

KHU'DATA 3조

# NEET족 구제하기

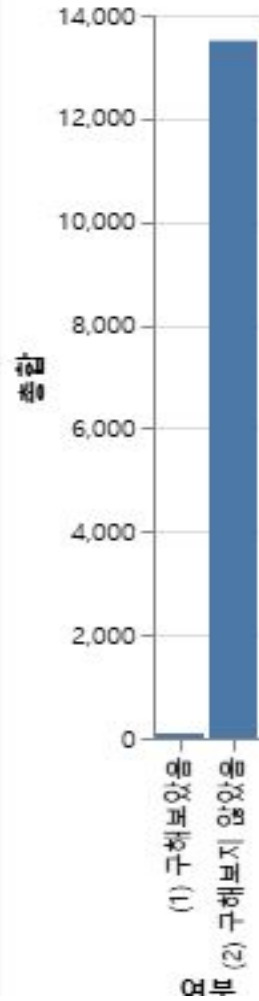
조민우, 성민지, 정시찬, 김현정, 유민아, 박태호

# Index

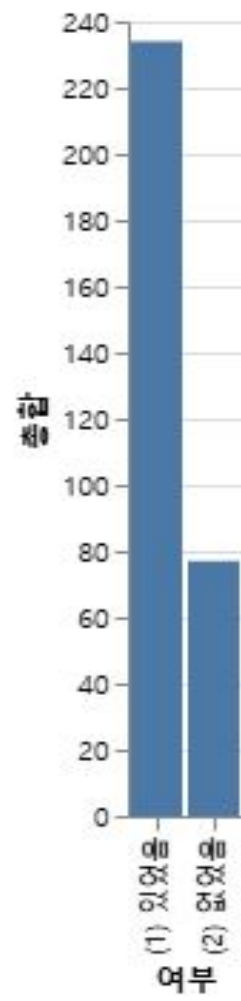
1. 프로젝트 개요
2. 데이터 전처리
3. 모델링
4. 결과 분석 및 방향성 제시

