

2024 MID - PROJECT DATA ANALYSIS

2024.04.03

목차

주제 1 : 연령대/특성별 행정동 추천

1. 주제선정 배경
2. 데이터 수집 및 전처리
 - b. 기본 전처리
 - c. 소분류 통합
 - d. Segement 추가 분석
3. 분석결과
 - a. 연령대별 선호지역
 - b. 특성별 선호지역

주제 2 : 소득수준별 가치관 차이 및 소득 예측

1. 주제선정 배경
2. 데이터 수집 및 전처리
 - a. 데이터 수집
 - b. 기본 전처리
3. 시각화
4. 모델링 준비
5. 모델링

Conclusion

1. 활용방안
2. 한계점 및 자체평가
3. 기술사양 및 프로젝트 공정계획표

01. Project Overview

프로젝트 팀구성 및 역할

구성 및 역할

강 건

주제선정 & 자료탐색
데이터 전처리
VS 코딩 / 소스코딩 정리
인사이트 도출

김효정

주제선정 & 자료탐색
VS 코딩작업
인사이트 도출
문서정리

Project

연령대/특성별 행정동 추천

1

Project 1 : 주제선정 배경

BACKGROUND

여러분은 어느 지역에서 살고 싶으신가요?

유명한 곳? 집값이 비싼 곳이 과연 나에게도 딱! 맞는 살기 좋을까요?

역세권

대중교통 역에
가까워 접근성이
좋은 지역

학세권

좋은 학교나 교육 시설에
가까운 지역교육
환경을 중시

숲세권

공원이나 숲과 같은
자연 친화적인 환경에
가까운 지역

몰세권

대형 쇼핑몰, 상업 시설
에 인접해 편의성이 높은
지역

Conclusion

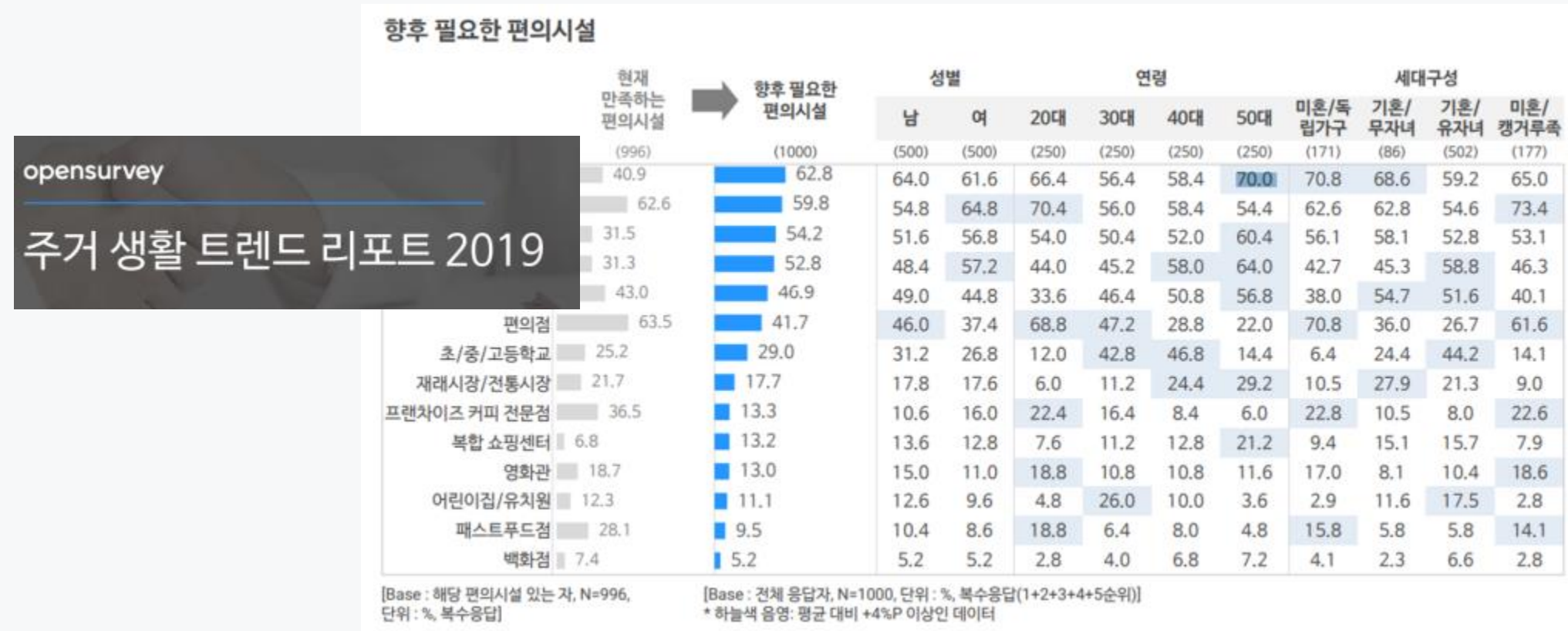
연령별/특성별 살기좋은 지역 추천

Project 1 : 데이터 수집

DATA COLLECTION

총 features 17 / segment 4개(20,30,40,50대) / row 426개(행정동)

변수	출처
주거환경 요인 선호도	오픈서베이
녹지비율	산림빅데이터거래소
교통사고안전	TAAS 교통사고분석시스템
미세먼지 / 집값 / 범죄안전 / 감염병안전 / 편의점 / 프랜차이즈커피전문점 / 패스트푸드점 / 시장 / 백화점 / 버스정류장 / 지하철역 / 초중고 / 병원 / 영화관 / 유치원	서울 열린데이터 광장



Project 1 : 데이터 전처리

DATA PREPROCESS

1 기본 전처리

- Null/이상치 처리
- 행정동 <-> 법정동 기준 통일
- 테이블 병합
- 단위 조정(동별 인구, 면적등)
- 스케일링 (Min-Max)

2 소분류 -> 대분류

3 Segment 추가 분석

```
9 from sklearn.preprocessing import MinMaxScaler
```

데이터 읽어오기
1 cell hidden --

전처리

기준이 되는 행정동 코드 자료 가져오기 (Code list)
1 cell hidden --

Features : 병원, 유치원, 초중고, 백화점, 버스정거장, 지하철역 작업
10 cells hidden --

Feature : 교통사고 수
16 cells hidden --

Feature : cctv개수 (전처리의 어려움 때문에 사용X)
15 cells hidden --

Feature : 방범지수 (행정동 단위의 방범지수 데이터는 행정동 단위로만 존재함)

Features : 패스트푸드점, 프렌차이즈전문커피점, 편의점(ppp)
18 cells hidden --

Feature : 시장
7 cells hidden --

Feature : 공원면적
10 cells hidden --

Feature : 공시지가
32 cells hidden --

주거환경 요인 선호도 데이터 작업 (Scale)
22 cells hidden --

행정동별 인구특성 수 및 면적 데이터 정리 : feature들을 절대적인 개수가 아닌 인구 혹은 면적으로 나눠서 비율을 사용하려는 의도
15 cells hidden --

Features : 영화관, 범죄, 감염병 (행정동별 데이터가 없고 구 단위 데이터 사용한 Features)

Project 1 : 데이터 전처리

DATA PREPROCESS

1 기본 전처리

-> 결과

점수 합산 예시:

$\text{sum}(\text{각 feature 선호도 가중치} * \text{행정동의 각 feature 값})$

2 소분류 -> 대분류

3 Segment 추가 분석

1 final_20.sort_valu
✓ 0.0s

	시군구	행정구역
214	마포구	서교동
149	도봉구	도봉2동
194	서대문구	충현동
198	서대문구	홍제3동
154	도봉구	방학1동
52	성동구	행당1동
147	도봉구	창5동
207	마포구	공덕동
210	마포구	대흥동
169	노원구	상계2동

1 final_30.sort_valu
✓ 0.0s

	시군구	행정구역
214	마포구	서교동
149	도봉구	도봉2동
148	도봉구	도봉1동
154	도봉구	방학1동
147	도봉구	창5동
194	서대문구	충현동
198	서대문구	홍제3동
2	종로구	삼청동
276	금천구	가산동
18	중구	회현동

1 final_40.sort_valu
✓ 0.0s

	시군구	행정구역
214	마포구	서교동
2	종로구	삼청동
149	도봉구	도봉2동
147	도봉구	창5동
198	서대문구	홍제3동
154	도봉구	방학1동
194	서대문구	충현동
148	도봉구	도봉1동
207	마포구	공덕동
52	성동구	행당1동

1 final_50.sort_valu
✓ 0.0s

	시군구	행정구역
214	마포구	서교동
2	종로구	삼청동
149	도봉구	도봉2동
147	도봉구	창5동
198	서대문구	홍제3동
207	마포구	공덕동
148	도봉구	도봉1동
154	도봉구	방학1동
194	서대문구	충현동
221	마포구	상암동

Project 1 : 데이터 전처리

DATA PREPROCESS

1 기본 전처리 + Features 선호도 Rank

1 features_scaled_df

	대분류	소분류	20대	30대	40대	50대
0	자연	미세먼지	0.455621	0.809160	0.852713	0.801205
1	자연	녹지비율	0.455621	0.809160	0.852713	0.801205
2	주택	집값	0.289941	0.274809	0.116279	0.204819
3	안전	범죄안전	0.615385	0.496183	0.441860	0.373494
4	안전	감염병안전	0.615385	0.496183	0.441860	0.373494
5	안전	교통사고안전	0.615385	0.496183	0.441860	0.373494
6	편의시설	편의점	0.976331	0.824427	0.426357	0.277108
7	편의시설	프렌차이즈커피전문점	0.289941	0.236641	0.031008	0.036145
8	편의시설	패스트푸드점	0.236686	0.045802	0.023256	0.018072
9	편의시설	시장	0.047337	0.137405	0.341085	0.385542
10	편의시설	백화점	0.000000	0.000000	0.000000	0.054217
11	교통	버스정류장	1.000000	0.992366	1.000000	0.765060
12	교통	지하철역	0.940828	1.000000	1.000000	1.000000
13	교육	초중고	0.136095	0.740458	0.775194	0.162651
14	복지문화	병원	0.609467	0.786260	0.992248	0.909639
15	복지문화	영화관	0.236686	0.129771	0.077519	0.120482
16	복지문화	유치원	0.029586	0.419847	0.062016	0.000000



