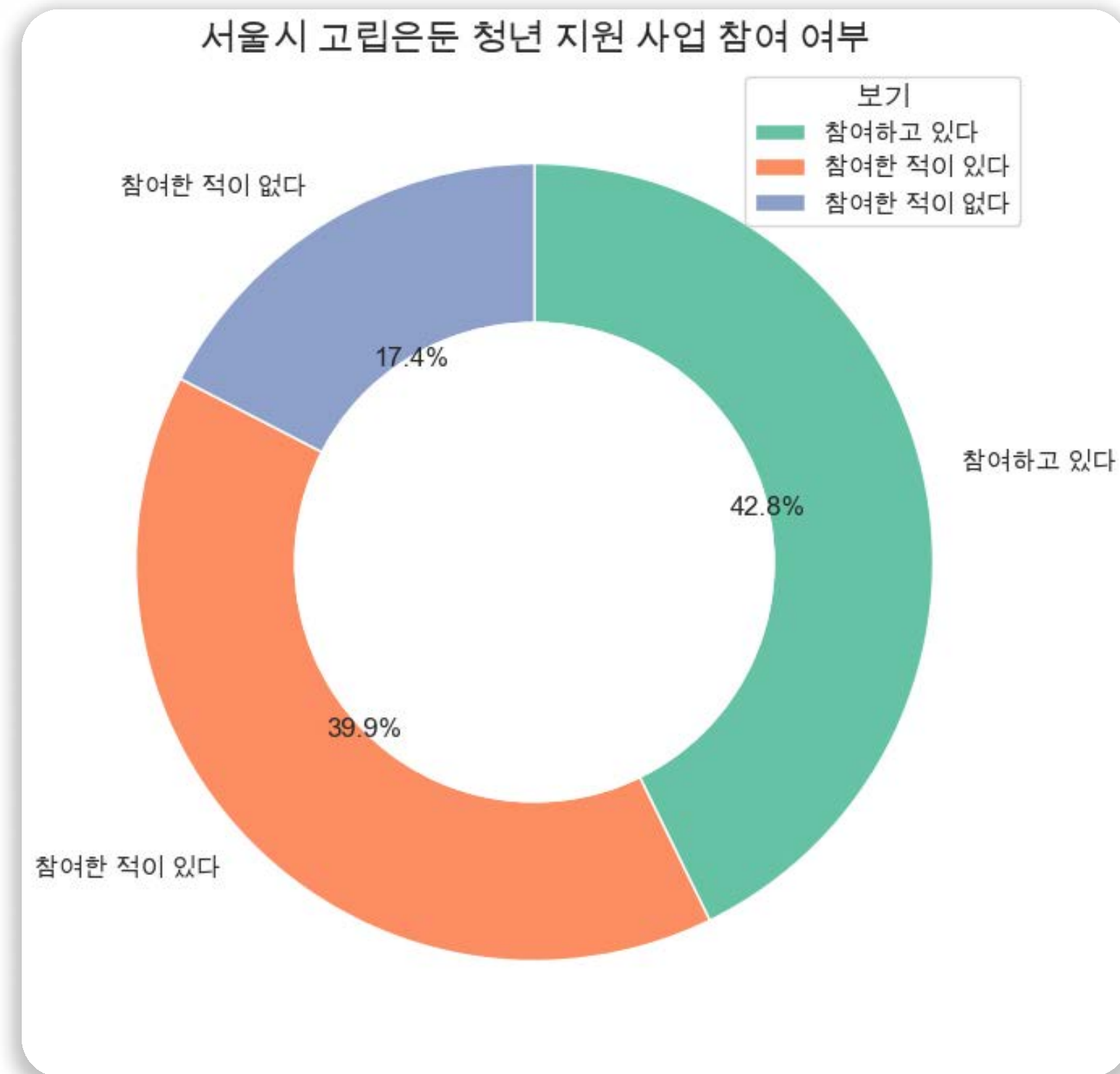
1위 : 일을 하지 않았다

2위 : 일을 하였다

3위 또한 휴가 또는 휴직 중이라는 상황을 확인해보았을 때,
일을 하지 않는 비율이 현저히 높다는 사실을 확인할 수 있다.