## 주제 선정 이유

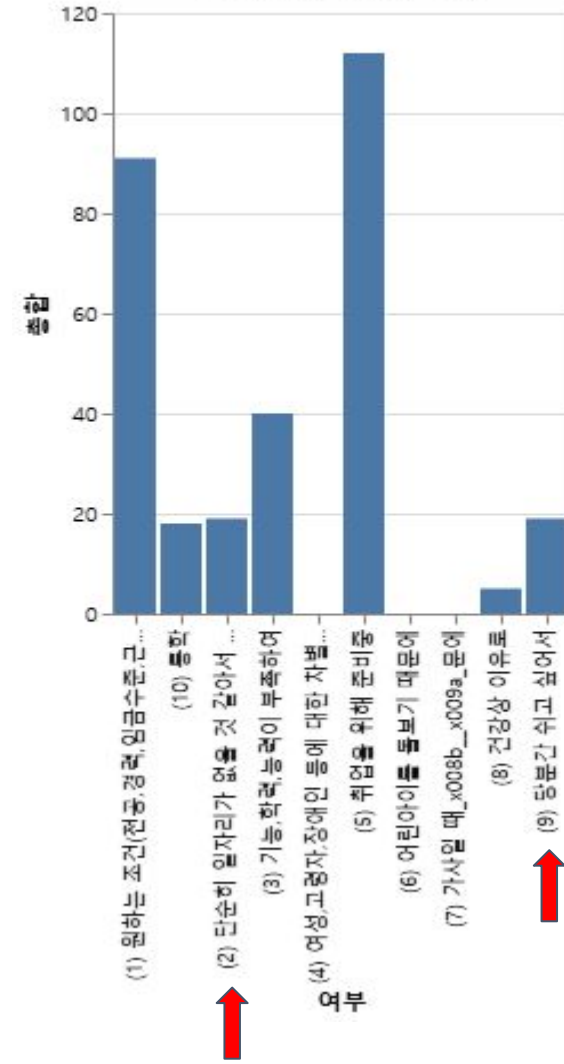
지난 1개월 간 구직 여부



일자리제공시 근로여부



구직활동 안하는 이유

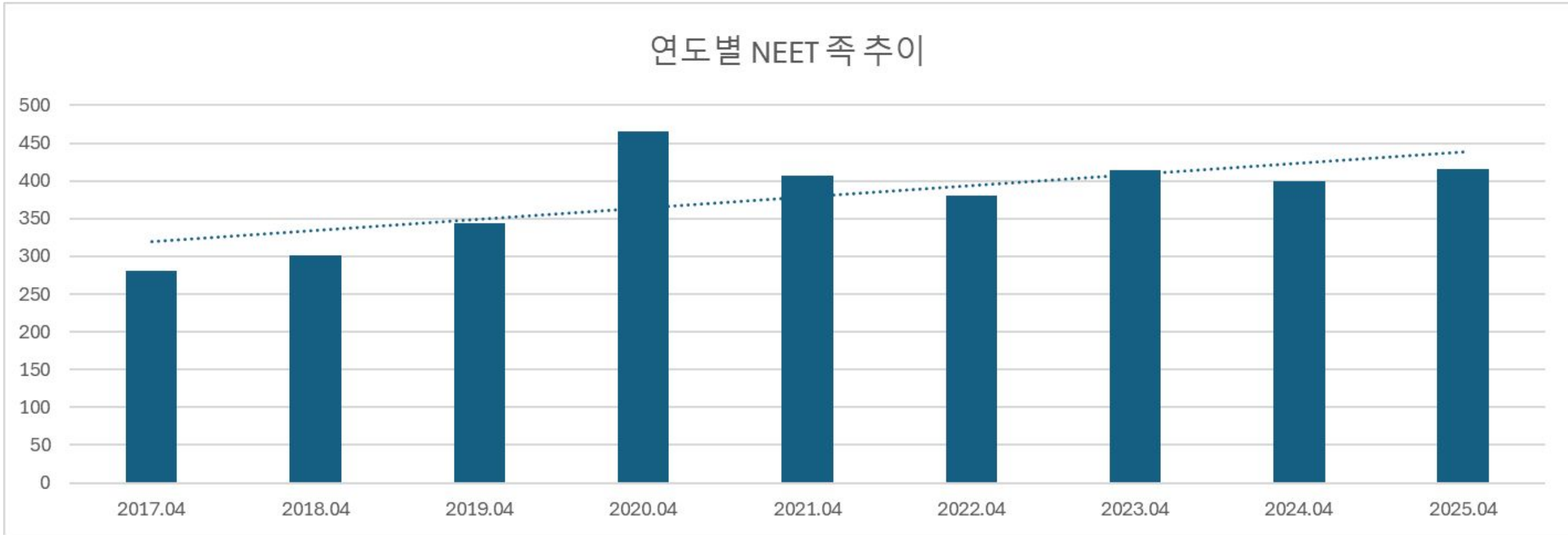


## NEET족이란?

# Not in Education, Employment, or Training

취업자, 일시휴직자, 정규교육기관 재학 학생,  
입시학원이나 취업을 위한 학원, 기관 통학 자가 아닌 사람

# NEET족이란?



출처 : 이규용·금재호·안주엽·윤자영·김기현·이지은·신선옥·이혜정·김기홍(2015). 제17차(2014)년도 한국 가가와 개인의 경제활동 - 한국노동패널 기초분석보고서 -. 세종: 한국노동연구원.