	대분류	소분류	20대_rank	30대_rank	40대_rank	50대_rank
0	자연	미세먼지	9.5	13.5	13.5	14.5
1	자연	녹지비율	9.5	13.5	13.5	14.5
2	주택	집값	7.5	6.0	6.0	7.0
3	안전	범죄안전	13.0	9.0	10.0	10.0
4	안전	감염병안전	13.0	9.0	10.0	10.0
5	안전	교통사고안전	13.0	9.0	10.0	10.0
6	편의시설	편의점	16.0	15.0	8.0	8.0
7	편의시설	프렌차이즈커피전문점	7.5	5.0	3.0	3.0
8	편의시설	패스트푸드점	5.5	2.0	2.0	2.0
9	편의시설	시장	3.0	4.0	7.0	12.0
10	편의시설	백화점	1.0	1.0	1.0	4.0
11	교통	버스정류장	17.0	16.0	16.5	13.0
12	교통	지하철역	15.0	17.0	16.5	17.0
13	교육	초중고	4.0	11.0	12.0	6.0
14	복지문화	병원	11.0	12.0	15.0	16.0
15	복지문화	영화관	5.5	3.0	5.0	5.0
16	복지문화	유치원	2.0	7.0	4.0	1.0

Project 1 : 데이터 전처리

DATA PREPROCESS

1 기본 전처리

+ Features 선호도 Rank 결과

1 final_20_rank.sort

✓ 0.0s

	시군구	행정구역
214	마포구	서교동
195	서대문구	신촌동
149	도봉구	도봉2동
154	도봉구	방학1동
147	도봉구	창5동
153	도봉구	쌍문4동
276	금천구	가산동
198	서대문구	홍제3동
194	서대문구	충현동
221	마포구	상암동

1 final_30_rank.sort

✓ 0.0s

	시군구	행정구역
148	도봉구	도봉1동
214	마포구	서교동
195	서대문구	신촌동
153	도봉구	쌍문4동
198	서대문구	홍제3동
154	도봉구	방학1동
221	마포구	상암동
147	도봉구	창5동
2	종로구	삼청동
149	도봉구	도봉2동

1 final_40_rank.sort

✓ 0.0s

	시군구	행정구역
148	도봉구	도봉1동
153	도봉구	쌍문4동
214	마포구	서교동
195	서대문구	신촌동
154	도봉구	방학1동
149	도봉구	도봉2동
147	도봉구	창5동
198	서대문구	홍제3동
221	마포구	상암동
276	금천구	가산동

1 final_50_rank.sort

✓ 0.0s

	시군구	행정구역
153	도봉구	쌍문4동
195	서대문구	신촌동
154	도봉구	방학1동
148	도봉구	도봉1동
149	도봉구	도봉2동
147	도봉구	창5동
214	마포구	서교동
198	서대문구	홍제3동
276	금천구	가산동
152	도봉구	쌍문3동

Project 1 : 데이터 전처리

DATA PREPROCESS

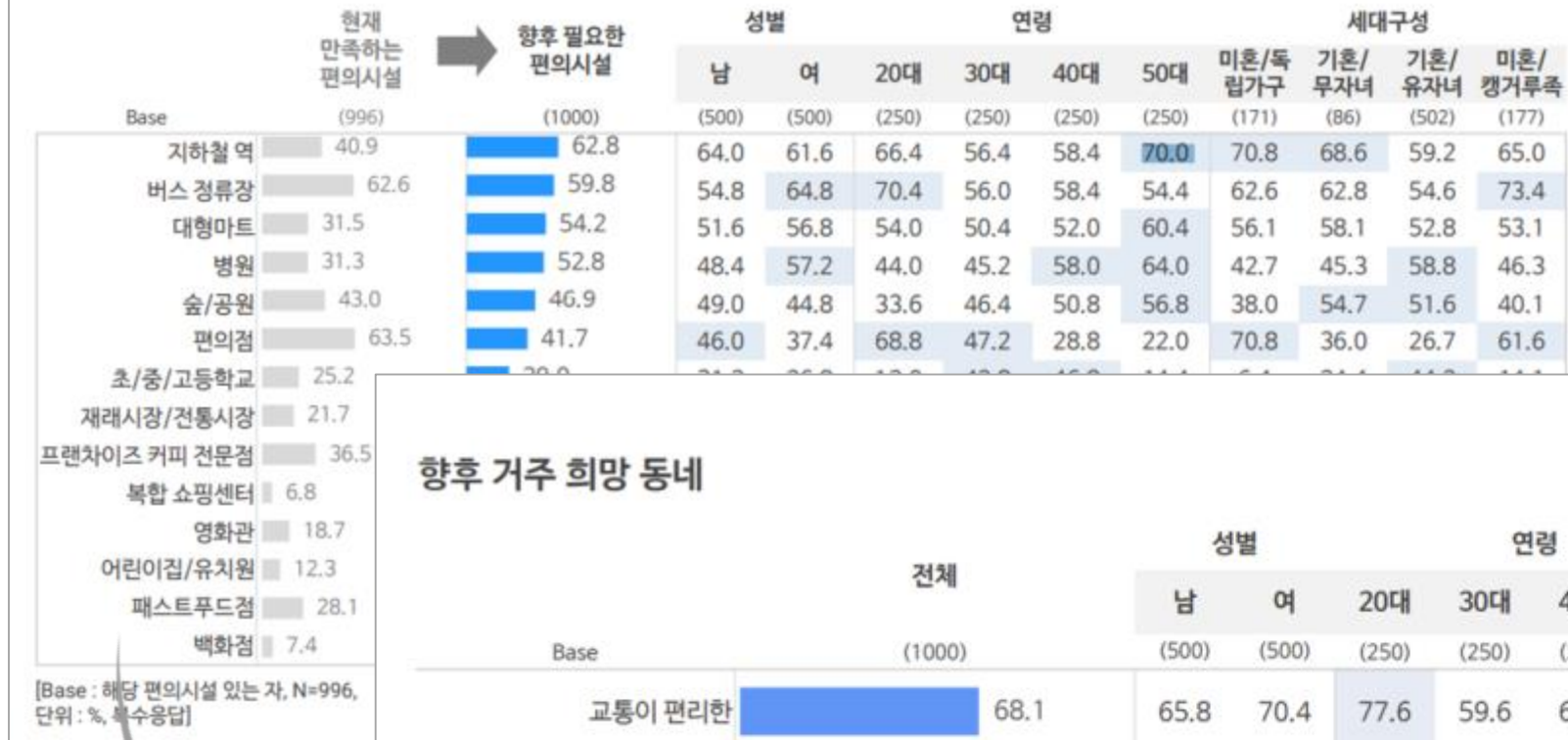
2 소분류 → 대분류

- 17개 소분류 -> 7개 대분류
- 무의미한 feature drop (예: 미세먼지)
- 유의미한 new feature add (예: 학원수, 음식점수 등)
- 특정 그룹으로 쏠림 방지

3 Segment 추가 분석

연령 (20,30,40,50)
+ 미혼(독립가구), 기혼(무자녀), 기혼(유자녀), 미혼(캥거루족)

향후 필요한 편의시설



향후 거주 희망 동네

Base	전체 (1000)	성별		연령				세대구성			
		남 (500)	여 (500)	20대 (250)	30대 (250)	40대 (250)	50대 (250)	미혼/독립가구 (171)	기혼/무자녀 (86)	기혼/유자녀 (502)	미혼/캥거루족 (177)
교통이 편리한	68.1	65.8	70.4	77.6	59.6	65.2	70.0	72.5	68.6	61.8	81.9
상업/편의시설이 잘 갖추어진	54.9	51.2	58.6	66.4	51.6	51.2	50.4	67.3	41.9	49.4	67.8
자연 환경이 좋은	46.5	49.2	43.8	27.6	36.8	53.6	68.0	30.4	52.3	55.8	33.3
주변 치안이 안전한	41.6	33.8	49.4	51.6	41.6	37.2	36.0	45.0	41.9	34.5	57.1
향후 발전/개발 가능성이 있는	26.9	30.4	23.4	21.6	29.2	26.4	30.4	19.3	39.5	28.7	24.9
아이를 키우기 좋은	25.4	28.6	22.2	16.0	45.6	33.6	6.4	12.9	29.1	36.5	8.5
집값이 저렴한	17.7	20.4	15.0	22.4	18.4	12.8	17.2	30.4	14.0	14.7	14.1

[Base : 전체 응답자, N=1000, 단위 : %, 복수응답(1+2+3순위)]

* 하늘색 음영: 평균 대비 +4%p 이상인 데이터

©opensurvey

Project 1 : 연령대별 선호지역

ANALYSIS RESULTS

20대

시군구	행정구역
37 금천구	가산동
96 강서구	가양1동
347 도봉구	도봉1동
367 성동구	성수2가3동
74 종로구	종로1.2.3.4가동
247 도봉구	창5동
142 송파구	문정2동
217 도봉구	방학1동
28 도봉구	쌍문3동
332 송파구	가락본동

30대

시군구	행정구역
347 도봉구	도봉1동
74 종로구	종로1.2.3.4가동
37 금천구	가산동
3 강동구	상일2동
96 강서구	가양1동
179 노원구	상계1동
340 노원구	중계4동
235 은평구	진관동
360 종로구	청운효자동
262 서초구	양재1동

40대

시군구	행정구역
347 도봉구	도봉1동
179 노원구	상계1동
340 노원구	중계4동
156 서초구	내곡동
360 종로구	청운효자동
74 종로구	종로1.2.3.4가동
234 강남구	일원본동
262 서초구	양재1동
3 강동구	상일2동
373 중랑구	면목4동

50대

시군구	행정구역
347 도봉구	도봉1동
179 노원구	상계1동
340 노원구	중계4동
156 서초구	내곡동
373 중랑구	면목4동
3 강동구	상일2동
360 종로구	청운효자동
234 강남구	일원본동
262 서초구	양재1동
299 송파구	오륜동

Project 1 : 특성별 선호지역

ANALYSIS RESULTS

미혼
(독립가구)

	시군구	행정구역
347	도봉구	도봉1동
37	금천구	가산동
156	서초구	내곡동
235	은평구	진관동
96	강서구	가양1동
340	노원구	중계4동
179	노원구	상계1동
28	도봉구	쌍문3동
242	강서구	공항동
217	도봉구	방학1동

기혼
(무자녀)

	시군구	행정구역
3	강동구	상일2동
0	성북구	길음2동
1	송파구	거여2동
4	서대문구	홍제3동
2	은평구	증산동
5	송파구	위례동
6	구로구	고척1동
347	도봉구	도봉1동
179	노원구	상계1동
7	종로구	망우본동

기혼
(유자녀)

	시군구	행정구역
347	도봉구	도봉1동
360	종로구	청운효자동
74	종로구	종로1.2.3.4가동
179	노원구	상계1동
262	서초구	양재1동
234	강남구	일원본동
318	마포구	상암동
340	노원구	중계4동
235	은평구	진관동
118	종로구	혜화동