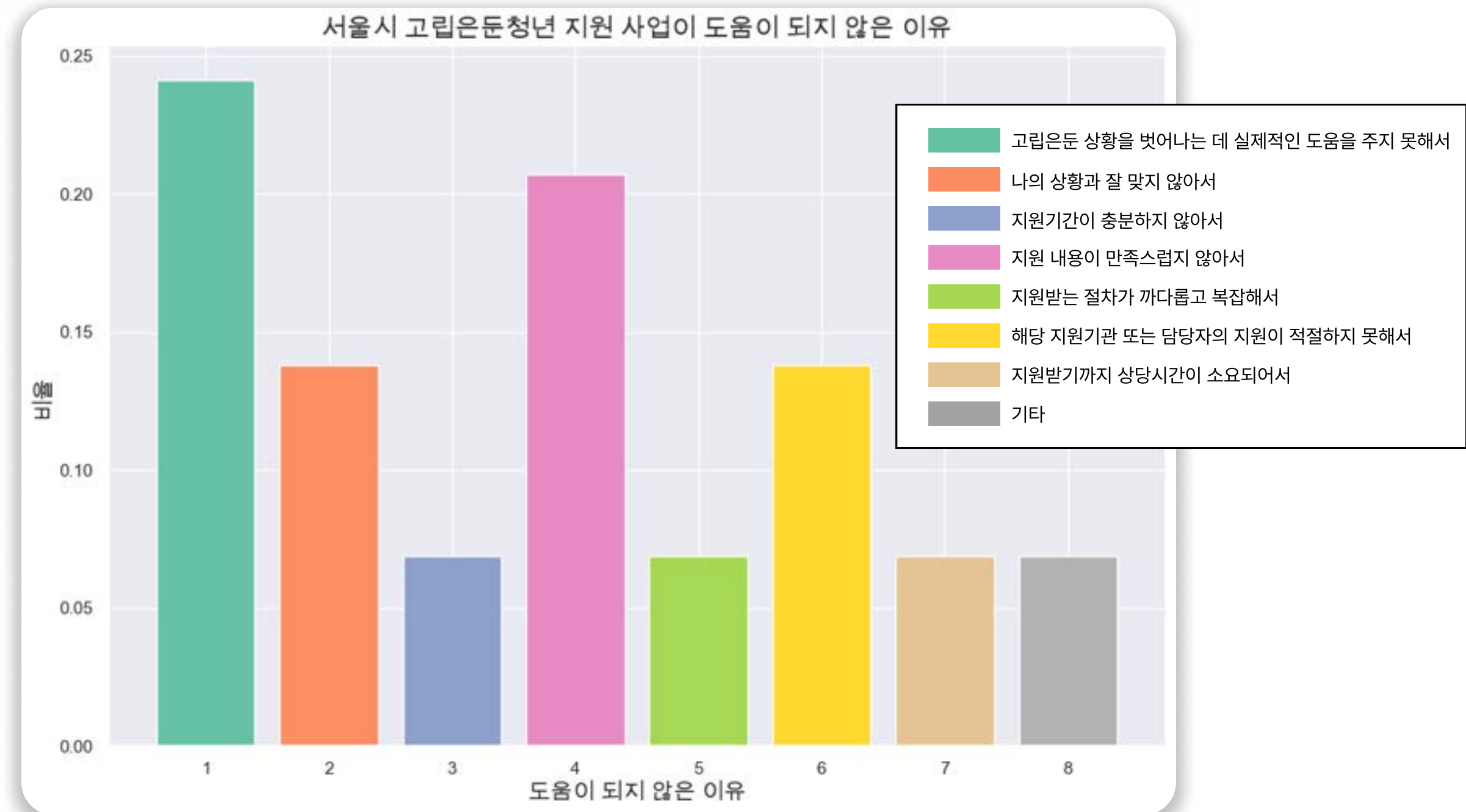
데이터 분석

01 | 데이터 분포



- 1위 : 참여하고 있다.
2위 : 참여한 적이 있다.

고립은둔청년 절반 이상이 서울시 지원 사업에 참여한 경험이 있다는 것을 확인할 수 있다.



- 1위 : 고립은둔 상황을 벗어나는 데 실제적인 도움을 주지 못해서
2위 : 지원 내용이 만족스럽지 않아서

지원 사업이 고립은둔청년에게 실질적인 도움을 주지 못한다는 것을 확인할 수 있다.