## NEET족이란?



출처 : <https://www.oecd.org/en/data/indicators.html?orderBy=mostRelevant&page=0>

## NEET족 취업난 문제를 해결해야하는 이유



출처 : 게티이미지뱅크

- 경제 성장과 생산성 유지
- 인구 및 미래 세대 문제
- 개인 삶의 질 저하



## 데이터 수집



### 청년 패널 설문

| 구직자 수, (데이터 전처리 끝나면 feature 적기)



### 청년 정책 API

| 현재 정부에서 시행하는 정책



# 데이터 전처리

## 1. 분석 목표 설정

NEET(니트) 족 여부를 식별하고,

NEET 상태에 있는 청년들이 취업을 위해 실제로 어떤 활동을 했는지를 변수 기반으로 분석

## 2. NEET 여부를 판별할 수 있는 변수 추출

"현재 일하지 않고, 교육·직업훈련에도 참여하지 않는 상태"

이를 확인할 수 있는 변수 15개를 선별

## 변수명 매핑

```
# 열 이름 중 한글이 포함된 것만 필터링
korean_columns = [col for col in df.columns if re.search("[가-힣]", str(col))]

# 해당 열만 추출한 새 데이터프레임 생성
df_korean = df[korean_columns]
```

✓ 0.0s

```
df_korean.to_excel("result.xlsx", index=False)
```

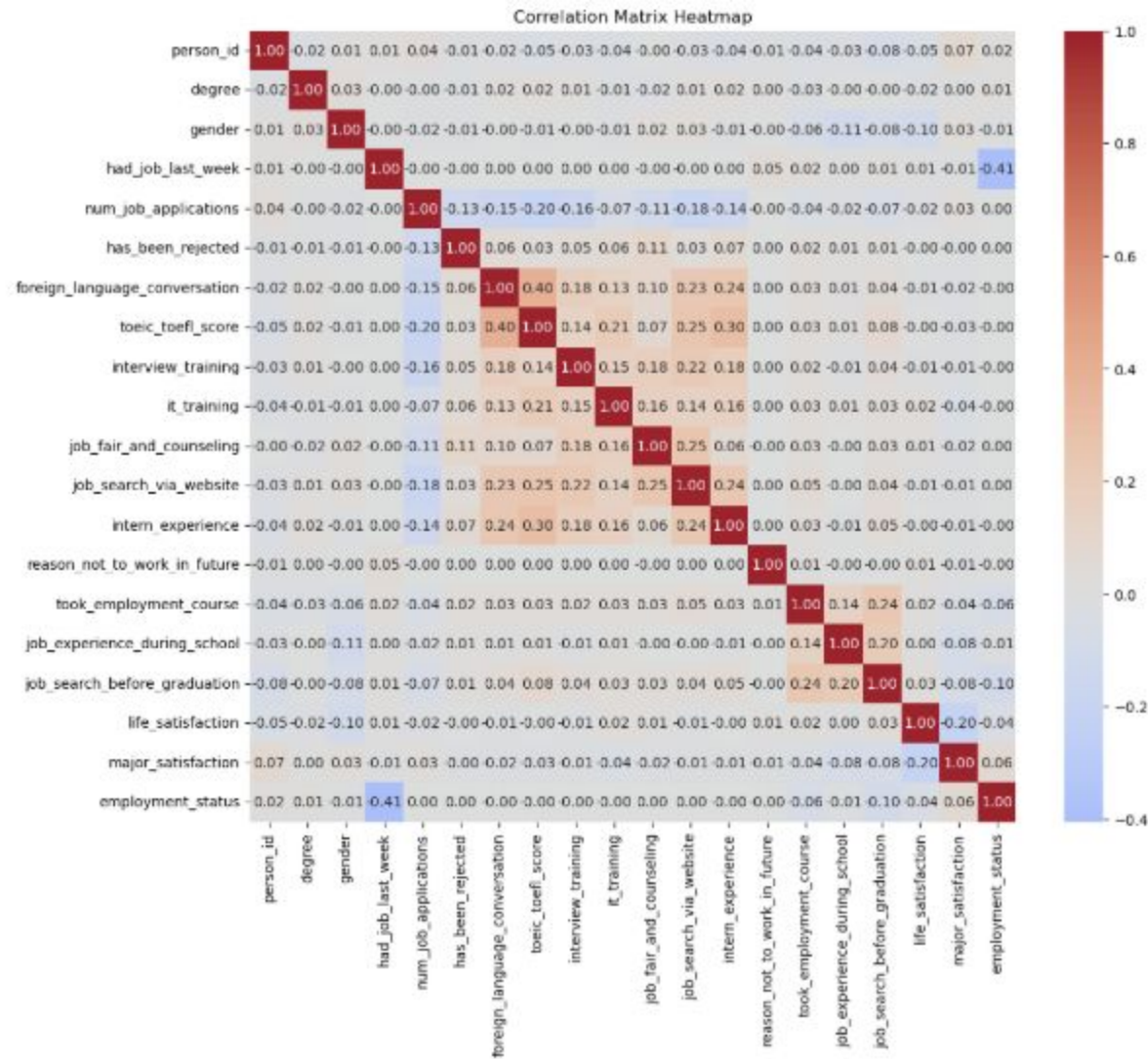
✓ 4.4s

## 결측치 제거&amp;중요피쳐 추출

	학 위	성 별	지난 주에 일했 었으면	자 리 원수 지 회	거 절경 험유무	외 국어 회화	토 익 도 점	면 접 교 과	IT 교 과	취 업 관 심 상 담	인 터 넷 이 구	인 턴 경 험	향 후 이 어 지 는 유 형	취 업 관 과 참 여 부	학 일 중 재 일 지 경 험	졸 업 후 직 동 행	살 의 만 족 도	전 공 만 족 여 부
0	3	2	1.0	1.0	2.0	2.0	1.0	2.0	1.0	1.0	2.0	2.0	0.0	2	1	1	4	1
1	3	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1	2	1	4	2
2	3	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2	1	2	4	2
3	3	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1	1	2	7	2
4	3	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2	2	2	4	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
18266	3	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1	2	2	6	1
18267	3	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1	1	1	7	2
18268	3	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2	2	2	3	2
18269	3	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1	1	2	6	1
18270	3	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1	1	2	3	2