미혼
(캥거루족)

	시군구	행정구역
347	도봉구	도봉1동
96	강서구	가양1동
179	노원구	상계1동
37	금천구	가산동
340	노원구	중계4동
3	강동구	상일2동
0	성북구	길음2동
247	도봉구	창5동
28	도봉구	쌍문3동
217	도봉구	방학1동

Project 1 : 한계점

LIMITATIONS

01 변별력

segment별
차별화된 요소부족
(예: 교통수단 선호)

02 중복응답

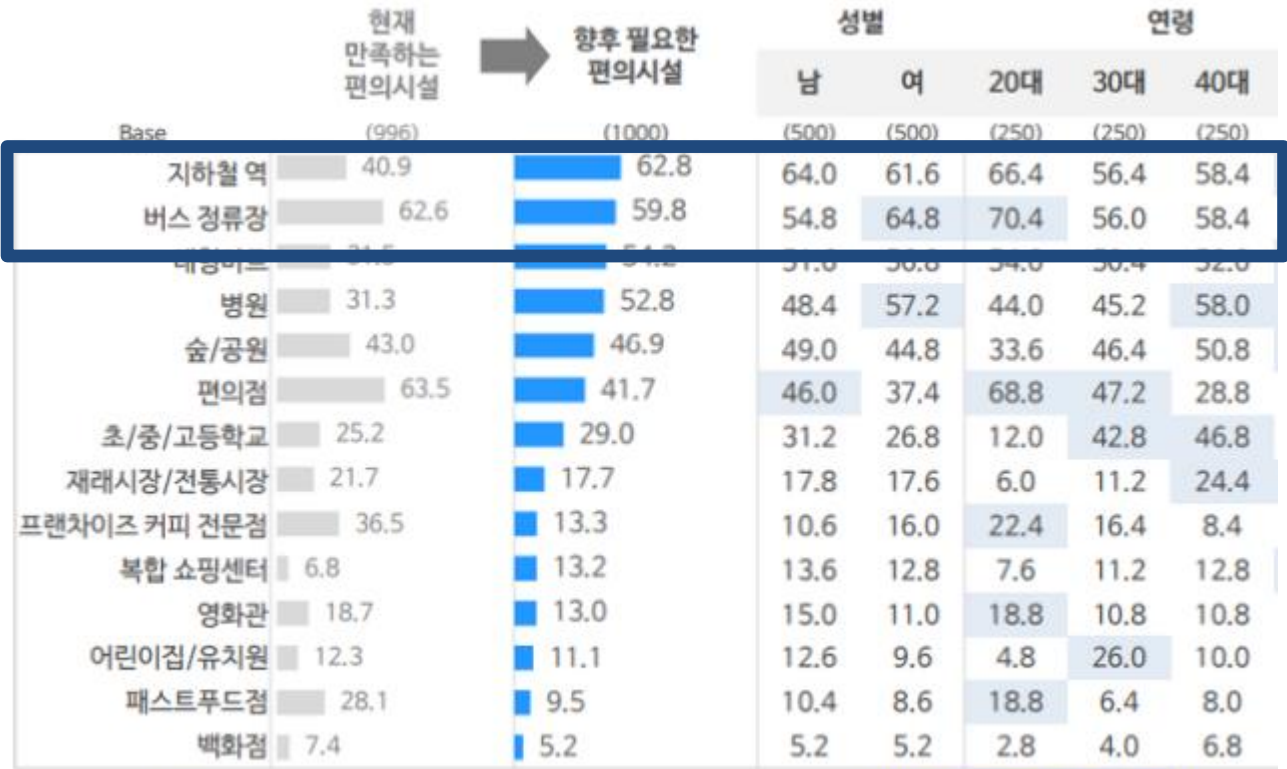
중복응답으로 인해
구분이 명확하지 않음

03 신뢰성

분석목적에 맞는 자료부족
및
신뢰도가 높지 않은 조사기관

opensurvey

향후 필요한 편의시설



[Base : 해당 편의시설 있는 자, N=996, 단위 : %, 복수응답]
[Base : 전체 응답자, N=1000, 단위 : %, 복수응답(1+2+3+4+5순위)]
* 하늘색 음영: 평균 대비 +4%p 이상인 데이터

04 통일성

매칭기준이 상이하여
전처리에 많은 시간 소요
예) 법정동 vs 행정동

D	E	F	G	H	I
행정동 (행정기관명)	법정동	행정구역 분류	행정기관코드	행정기관 생성일	법정동코드
청운호자동	청운동	11010720	1111051500	20081101	1111010100
청운호자동	신교동	11010720	1111051500	20081101	1111010200
청운호자동	궁정동	11010720	1111051500	20081101	1111010300
청운호자동	호자동	11010720	1111051500	20081101	1111010400
청운호자동	찰성동	11010720	1111051500	20081101	1111010500
청운호자동	통인동	11010720	1111051500	20081101	1111010800
청운호자동	누상동	11010720	1111051500	20081101	1111010900
청운호자동	누하동	11010720	1111051500	20081101	1111011000
청운호자동	옥인동	11010720	1111051500	20081101	1111011100
청운호자동	세종로	11010720	1111051500	20081101	1111011900

Project

2

소득수준별 가치관 차이 및 소득 예측

Project 2 : 주제선정 배경

BACKGROUND

소득수준별로 삶의 질, 의식, 가치관의 유의미한 차이가 있을까?

가치관

예: 신뢰도(가족, 친구 등),
계층 이동 가능성 등

여가활동

예: 여가 유형,
여가 시간 등

.....

삶의질

예: 일상 생활 스트레스,
행복지수 등

행동유형

예: 통근/통학 여부,
물품 구매 경로 등

2022 서울서베이를 통한 소득수준 예측

Project 2 : 데이터 수집 및 전처리

DATA COLLECTION & PREPROCESSING

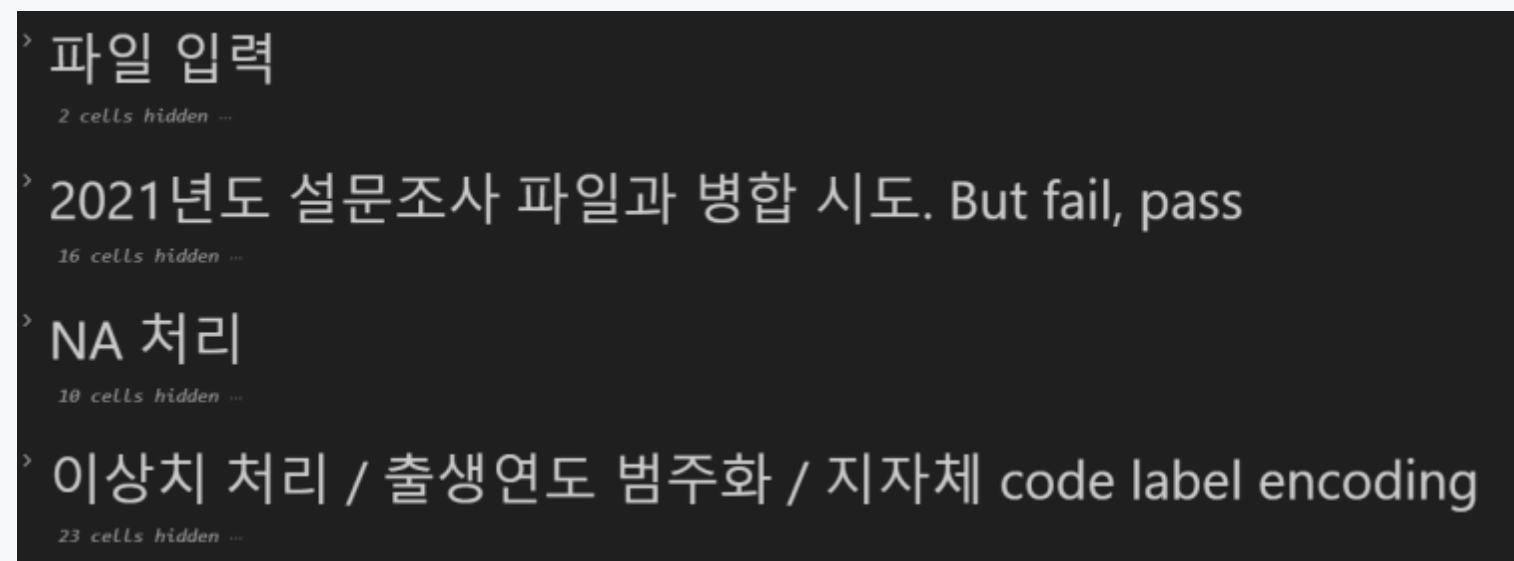
1 데이터 수집

- 서울서베이 (2022)
- row 약 4만개 (응답자 수)
- feature 약 200개 (설문항목)



2 기본 전처리

- 결측치/이상치 처리
- 수치형 → 범주화 등
- Feature 자체 선별(예: 통근/통학 지역, 여가활동 유형 2순위~ 등)



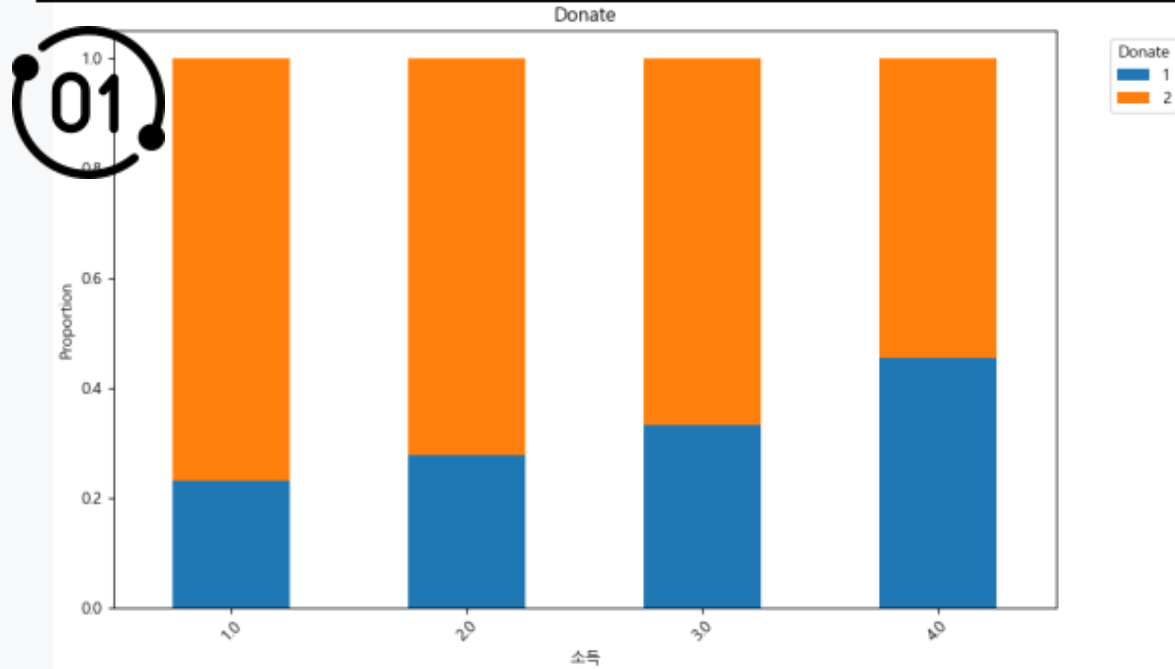
소득수준 (저소득, 중하, 중상, 고소득) 분류

'저소득' 중위소득 75% 이하, '고소득' 중위소득 200% 이상 (OECD 기준)