데이터 분석

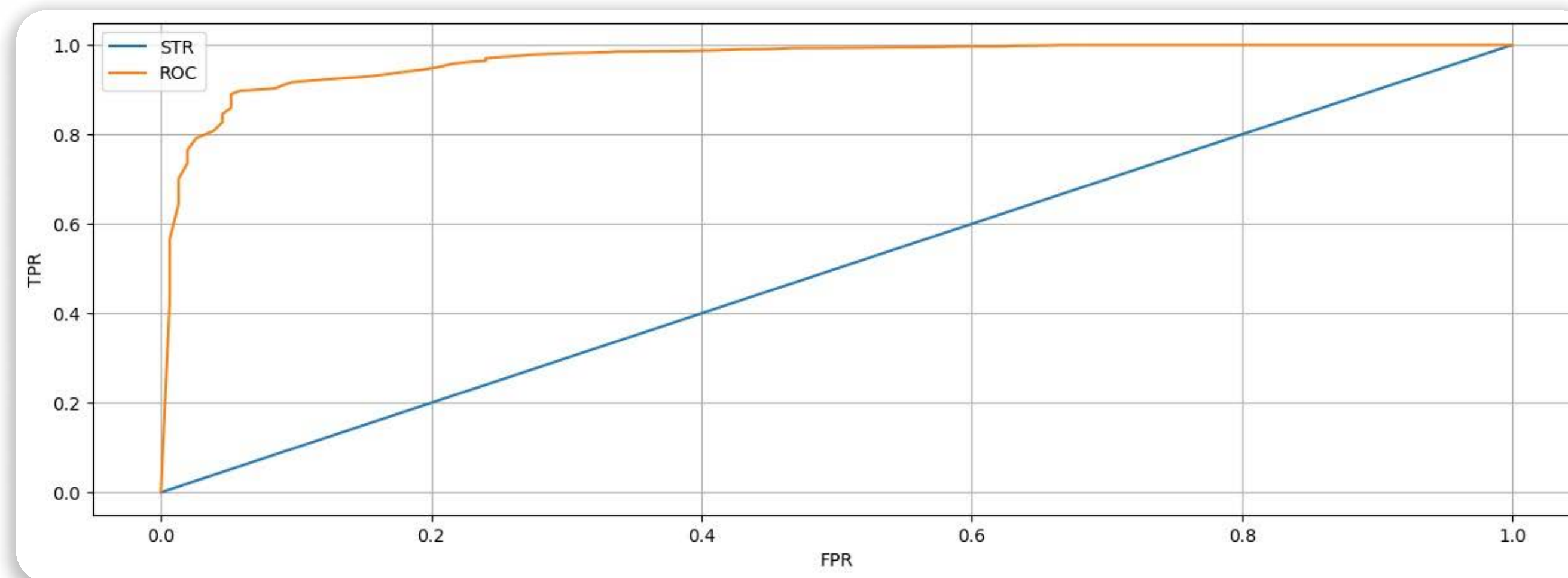
RandomForest를 이용한 예측모델

독립변수 : 결측값 처리를 마친 변수 중, 통계적 검정을 통해 선정된 변수

종속변수: 고립은둔여부인 'KEY_1' (1: 고립은둔 청년, 2: 미해당)