## 데이터 전처리

- 상관관계 분석



## 모델링

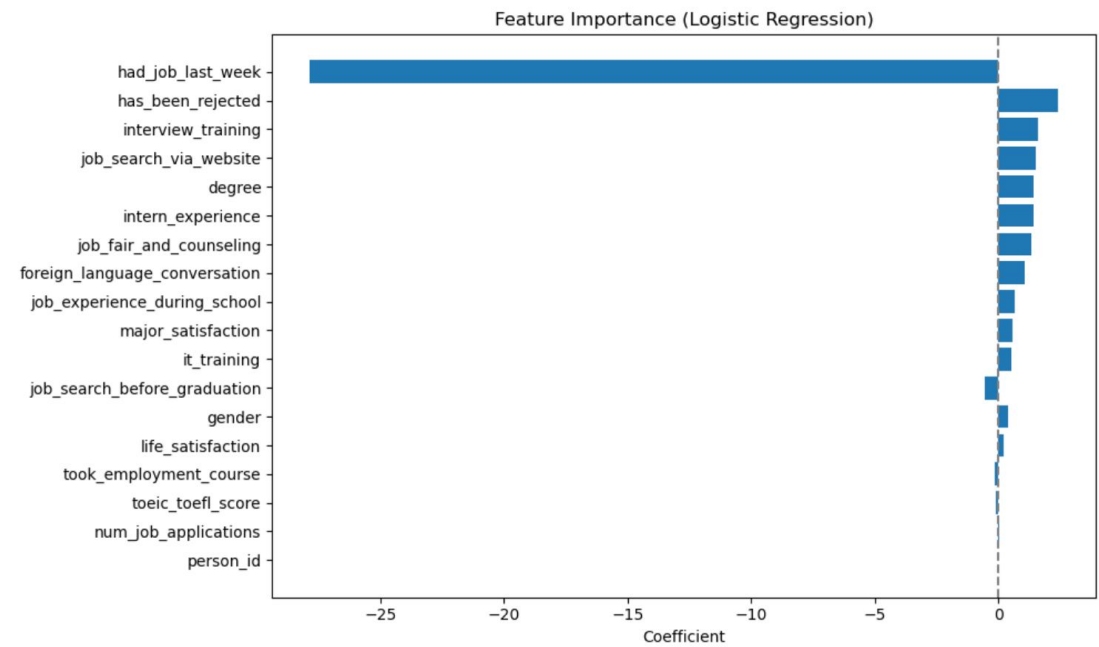
---

- 모델 선정  
Logistic Regression
- 성능 지표
  - $R^2$
  - MAE

# 모델링

- 최종 모델

R <sup>2</sup> Score	Mean Absolute Error
0.5951	0.144



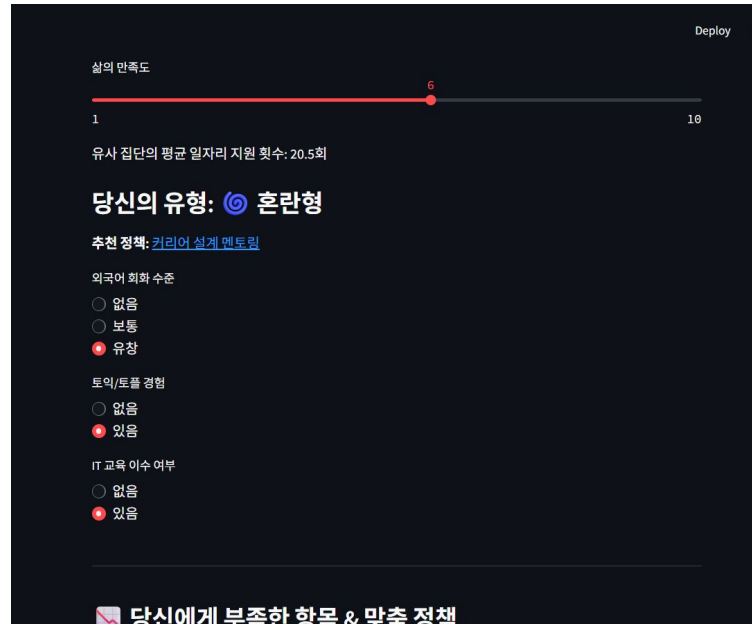
## 한계점

---

- 신규 사용자에게 대해 예측
- 극단적인 값에 대한 성능
- 메뉴 검색 시스템
- 한정된 서비스 범위



- 기존 청년 구직 정책과 연결
  - 기존 정책의 청년 타겟팅 비판,  
NEET 지수를 기반으로 한 새로운 청년 정책 제시
- 웹사이트를 통해 개인별 NEET 지수 확인




Deploy

삶의 만족도

1 6 10

유사 집단의 평균 일자리 지원 횟수: 20.5회

당신의 유형:  혼란형

추천 정책: [커리어 설계 멘토링](#)

외국어 회화 수준

☐ 없음

☐ 보통

☒ 유창

토익/토플 경험


☐ 없음

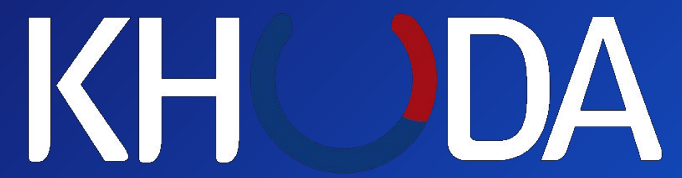
☒ 있음

IT 교육 이수 여부

☐ 없음

☒ 있음

 당신에게 부족한 항목 & 맞춤 정책



3조 Q&A

---

3조 NEET족 구제하기

감사합니다