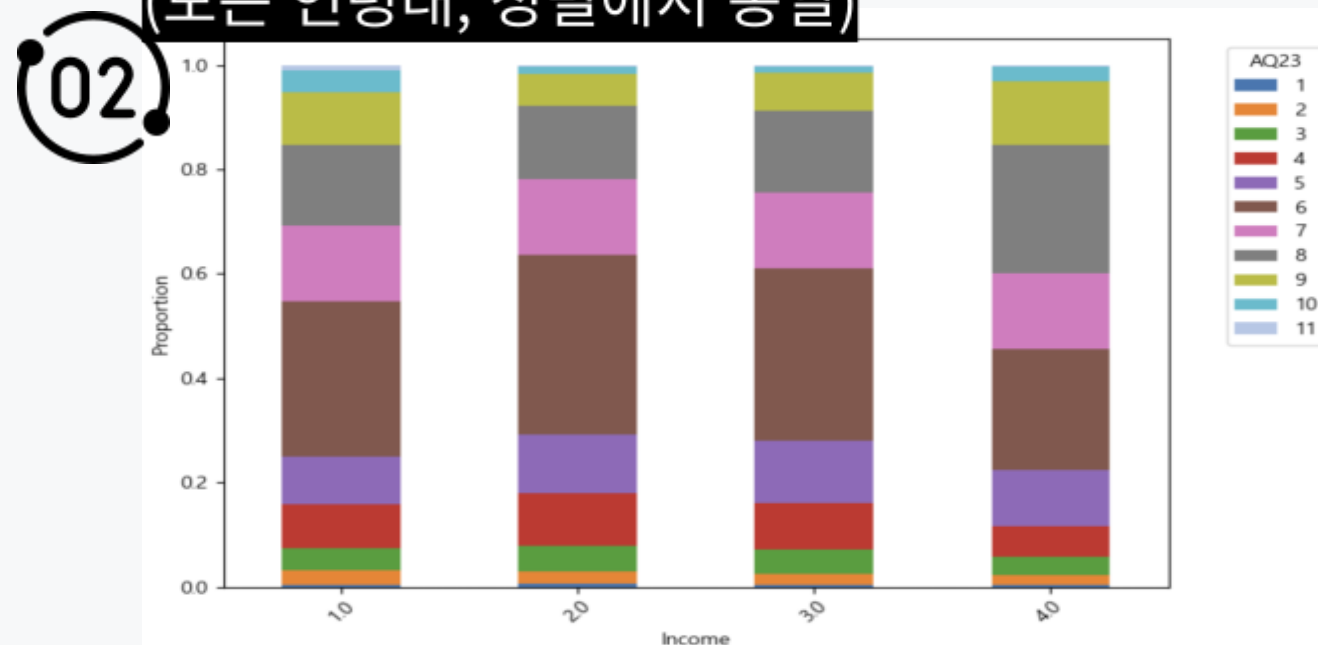
Project 2 : 시각화 - 소득간의 유의미한 차이?

VISUALIZATION

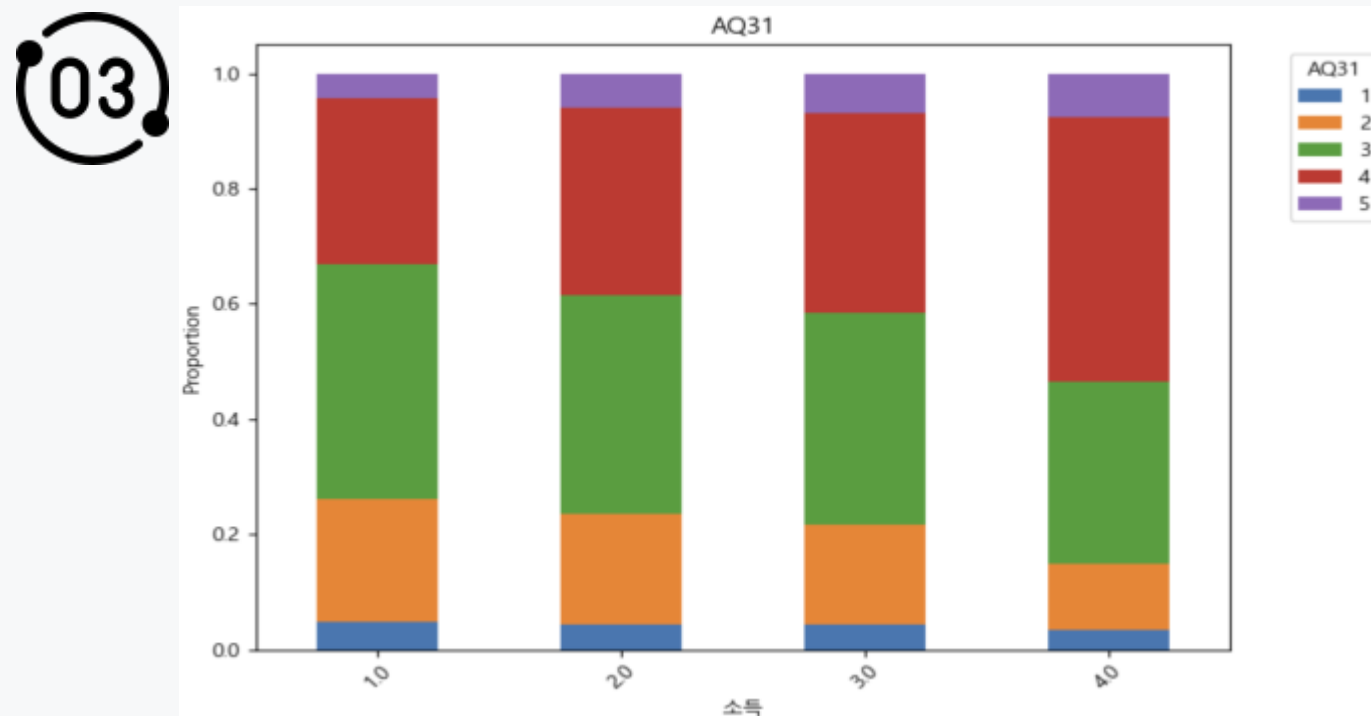
1년간 기부 여부 : 고소득일수록 기부 경험 많음



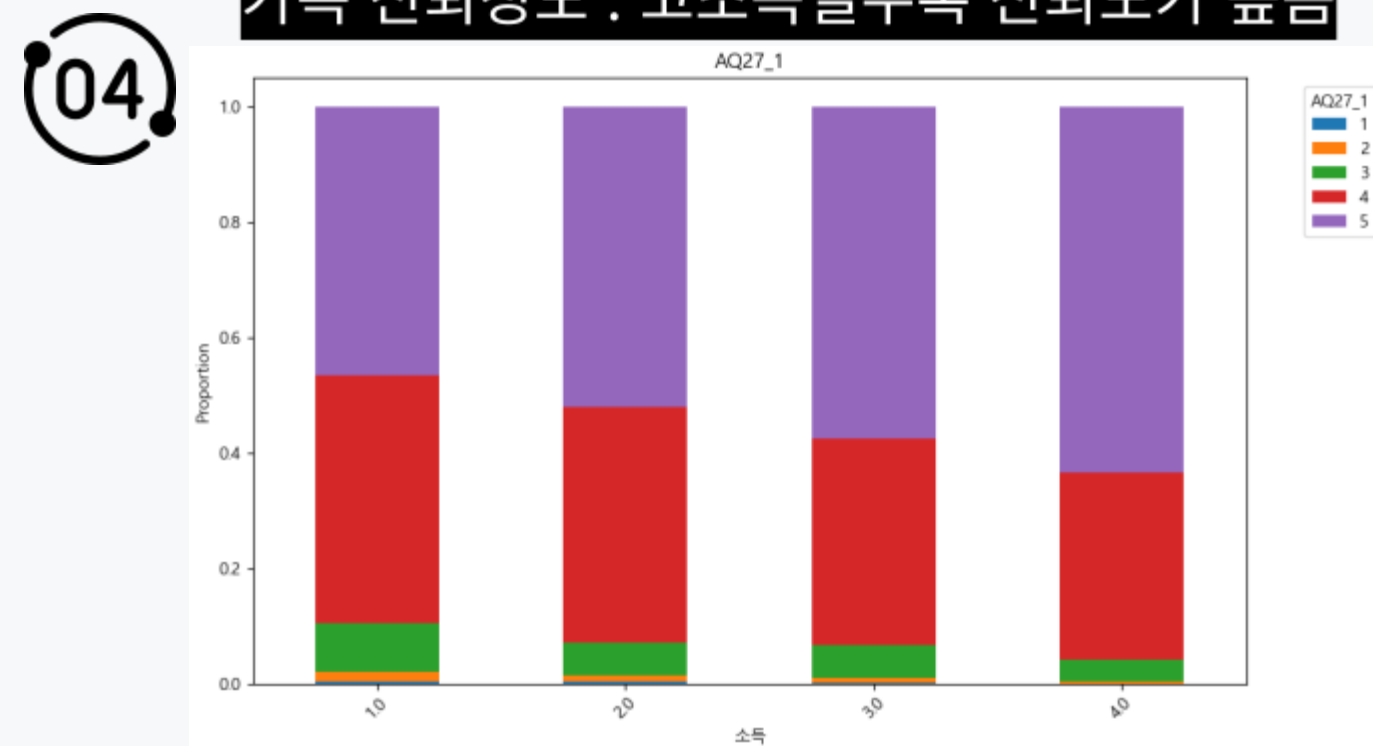
정치성향 : 고소득층(그룹4)이 조금 더 보수적인 성향
(모든 연령대, 성별에서 동일)



자녀의 계층이동 가능성 : 고소득일수록 가능성 높게 봄



가족 신뢰정도 : 고소득일수록 신뢰도가 높음



Project 2 : 모델링 준비 - ONEHOTENCODING

MODELING

1 OneHotEncoding

- 범주형 변수를 처리하는 주요 기법 중 하나로, 각 범주를 독립된 이진 특성으로 변환
- 모델이 범주 간 수치적 거리를 오해하지 않도록 하는 것이 장점.