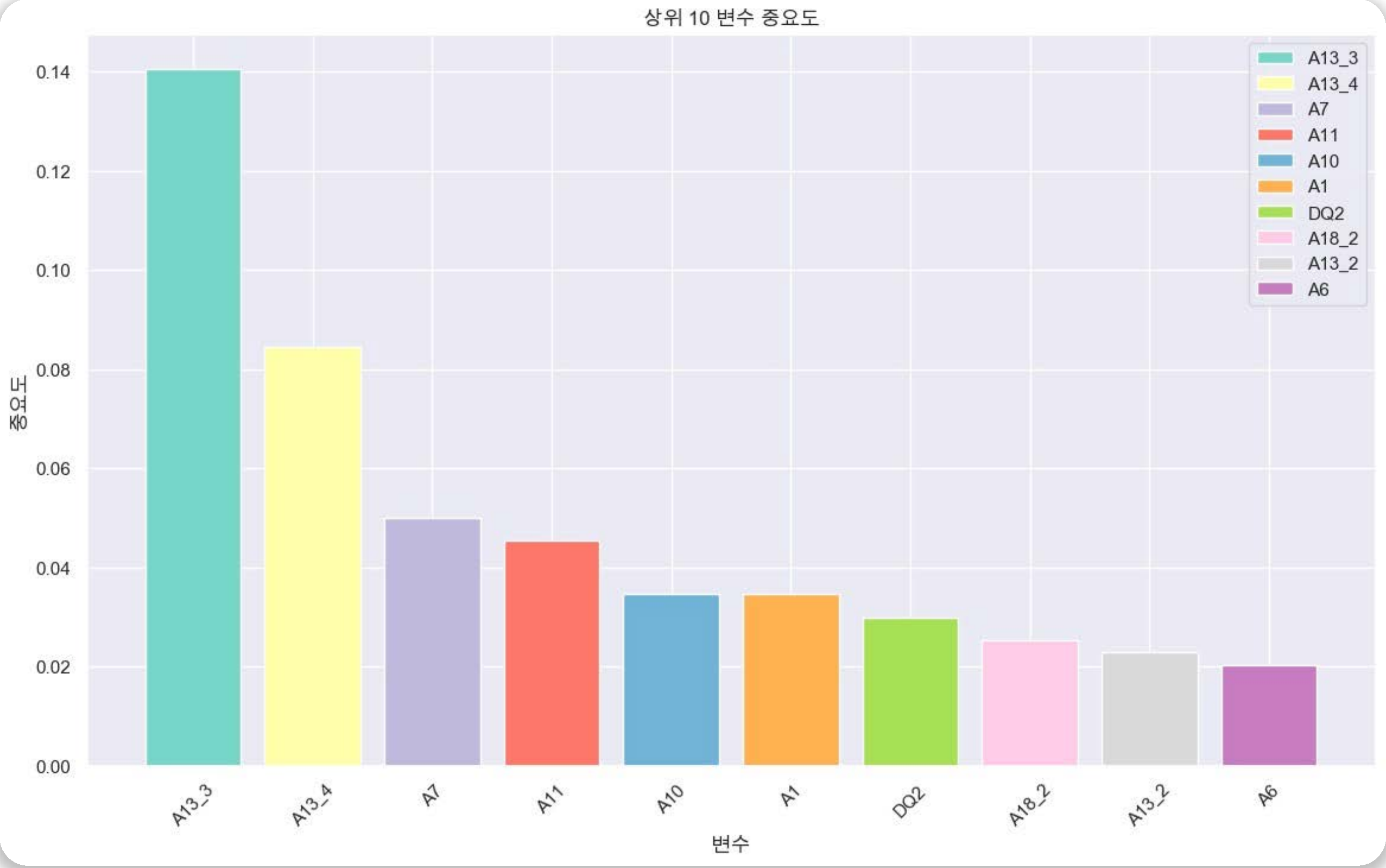
분류모델 : RandomForestClassifier

ROC 그래프



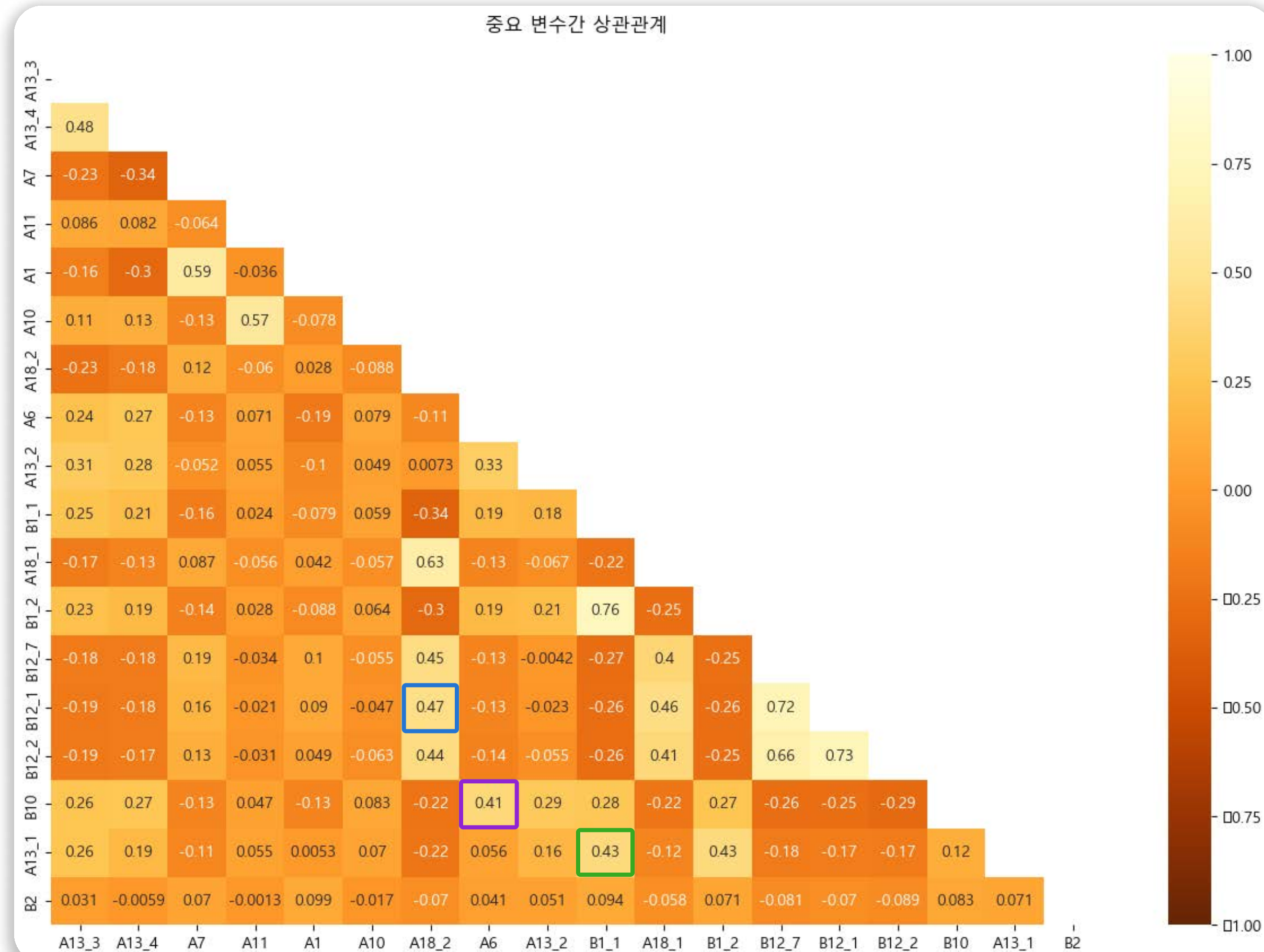
검증 데이터에 대한 성능 → ROC-AUC Score는 약 0.967

분류모델을 활용한 중요변수에 대한 시각화



순위	변수	설명
Top 1	A13_3	평소 다음의 사람들과 얼마나 자주 직접 만나 교류를 하시나요? ▶친한 친구나 친한 사람과의 대면 교류
Top 2	A13_4	평소 다음의 사람들과 얼마나 자주 직접 만나 교류를 하시나요? ▶이외 직장, 학교, 동네에서 알고 지내는 사람들과 대면 교류(업무 상 교류 제외)
Top 3	A7	여러분은 평소에 얼마나 자주 외출하시나요?
Top 4	A11	여러분이 지난 2주 동안 일상생활에서 누군가와 만나 대화를 나누거나 무언가를 같이 하며 함께 시간을 보낸 것은 전부 몇 번이었나요?
Top 5	A10	여러분이 지난 2주 동안 일상생활에서 만나 이야기를 나눈 사람은 전부 몇 명인가요?

각 변수간 상관관계를 알아보기 위한 시각화 진행