One hot Encoding

```
1 survey_encoded_df = pd.get_dummies(survey_df.drop(columns='Income2')).astype(np.int32)
```

```
1 survey_encoded_df
```

	sq0_2_1	sq0_2_2	sq0_2_3	sq0_2_4	sq0_3_1	sq0_3_2	sq0_3_3	sq0_3_4	sq0_3_5	sq0_3_6	...	Donate_2	Sc
0	1	0	0	0	1	0	0	0	0	0	...	1	
1	1	0	0	0	1	0	0	0	0	0	...	1	
2	1	0	0	0	1	0	0	0	0	0	...	1	
3	1	0	0	0	1	0	0	0	0	0	...	1	
4	1	0	0	0	1	0	0	0	0	0	...	1	
...
39335	0	0	0	1	1	0	0	0	0	0	...	0	
39336	0	0	0	1	0	1	0	0	0	0	...	1	
39337	0	0	0	1	0	1	0	0	0	0	...	1	
39338	0	0	0	1	0	1	0	0	0	0	...	1	
39339	0	0	0	1	0	1	0	0	0	0	...	1	

38607 rows × 848 columns

Project 2 : 모델링 준비 - UNDERSAMPLING / FEATURE SELECTION

MODELING

2 UnderSampling

클래스 데이터 불균형 해소

클래스 데이터 불균형 : 클래스 간의 데이터 양의 차이가 클 때, 즉 특정 클래스는 데이터 비중이 크고 특정 클래스는 적을 때. 편향 되고 정확하지 않은 모델 결과를 낼 수 있음.

```
1 # Undersampling
2 from imblearn.under_sampling import RandomUnderSampler
3
4 # *0.15정도 규모로 해주며 4번 그룹이 전체 10%정도 되게끔 맞춰주기
5 undersampler = RandomUnderSampler(sampling_strategy={'2.0': 3900, '3.0': 1000, '1.0': 780})
6 X_train_under, y_train_under = undersampler.fit_resample(X_train, y_train)
```

```
1 survey_df['Income2'].value_counts()
✓ 0.0s
```

Income2	count
2.0	25819
3.0	6818
1.0	5163
4.0	807

Name: count, dtype: Int64

```
1 y_train_under.value_counts()
✓ 0.0s
```

Income2	count
2.0	3900
3.0	1000
1.0	780
4.0	605

Name: count, dtype: Int64

3 Feature Selection

Chi-square(SelectKBest) Feature Selection

각 피처가 타겟 변수와 독립적인지를 테스트하여 가장 관련성이 높은 K개의 피처를 선택함. 이는 모델의 예측력을 향상시키고 오버피팅을 줄이는 데 도움을 줌

```
1 # 카이제곱 feature selection
2 from sklearn.feature_selection import SelectKBest
3 from sklearn.feature_selection import chi2
4
5 selector = SelectKBest(score_func=chi2, k=80)
6 X_new = selector.fit_transform(X_train_under, y_train_under)
7 X_test_new = selector.transform(X_test)
```


Project 2 : RANDOM FOREST CLASSIFIER

MODELING

Random Forest Classifier

여러 결정 트리를 조합하여 분류 문제를 해결하는 앙상블 학습 방법
각 트리의 예측을 통합해 더 정확한 결과 도출

```
1 from sklearn.ensemble import RandomForestClassifier #Undersampled + feature selected
2
3 rfc = RandomForestClassifier(n_estimators=100, random_state=42)
4 rfc.fit(X_new, y_train_under)
5 feature_importances = rfc.feature_importances_
6 rfc.score(X_new,y_train_under), rfc.score(X_test_new,y_test)
```

✓ 1.1s

(0.9978763272954404, 0.7194363862411935)

Project 2 : GRADIENT BOOSTING CLASSIFIER & CNN

MODELING

Gradient Boosting classifier

약한 예측 모델을 순차적으로 학습시켜 오차를 줄여
나가는 앙상블 기법
이전 모델의 오류를 보완해 강력한 예측 모델 구축

```
1 from sklearn.ensemble import GradientBoostingClassifier #Undersample + feature selected
2
3 gbc = GradientBoostingClassifier(n_estimators=50, max_depth=10, random_state=42)
4 gbc.fit(X_new, y_train_under)
5 gbc.score(X_new, y_train_under), gbc.score(X_test_new, y_test)
6
7 # gbc = GradientBoostingClassifier(n_estimators=20, max_depth=10, random_state=42) : (0.86433479
8 #
```

[56] ✓ 8.2s

... (0.9948920744768496, 0.6969539991711562)

Epoch 99/100
773/773 [=====] - 2s 2ms/step - loss: 0.2793 - accuracy: 0.8879 - val_loss: 1.6045 - val_accuracy: 0.7256
Epoch 100/100
773/773 [=====] - 2s 2ms/step - loss: 0.2827 - accuracy: 0.8867 - val_loss: 1.7443 - val_accuracy: 0.7306
Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

CNN

이미지 인식과 처리에 주로 사용되는 딥러닝 구조
여러 층의 컨볼루션(필터링)과 풀링(축소) 과정을 거쳐,
특징을 추출하고 이를 분류하거나 인식

```
1 import numpy as np
2 from sklearn.model_selection import train_test_split
3 from tensorflow.keras.models import Sequential
4 from tensorflow.keras.layers import Conv1D, MaxPooling1D, Flatten, Dense
5 import tensorflow.keras as tf_keras
6
7
8 X = survey_encoded_df.values
9 y = survey_df['Income2'].values-1
10
11 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
12
13 model = Sequential([
14     Dense(128, activation='relu', input_shape=(848,)), #k
15     tf_keras.layers.Dropout(rate=0.5),
16     Dense(64, activation='relu'),
17     tf_keras.layers.Dropout(rate=0.5),
18     Dense(32, activation='relu'),
19     tf_keras.layers.Dropout(rate=0.5),
20     Dense(4, activation='softmax')
21 ])
22
23 model.compile(optimizer=tf_keras.optimizers.Adam(learning_rate=0.001),
24               loss='sparse_categorical_crossentropy',
25               metrics=['accuracy'])
26
27 history = model.fit(X_train, y_train, epochs=100, batch_size=32)
```

Epoch 1/100
773/773 [=====] - 2s 2ms/step - loss: 1.1839 - accuracy: 0.5682 - val_loss: 0.9249 - val_accuracy: 0.6662
Epoch 2/100
773/773 [=====] - 2s 2ms/step - loss: 1.0525 - accuracy: 0.6521 - val_loss: 0.9451 - val_accuracy: 0.6662
Epoch 3/100
773/773 [=====] - 2s 2ms/step - loss: 1.0012 - accuracy: 0.6625 - val_loss: 0.9001 - val_accuracy: 0.6662
Epoch 4/100
773/773 [=====] - 1s 2ms/step - loss: 0.9658 - accuracy: 0.6664 - val_loss: 0.8747 - val_accuracy: 0.6662
Epoch 5/100
773/773 [=====] - 1s 2ms/step - loss: 0.9363 - accuracy: 0.6685 - val_loss: 0.8473 - val_accuracy: 0.6667
Epoch 6/100
773/773 [=====] - 2s 2ms/step - loss: 0.9044 - accuracy: 0.6718 - val_loss: 0.8211 - val_accuracy: 0.6722
Epoch 7/100
773/773 [=====] - 1s 2ms/step - loss: 0.8780 - accuracy: 0.6757 - val_loss: 0.8029 - val_accuracy: 0.6785
Epoch 8/100
773/773 [=====] - 2s 2ms/step - loss: 0.8495 - accuracy: 0.6767 - val_loss: 0.7868 - val_accuracy: 0.6778
Epoch 9/100
773/773 [=====] - 2s 2ms/step - loss: 0.8297 - accuracy: 0.6776 - val_loss: 0.7723 - val_accuracy: 0.6824
Epoch 10/100
773/773 [=====] - 1s 2ms/step - loss: 0.8242 - accuracy: 0.6796 - val_loss: 0.7684 - val_accuracy: 0.6850
Epoch 11/100
773/773 [=====] - 1s 2ms/step - loss: 0.8088 - accuracy: 0.6848 - val_loss: 0.7510 - val_accuracy: 0.6879
Epoch 12/100
773/773 [=====] - 2s 2ms/step - loss: 0.7945 - accuracy: 0.6864 - val_loss: 0.7364 - val_accuracy: 0.6871
Epoch 13/100
773/773 [=====] - 2s 2ms/step - loss: 0.7793 - accuracy: 0.8879 - val_loss: 1.6045 - val_accuracy: 0.7256
Epoch 100/100
773/773 [=====] - 2s 2ms/step - loss: 0.2827 - accuracy: 0.8867 - val_loss: 1.7443 - val_accuracy: 0.7306
Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

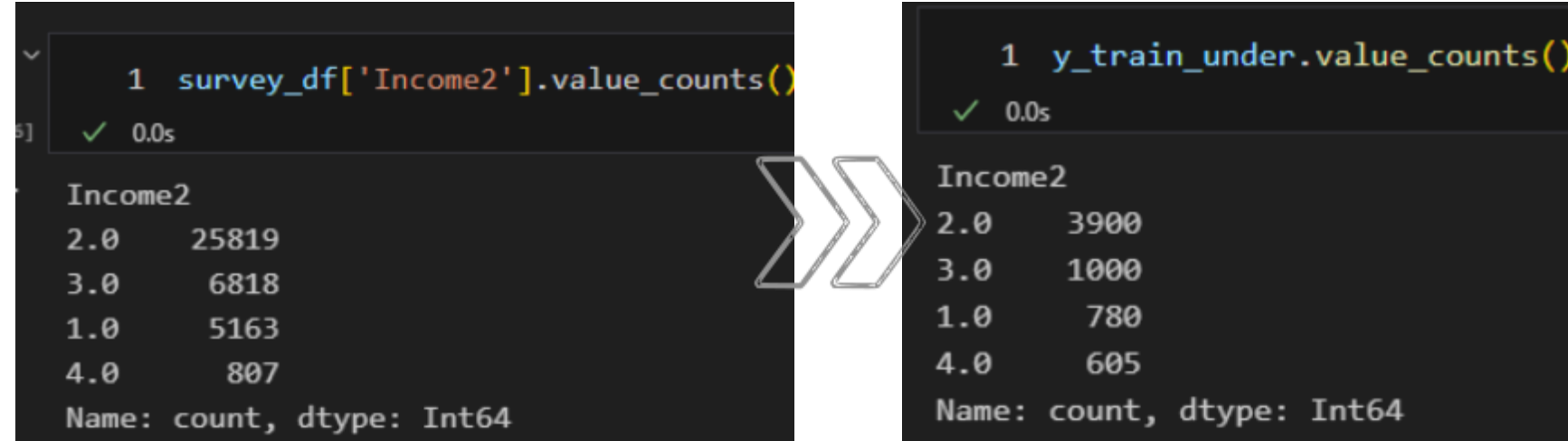
Project 2 : 한계

LIMITATIONS

01

불균형

타겟클래스의 불균형으로
왜곡이 발생 (고소득 4번
그룹이 전체 2%)