유사한 변수 간 관계를 제외하고, 의미있는 관계를 추출한 결과

1. A18_2와 B12_1 (0.47)

외로움 척도, 나는 도움을 청할 사람이 아무도 없다
 ↓
 우울 척도, 기분이 가라앉거나, 우울하거나, 희망이 없다고 느꼈다.
 도움을 청할 사람이 없다고 느낄 때, 우울감이 심해진다.

2. B1_1과 A13_1 (0.43)

평소 생활 패턴, 나는 우리 가족과 친밀한 편이다
 ↓
 물리적 고립, 가족들과의 대면 교류
 가족과의 대면 교류가 줄어들면, 친밀도 역시 떨어진다.

3. A6과 B10 (0.41)

여러분이 생각할 때 나 자신의 경제 수준은 다음 중 어디 정도인가요?
 ↓
 여러분이 생각하시기에 자신의 건강상태는 어떠한가요?

피쳐 중요도 상위 20개 변수 중, 명목형 변수인 'DQ2', 'SQ4'는 제거하고 진행

💡 예측모델을 통한 중요도가 높은 피쳐

친한 친구나 친한 사람과의 대면 교류
직장, 학교, 동네에서 알고 지내는 사람들과 대면 교류
외출 횟수

Problem

고립은둔청년은 대면 활동 참여 및 타인과의 소통 부족
으로 인해 생겨날 가능성이 높다.