02

통일성

매칭기준이 상이하여
데이터 통합에 많은 시간 소요
예) 2021년 vs 2022년

```
2021년도 설문조사 파일과 병합 시도. But fail, pass

1 survey_2021_df = pd.read_excel('data-files/2021_survey.xlsx')

1 survey_2021_df.to_pickle('survey_2021_df')

1 # 63
2 miss_col = survey_2021_df.columns[survey_2021_df.columns.isin(survey_2022_df.columns)].tolist()
3 print(miss_col)

1 # col이 겹치는 것
2 b = ['ID', 'FAM1', 'FAM15', 'SQ1_1', 'SQ1_2', 'SQ1_3', 'SQ1_4', 'SQ1_5', 'SQ1_6', 'SQ1_7', 'AQ1_1', \
3      'AQ1_2', 'AQ1_3', 'AQ1_4', 'AQ2_1', 'AQ2_2', 'AQ2_3', 'AQ2_4', 'AQ2_5', 'AQ4', 'AQ5_1', 'AQ5_2', 'AQ5_3', 'AQ5_4', 'AQ6A_1', \
4      'AQ6A_2', 'AQ6A_3', 'AQ6B_1', 'AQ6B_2', 'AQ6B_3', 'AQ7_1', 'AQ7_2', 'AQ7_3', 'AQ7_4', 'AQ7_5', 'AQ7_6', 'AQ7_7', 'AQ7_8', \
5      'AQ7_9', 'AQ7_10', 'AQ8_1', 'AQ8_2', 'AQ8_3', 'AQ8_4', 'AQ9_1', 'AQ9_2', 'AQ9_3', 'AQ10', 'AQ10A', 'AQ10B', \
6      'AQ10C_1', 'AQ10C_2', 'AQ10D', 'AQ13A_1', 'AQ13A_2', 'AQ13B_1', 'AQ13B_2', 'AQ15', 'AQ16', 'AQ24', 'AQ25', 'AQ01', \
7      'AQ01A', 'AQ02_1_1', 'AQ02_1_2', 'AQ02_2_1', 'AQ02_2_2', 'AQ03', 'AQ03A', 'AQ03AA_1', 'AQ03AA_2', 'AQ03B_1', 'AQ03B_2', \
8      'AQ03C_1', 'AQ03C_2', 'AQ03D', 'AQ03E', 'DEW1', 'DEW2', 'DEW3', 'DEW4', 'DEW5', 'DEW6', 'DEW7', 'DEW8', 'DEW9', 'DEW10', 'wtb1', 'wtb2']
9 print(b)
10 len(b)

1 # 2022에 없는 것
2 a = ['GU', 'SQ0_2', 'SQ0_3', 'SQ1_8A', 'AQ3', 'AQ11_1', 'AQ11_2', 'AQ11_3', 'AQ12A_1', 'AQ12A_2', 'AQ12A_3', 'AQ12A_4', 'AQ12A_5', \
3      'AQ12A_6', 'AQ12A_7', 'AQ12A_8', 'AQ12AA_1', 'AQ12AA_2', 'AQ12AA_3', 'AQ12AA_4', 'AQ12AA_5', \
4      'AQ12AA_6', 'AQ12AA_7', 'AQ12AA_8', 'AQ12B_1', 'AQ12B_2', 'AQ12B_3', 'AQ12B_4', 'AQ12B_5', 'AQ12B_6', 'AQ12B_7', 'AQ12B_8', \
5      'AQ13C_1', 'AQ13C_2', 'AQ13D_1', 'AQ13D_2', 'AQ14', 'AQ16A_1', 'AQ16A_2', 'AQ17_1', 'AQ17_2', 'AQ18', 'AQ18A_1', 'AQ18A_2', \
6      'AQ18A_3', 'AQ18A_4', 'AQ18A_5', 'AQ18A_6', 'AQ18A_7', 'AQ19_1', 'AQ19_2', 'AQ19_3', 'AQ19_4', 'AQ19_5', 'AQ19_6', 'AQ19_7', \
7      'AQ20', 'AQ20A_1', 'AQ20A_2', 'AQ20A_3', 'AQ20A_4', 'AQ20A_5', 'AQ20A_6', 'AQ20A_7', 'AQ21', 'AQ22', 'AQ22A', 'AQ22A_1', \
8      'AQ23A_1', 'AQ23A_2', 'AQ23A_3', 'AQ23A_4', 'AQ23A_5', 'AQ23A_6', 'AQ23A_7', 'AQ23A_8', 'AQ23A_9', 'AQ23A_10', 'AQ23A_11', \
9      'AQ23A_12', 'AQ23A_13', 'AQ23A_14', 'AQ23A_15', 'AQ23A_16', 'AQ23A_17', 'AQ23A_18', 'AQ23A_19', 'AQ23A_20', 'AQ23A_21', \
10     'AQ23A_22', 'AQ23A_23', 'AQ23A_24', 'AQ23A_25', 'AQ23A_26', 'AQ23A_27', 'AQ23A_28', 'AQ23A_29', 'AQ23A_30', 'AQ23A_31', \
11     'AQ23A_32', 'AQ23A_33', 'AQ23A_34', 'AQ23A_35', 'AQ23A_36', 'AQ23A_37', 'AQ23A_38', 'AQ23A_39', 'AQ23A_40', 'AQ23A_41', \
12     'AQ23A_42', 'AQ23A_43', 'AQ23A_44', 'AQ23A_45', 'AQ23A_46', 'AQ23A_47', 'AQ23A_48', 'AQ23A_49', 'AQ23A_50', 'AQ23A_51', \
13     'AQ23A_52', 'AQ23A_53', 'AQ23A_54', 'AQ23A_55', 'AQ23A_56', 'AQ23A_57', 'AQ23A_58', 'AQ23A_59', 'AQ23A_60', 'AQ23A_61', \
14     'AQ23A_62', 'AQ23A_63', 'AQ23A_64', 'AQ23A_65', 'AQ23A_66', 'AQ23A_67', 'AQ23A_68', 'AQ23A_69', 'AQ23A_70', 'AQ23A_71', \
15     'AQ23A_72', 'AQ23A_73', 'AQ23A_74', 'AQ23A_75', 'AQ23A_76', 'AQ23A_77', 'AQ23A_78', 'AQ23A_79', 'AQ23A_80', 'AQ23A_81', \
16     'AQ23A_82', 'AQ23A_83', 'AQ23A_84', 'AQ23A_85', 'AQ23A_86', 'AQ23A_87', 'AQ23A_88', 'AQ23A_89', 'AQ23A_90', 'AQ23A_91', \
17     'AQ23A_92', 'AQ23A_93', 'AQ23A_94', 'AQ23A_95', 'AQ23A_96', 'AQ23A_97', 'AQ23A_98', 'AQ23A_99', 'AQ23A_100', 'AQ23A_101', \
18     'AQ23A_102', 'AQ23A_103', 'AQ23A_104', 'AQ23A_105', 'AQ23A_106', 'AQ23A_107', 'AQ23A_108', 'AQ23A_109', 'AQ23A_110', \
19     'AQ23A_111', 'AQ23A_112', 'AQ23A_113', 'AQ23A_114', 'AQ23A_115', 'AQ23A_116', 'AQ23A_117', 'AQ23A_118', 'AQ23A_119', \
20     'AQ23A_120', 'AQ23A_121', 'AQ23A_122', 'AQ23A_123', 'AQ23A_124', 'AQ23A_125', 'AQ23A_126', 'AQ23A_127', 'AQ23A_128', \
21     'AQ23A_129', 'AQ23A_130', 'AQ23A_131', 'AQ23A_132', 'AQ23A_133', 'AQ23A_134', 'AQ23A_135', 'AQ23A_136', 'AQ23A_137', \
22     'AQ23A_138', 'AQ23A_139', 'AQ23A_140', 'AQ23A_141', 'AQ23A_142', 'AQ23A_143', 'AQ23A_144', 'AQ23A_145', 'AQ23A_146', \
23     'AQ23A_147', 'AQ23A_148', 'AQ23A_149', 'AQ23A_150', 'AQ23A_151', 'AQ23A_152', 'AQ23A_153', 'AQ23A_154', 'AQ23A_155', \
24     'AQ23A_156', 'AQ23A_157', 'AQ23A_158', 'AQ23A_159', 'AQ23A_160', 'AQ23A_161', 'AQ23A_162', 'AQ23A_163', 'AQ23A_164', \
25     'AQ23A_165', 'AQ23A_166', 'AQ23A_167', 'AQ23A_168', 'AQ23A_169', 'AQ23A_170', 'AQ23A_171', 'AQ23A_172', 'AQ23A_173', \
26     'AQ23A_174', 'AQ23A_175', 'AQ23A_176', 'AQ23A_177', 'AQ23A_178', 'AQ23A_179', 'AQ23A_180', 'AQ23A_181', 'AQ23A_182', \
27     'AQ23A_183', 'AQ23A_184', 'AQ23A_185', 'AQ23A_186', 'AQ23A_187', 'AQ23A_188', 'AQ23A_189', 'AQ23A_190', 'AQ23A_191', \
28     'AQ23A_192', 'AQ23A_193', 'AQ23A_194', 'AQ23A_195', 'AQ23A_196', 'AQ23A_197', 'AQ23A_198', 'AQ23A_199', 'AQ23A_200', \
29     'AQ23A_201', 'AQ23A_202', 'AQ23A_203', 'AQ23A_204', 'AQ23A_205', 'AQ23A_206', 'AQ23A_207', 'AQ23A_208', 'AQ23A_209', \
30     'AQ23A_210', 'AQ23A_211', 'AQ23A_212', 'AQ23A_213', 'AQ23A_214', 'AQ23A_215', 'AQ23A_216', 'AQ23A_217', 'AQ23A_218', \
31     'AQ23A_219', 'AQ23A_220', 'AQ23A_221', 'AQ23A_222', 'AQ23A_223', 'AQ23A_224', 'AQ23A_225', 'AQ23A_226', 'AQ23A_227', \
32     'AQ23A_228', 'AQ23A_229', 'AQ23A_230', 'AQ23A_231', 'AQ23A_232', 'AQ23A_233', 'AQ23A_234', 'AQ23A_235', 'AQ23A_236', \
33     'AQ23A_237', 'AQ23A_238', 'AQ23A_239', 'AQ23A_240', 'AQ23A_241', 'AQ23A_242', 'AQ23A_243', 'AQ23A_244', 'AQ23A_245', \
34     'AQ23A_246', 'AQ23A_247', 'AQ23A_248', 'AQ23A_249', 'AQ23A_250', 'AQ23A_251', 'AQ23A_252', 'AQ23A_253', 'AQ23A_254', \
35     'AQ23A_255', 'AQ23A_256', 'AQ23A_257', 'AQ23A_258', 'AQ23A_259', 'AQ23A_260', 'AQ23A_261', 'AQ23A_262', 'AQ23A_263', \
36     'AQ23A_264', 'AQ23A_265', 'AQ23A_266', 'AQ23A_267', 'AQ23A_268', 'AQ23A_269', 'AQ23A_270', 'AQ23A_271', 'AQ23A_272', \
37     'AQ23A_273', 'AQ23A_274', 'AQ23A_275', 'AQ23A_276', 'AQ23A_277', 'AQ23A_278', 'AQ23A_279', 'AQ23A_280', 'AQ23A_281', \
38     'AQ23A_282', 'AQ23A_283', 'AQ23A_284', 'AQ23A_285', 'AQ23A_286', 'AQ23A_287', 'AQ23A_288', 'AQ23A_289', 'AQ23A_290', \
39     'AQ23A_291', 'AQ23A_292', 'AQ23A_293', 'AQ23A_294', 'AQ23A_295', 'AQ23A_296', 'AQ23A_297', 'AQ23A_298', 'AQ23A_299', \
40     'AQ23A_300', 'AQ23A_301', 'AQ23A_302', 'AQ23A_303', 'AQ23A_304', 'AQ23A_305', 'AQ23A_306', 'AQ23A_307', 'AQ23A_308', \
41     'AQ23A_309', 'AQ23A_310', 'AQ23A_311', 'AQ23A_312', 'AQ23A_313', 'AQ23A_314', 'AQ23A_315', 'AQ23A_316', 'AQ23A_317', \
42     'AQ23A_318', 'AQ23A_319', 'AQ23A_320', 'AQ23A_321', 'AQ23A_322', 'AQ23A_323', 'AQ23A_324', 'AQ23A_325', 'AQ23A_326', \
43     'AQ23A_327', 'AQ23A_328', 'AQ23A_329', 'AQ23A_330', 'AQ23A_331', 'AQ23A_332', 'AQ23A_333', 'AQ23A_334', 'AQ23A_335', \
44     'AQ23A_336', 'AQ23A_337', 'AQ23A_338', 'AQ23A_339', 'AQ23A_340', 'AQ23A_341', 'AQ23A_342', 'AQ23A_343', 'AQ23A_344', \
45     'AQ23A_345', 'AQ23A_346', 'AQ23A_347', 'AQ23A_348', 'AQ23A_349', 'AQ23A_350', 'AQ23A_351', 'AQ23A_352', 'AQ23A_353', \
46     'AQ23A_354', 'AQ23A_355', 'AQ23A_356', 'AQ23A_357', 'AQ23A_358', 'AQ23A_359', 'AQ23A_360', 'AQ23A_361', 'AQ23A_362', \
47     'AQ23A_363', 'AQ23A_364', 'AQ23A_365', 'AQ23A_366', 'AQ23A_367', 'AQ23A_368', 'AQ23A_369', 'AQ23A_370', 'AQ23A_371', \
48     'AQ23A_372', 'AQ23A_373', 'AQ23A_374', 'AQ23A_375', 'AQ23A_376', 'AQ23A_377', 'AQ23A_378', 'AQ23A_379', 'AQ23A_380', \
49     'AQ23A_381', 'AQ23A_382', 'AQ23A_383', 'AQ23A_384', 'AQ23A_385', 'AQ23A_386', 'AQ23A_387', 'AQ23A_388', 'AQ23A_389', \
50     'AQ23A_390', 'AQ23A_391', 'AQ23A_392', 'AQ23A_393', 'AQ23A_394', 'AQ23A_395', 'AQ23A_396', 'AQ23A_397', 'AQ23A_398', \
51     'AQ23A_399', 'AQ23A_400', 'AQ23A_401', 'AQ23A_402', 'AQ23A_403', 'AQ23A_404', 'AQ23A_405', 'AQ23A_406', 'AQ23A_407', \
52     'AQ23A_408', 'AQ23A_409', 'AQ23A_410', 'AQ23A_411', 'AQ23A_412', 'AQ23A_413', 'AQ23A_414', 'AQ23A_415', 'AQ23A_416', \
53     'AQ23A_417', 'AQ23A_418', 'AQ23A_419', 'AQ23A_420', 'AQ23A_421', 'AQ23A_422', 'AQ23A_423', 'AQ23A_424', 'AQ23A_425', \
54     'AQ23A_426', 'AQ23A_427', 'AQ23A_428', 'AQ23A_429', 'AQ23A_430', 'AQ23A_431', 'AQ23A_432', 'AQ23A_433', 'AQ23A_434', \
55     'AQ23A_435', 'AQ23A_436', 'AQ23A_437', 'AQ23A_438', 'AQ23A_439', 'AQ23A_440', 'AQ23A_441', 'AQ23A_442', 'AQ23A_443', \
56     'AQ23A_444', 'AQ23A_445', 'AQ23A_446', 'AQ23A_447', 'AQ23A_448', 'AQ23A_449', 'AQ23A_450', 'AQ23A_451', 'AQ23A_452', \
57     'AQ23A_453', 'AQ23A_454', 'AQ23A_455', 'AQ23A_456', 'AQ23A_457', 'AQ23A_458', 'AQ23A_459', 'AQ23A_460', 'AQ23A_461', \
58     'AQ23A_462', 'AQ23A_463', 'AQ23A_464', 'AQ23A_465', 'AQ23A_466', 'AQ23A_467', 'AQ23A_468', 'AQ23A_469', 'AQ23A_470', \
59     'AQ23A_471', 'AQ23A_472', 'AQ23A_473', 'AQ23A_474', 'AQ23A_475', 'AQ23A_476', 'AQ23A_477', 'AQ23A_478', 'AQ23A_479', \
60     'AQ23A_480', 'AQ23A_481', 'AQ23A_482', 'AQ23A_483', 'AQ23A_484', 'AQ23A_485', 'AQ23A_486', 'AQ23A_487', 'AQ23A_488', \
61     'AQ23A_489', 'AQ23A_490', 'AQ23A_491', 'AQ23A_492', 'AQ23A_493', 'AQ23A_494', 'AQ23A_495', 'AQ23A_496', 'AQ23A_497', \
62     'AQ23A_498', 'AQ23A_499', 'AQ23A_500', 'AQ23A_501', 'AQ23A_502', 'AQ23A_503', 'AQ23A_504', 'AQ23A_505', 'AQ23A_506', \
63     'AQ23A_507', 'AQ23A_508', 'AQ23A_509', 'AQ23A_510', 'AQ23A_511', 'AQ23A_512', 'AQ23A_513', 'AQ23A_514', 'AQ23A_515', \
64     'AQ23A_516', 'AQ23A_517', 'AQ23A_518', 'AQ23A_519', 'AQ23A_520', 'AQ23A_521', 'AQ23A_522', 'AQ23A_523', 'AQ23A_524', \
65     'AQ23A_525', 'AQ23A_526', 'AQ23A_527', 'AQ23A_528', 'AQ23A_529', 'AQ23A_530', 'AQ23A_531', 'AQ23A_532', 'AQ23A_533', \
66     'AQ23A_534', 'AQ23A_535', 'AQ23A_536', 'AQ23A_537', 'AQ23A_538', 'AQ23A_539', 'AQ23A_540', 'AQ23A_541', 'AQ23A_542', \
67     'AQ23A_543', 'AQ23A_544', 'AQ23A_545', 'AQ23A_546', 'AQ23A_547', 'AQ23A_548', 'AQ23A_549', 'AQ23A_550', 'AQ23A_551', \
68     'AQ23A_552', 'AQ23A_553', 'AQ23A_554', 'AQ23A_555', 'AQ23A_556', 'AQ23A_557', 'AQ23A_558', 'AQ23A_559', 'AQ23A_560', \
69     'AQ23A_561', 'AQ23A_562', 'AQ23A_563', 'AQ23A_564', 'AQ23A_565', 'AQ23A_566', 'AQ23A_567', 'AQ23A_568', 'AQ23A_569', \
70     'AQ23A_570', 'AQ23A_571', 'AQ23A_572', 'AQ23A_573', 'AQ23A_574', 'AQ23A_575', 'AQ23A_576', 'AQ23A_577', 'AQ23A_578', \
71     'AQ23A_579', 'AQ23A_580', 'AQ23A_581', 'AQ23A_582', 'AQ23A_583', 'AQ23A_584', 'AQ23A_585', 'AQ23A_586', 'AQ23A_587', \
72     'AQ23A_588', 'AQ23A_589', 'AQ23A_590', 'AQ23A_591', 'AQ23A_592', 'AQ23A_593', 'AQ23A_594', 'AQ23A_595', 'AQ23A_596', \
73     'AQ23A_597', 'AQ23A_598', 'AQ23A_599', 'AQ23A_600', 'AQ23A_601', 'AQ23A_602', 'AQ23A_603', 'AQ23A_604', 'AQ23A_605', \
74     'AQ23A_606', 'AQ23A_607', 'AQ23A_608', 'AQ23A_609', 'AQ23A_610', 'AQ23A_611', 'AQ23A_612', 'AQ23A_613', 'AQ23A_614', \
75     'AQ23A_615', 'AQ23A_616', 'AQ23A_617', 'AQ23A_618', 'AQ23A_619', 'AQ23A_620', 'AQ23A_621', 'AQ23A_622', 'AQ23A_623', \
76     'AQ23A_624', 'AQ23A_625', 'AQ23A_626', 'AQ23A_627', 'AQ23A_628', 'AQ23A_629', 'AQ23A_630', 'AQ23A_631', 'AQ23A_632', \
77     'AQ23A_633', 'AQ23A_634', 'AQ23A_635', 'AQ23A_636', 'AQ23A_637', 'AQ23A_638', 'AQ23A_639', 'AQ23A_640', 'AQ23A_641', \
78     'AQ23A_642', 'AQ23A_643', 'AQ23A_644', 'AQ23A_645', 'AQ23A_646', 'AQ23A_647', 'AQ23A_648', 'AQ23A_649', 'AQ23A_650', \
79     'AQ23A_651', 'AQ23A_652', 'AQ23A_653', 'AQ23A_654', 'AQ23A_655', 'AQ23A_656', 'AQ23A_657', 'AQ23A_658', 'AQ23A_659', \
80     'AQ23A_660', 'AQ23A_661', 'AQ23A_662', 'AQ23A_663', 'AQ23A_664', 'AQ23A_665', 'AQ23A_666', 'AQ23A_667', 'AQ23A_668', \
81     'AQ23A_669', 'AQ23A_670', 'AQ23A_671', 'AQ23A_672', 'AQ23A_673', 'AQ23A_674', 'AQ23A_675', 'AQ23A_676', 'AQ23A_677', \
82     'AQ23A_678', 'AQ23A_679', 'AQ23A_680', 'AQ23A_681', 'AQ23A_682', 'AQ23A_683', 'AQ23A_684', 'AQ23A_685', 'AQ23A_686', \
83     'AQ23A_687', 'AQ23A_688', 'AQ23A_689', 'AQ23A_690', 'AQ23A_691', 'AQ23A_692', 'AQ23A_693', 'AQ23A_694', 'AQ23A_695', \
84     'AQ23A_696', 'AQ23A_697', 'AQ23A_698', 'AQ23A_699', 'AQ23A_700', 'AQ23A_701', 'AQ23A_702', 'AQ23A_703', 'AQ23A_704', \
85     'AQ23A_705', 'AQ23A_706', 'AQ23A_707', 'AQ23A_708', 'AQ23A_709', 'AQ23A_710', 'AQ23A_711', 'AQ23A_712', 'AQ23A_713', \
86     'AQ23A_714', 'AQ23A_715', 'AQ23A_716', 'AQ23A_717', 'AQ23A_718', 'AQ23A_719', 'AQ23A_720', 'AQ23A_721', 'AQ23A_722', \
87     'AQ23A_723', 'AQ23A_724', 'AQ23A_725', 'AQ23A_726', 'AQ23A_727', 'AQ23A_728', 'AQ23A_729', 'AQ23A_730', 'AQ23A_731', \
88     'AQ23A_732', 'AQ23A_733', 'AQ23A_734', 'AQ23A_735', 'AQ23A_736', 'AQ23A_737', 'AQ23A_738', 'AQ23A_739', 'AQ23A_740', \
89     'AQ23A_741', 'AQ23A_742', 'AQ23A_743', 'AQ23A_744', 'AQ23A_745', 'AQ23A_746', 'AQ23A_747', 'AQ23A_748', 'AQ23A_749', \
90     'AQ23A_750', 'AQ23A_751', 'AQ23A_752', 'AQ23A_753', 'AQ23A_754', 'AQ23A_755', 'AQ23A_756', 'AQ23A_757', 'AQ23A_758', \
91     'AQ23A_759', 'AQ23A_760', 'AQ23A_761', 'AQ23A_762', 'AQ23A_763', 'AQ23A_764', 'AQ23A_765', 'AQ23A_766', 'AQ23A_767', \
92     'AQ23A_768', 'AQ23A_769', 'AQ23A_770', 'AQ23A_771', 'AQ23A_772', 'AQ23A_773', 'AQ23A_774', 'AQ23A_775', 'AQ23A_776', \
93     'AQ23A_777', 'AQ23A_778', 'AQ23A_779', 'AQ23A_780', 'AQ23A_781', 'AQ23A_782', 'AQ23A_783', 'AQ23A_784', 'AQ23A_785', \
94     'AQ23A_786', 'AQ23A_787', 'AQ23A_788', 'AQ23A_789', 'AQ23A_790', 'AQ23A_791', 'AQ23A_792', 'AQ23A_793', 'AQ23A_794', \
95     'AQ23A_795', 'AQ23A_796', 'AQ23A_797', 'AQ23A_798', 'AQ23A_799', 'AQ23A_800', 'AQ23A_801', 'AQ23A_802', 'AQ23A_803', \
96     'AQ23A_804', 'AQ23A_805', 'AQ23A_806', 'AQ23A_807', 'AQ23A_808', 'AQ23A_809', 'AQ23A_810', 'AQ23A_811', 'AQ23A_812', \
97     'AQ23A_813', 'AQ23A_814', 'AQ23A_815', 'AQ23A_816', 'AQ23A_817', 'AQ23A_818', 'AQ23A_819', 'AQ23A_820', 'AQ23A_821', \
98     'AQ23A_822', 'AQ23A_823', 'AQ23A_824', 'AQ23A_825', 'AQ23A_826', 'AQ23A_827', 'AQ23A_828', 'AQ23A_829', 'AQ23A_830', \
99     'AQ23A_831', 'AQ23A_832', 'AQ23A_833', 'AQ23A_834', 'AQ23A_835', 'AQ23A_836', 'AQ23A_837', 'AQ23A_838', 'AQ23A_839', \
100    'AQ23A_840', 'AQ23A_841', 'AQ23A_842', 'AQ23A_843', 'AQ23A_844', 'AQ23A_845', 'AQ23A_846', 'AQ23A_847', 'AQ23A_848', \
101    'AQ23A_849', 'AQ23A_850', 'AQ23A_851', 'AQ23A_852', 'AQ23A_853', 'AQ23A_854', 'AQ23A_855', 'AQ23A_856', 'AQ23A_857', \
102    'AQ23A_858', 'AQ23A_859', 'AQ23A_860', 'AQ23A_861', 'AQ23A_862', 'AQ23A_863', 'AQ23A_864', 'AQ23A_865', 'AQ23A_866', \
103    'AQ23A_867', 'AQ23A_868', 'AQ23A_869', 'AQ23A_870', 'AQ23A_871', 'AQ23A_872', 'AQ23A_873', 'AQ23A_874', 'AQ23A_875', \
104    'AQ23A_876', 'AQ23A_877', 'AQ23A_878', 'AQ23A_879', 'AQ23A_880', 'AQ23A_881', 'AQ23A_882', 'AQ23A_883', 'AQ23A_884', \
105    'AQ23A_885', 'AQ23A_886', 'AQ23A_887', 'AQ23A_888', 'AQ23A_889', 'AQ23A_890', 'AQ23A_891', 'AQ23A_892', 'AQ23A_893', \
106    'AQ23A_894', 'AQ23A_895', 'AQ23A_896', 'AQ23A_897', 'AQ23A_898', 'AQ23A_899', 'AQ23A_900', 'AQ23A_901', 'AQ23A_902', \
107    'AQ23A_903', 'AQ23A_904', 'AQ23A_905', 'AQ23A_906', 'AQ23A_907', 'AQ23A_908', 'AQ23A_909', 'AQ23A_910', 'AQ23A_911', \
108    'AQ23A_912', 'AQ23A_913', 'AQ23A_914', 'AQ23A_915', 'AQ23A_916', 'AQ23A_917', 'AQ23A_918', 'AQ23A_919', 'AQ23A_920', \
109    'AQ23A_921', 'AQ23A_922', 'AQ23A_923', 'AQ23A_924', 'AQ23A_925', 'AQ23A_926', 'AQ23A_927', 'AQ23A_928', 'AQ23A_929', \
110    'AQ23A_930', 'AQ23A_931', 'AQ23A_932', 'AQ23A_933', 'AQ23A_934', 'AQ23A_935', 'AQ23A_936', 'AQ23A_937', 'AQ23A_938', \
111    'AQ23A_939', 'AQ23A_940', 'AQ23A_941', 'AQ23A_942', 'AQ23A_943', 'AQ23A_944', 'AQ23A_945', 'AQ23A_946', 'AQ23A_947', \
112    'AQ23A_948', 'AQ23A_949', 'AQ23A_950', 'AQ23A_951', 'AQ23A_952', 'AQ23A_953', 'AQ23A_954', 'AQ23A_955', 'AQ23A_956', \
113    'AQ23A_957', 'AQ23A_958', 'AQ23A_959', 'AQ23A_960', 'AQ23A_961', 'AQ23A_962', 'AQ23A_963', 'AQ23A_964', 'AQ23A_965', \
114    'AQ23A_966', 'AQ23A_967', 'AQ23A_968', 'AQ23A_969', 'AQ23A_970', 'AQ23A_971', 'AQ23A_972', 'AQ23A_973', 'AQ23A_974', \
115    'AQ23A_975', 'AQ23A_976', 'AQ23A_977', 'AQ23A_978', 'AQ23A_979', 'AQ23A_980', 'AQ23A_981', 'AQ23A_982', 'AQ23A_983', \
116    'AQ2
```