Solution

고립은둔청년이 물리적으로 고립되어 있으므로
사람들과의 교류를 증진시킬 방안이 필요하다.

여기|여기|모여라

다양한 크루와 함께 사람들과의 교류

세미나 진행을 통해 팀을 매칭한다.
공통 관심사를 가진 사람들과의 취미 활동을 진행한다.

함께하기만 해도 성과금 지급

꾸준히 출석하여 활동을 기록하면 성과금을 지급한다.
사람들과 외출 후기를 작성한 후, 세미나 참여시 상품을 지급한다.

  여기여기 모여라 프로그램은 성과금이라는 목표 의식을 만들어줌으로써 고립은둔청년들이 집 밖으로 나올 계기를 만들어준다.

기존 사업들과 차이점

내 손안에 서울

2023 서울시
고립·은둔청년 지원사업 참여자 모집

괜찮아? 괜찮아

“나와, 함께”

사업 대상
서울시 거주 만 19세~39세
고립·은둔청년

신청 기간
23년 4월 25일 (화) 부터 상시모집

신청 방법
청년몽땅정보통 홈페이지
(<https://youth.seoul.go.kr>)

지원 내용
"고립·은둔청년의 상태와 욕구에 따른 맞춤형 ONE-STOP 지원"

일상회복 및 심리정서
사회성 및 관계 향상
진로·취업·일 역량 향상
일경험 지원

서울특별시

내 손안에 서울

2023년
**고립·은둔 청년
반려식물 보급사업**

지원대상
만 19세~39세 서울 거주 청년 중
고립·은둔청년 지원사업 대상자

지원내용
반려식물 보급 및
원예프로그램 운영

신청방법
청년몽땅정보통 온라인 신청
youth.seoul.go.kr

문의 서울 다산콜센터(02-120) 청년몽땅정보통 youth.seoul.go.kr

고립·은둔청년이라는 단어를 직접적으로 사용하지 않는다.

당사자들에게 거부감을 불러일으킬 수 있으므로, 친근한 명칭을 만들어 접근한다.

명목상으로 지급되는 지원비보단 성과금 형식으로 지급한다.

출석시 성과금 지급으로 청년들이 밖으로 나오도록 유도한다.

정서적 기대효과

- 이용자는 프로그램의 참여를 통해 대면으로 고립은둔청년을 만나고 정서적 지지를 얻음으로써 사회 관계망을 형성할 수 있다.
- 고립은둔청년들은 프로그램을 통해 소속감을 갖고, 서로에 대한 유대감을 형성할 수 있다.

사회적 기대효과

- 저출산 등 사회 전반의 문제들을 완화하고 축소시킬 수 있다.
- 출석시 혜택을 지급하는 시스템을 통해 프로그램으로써 비용의 부담을 줄이고 영리 목적의 이용을 차단할 수 있다.
- 이후 '여기여기 모여라'를 통해 극복의지를 보인 고립은둔청년들의 이야기를 담아 사회인식개선에 기여한다.

참고 자료 및 문헌

조용성, [취재N팩트] '취업실패·대인관계' 때문에...고립·은둔청년 54만 명, YTN 뉴스, 2023년 12월 14일자

김다솜, 그냥 쉬는 '니트' 청년 증가세...지원책은 충분할까?, 데일리팝, 2023년 5월 14일자

홍예지, "절반이 중증 우울증" 어둠속에 방치된 61만명, 파이낸셜 뉴스, 2023년 8월 15일자

최훈진 외 3명, 외로움이 불러올 미래는...코로나 기간 우울증 환자 10년 새 최대폭 '증가', 서울신문, 2022년 1월 3일자

임종언, "외로움은 사회의 질병"...뉴욕, 외로움 명예대사 임명, 코메디닷컴, 2023년 11월 14일자

김종화, [뉴스속 용어]장애 인정 추진되는 '사회적 장애', 아시아경제, 2023년 3월 10일자

활용 데이터셋

파일데이터명	서울특별시_고립은둔 청년 실태조사 결과
제공기관명	서울특별시
출처	공공데이터포털 ("data.go.kr")
활용 파일	★DATA_서울시 고립은둔청년 실태조사(청년조사)_v1_230127.xlsx (5513행, 252열)