CONCLUSION 3

03. 활용방안

CONCLUSION



프로젝트 1

개인

일반사용자에게는 선호하는 조건에 따라 400여개의 행정구역중 적절한 지역을 추천하는데 활용

부동산

업체와 중개인은 이 인사이트를 활용하여, 타겟에 맞는 주거지역을 추천하고 마케팅 전략 수립

정부

정책적인 부분에서는 지역적인 특성을 이해하고 상대적으로 점수가 낮은 분야를 개발하거나 차별화 하여 도시계획의 기초 데이터로 활용할 수 있음.



프로젝트 2

추후 설문조사에서 설문 응답자들의 소득 정보가 누락 되어도 나머지 항목들로 어느 정도 소득 유추 가능

특정 회사가 고객의 민감한 급여정보를 수집하지 않아도 나머지 속성 10~20개로 소득 수준을 어느 정도 예측해 영업 및 타겟팅에 활용 가능

03. 기술사양

CONCLUSION

01 적용 기술

개발언어 / 데이터 전처리

python

탐색적 분석

데이터 시각화, 통계 분석

예측 모델링

머신러닝(RFC, GBC), 딥러닝 (CNN)

02 적용기술 및 도구

코드개발

Visual Studio Code

형상관리

github

탐색적분석

numpy, pandas , matplotlib, seaborn 등

머신러닝

scikit-learn, xgboost, tensorflow

[illegible]

03. 문제점 및 개선점

CONCLUSION

**Good quality data leads to
good results.**

좋은 품질의 데이터는 좋은 결과로 이어진다

Q & A