

세상의 모든 금쪽이들을 위한

---

# 우리가족 Golden Care

---

4조 김나래 이유나 차민수

# 목차

---

01

서론

---

02

자료 수집 및 프로젝트 목표

---

03

텍스트 마이닝 정보 추출

---

04

머신 러닝을 이용한 예측

---

05

서비스 시뮬레이션

---

06

결론

---

---

# 01 서론

- 사용 Tool 및 데이터 소개
  - 주제 선정 과정
  - ‘금쪽이’에 대한 정의
-

# 01 서론

---

## 사용 Tool

---

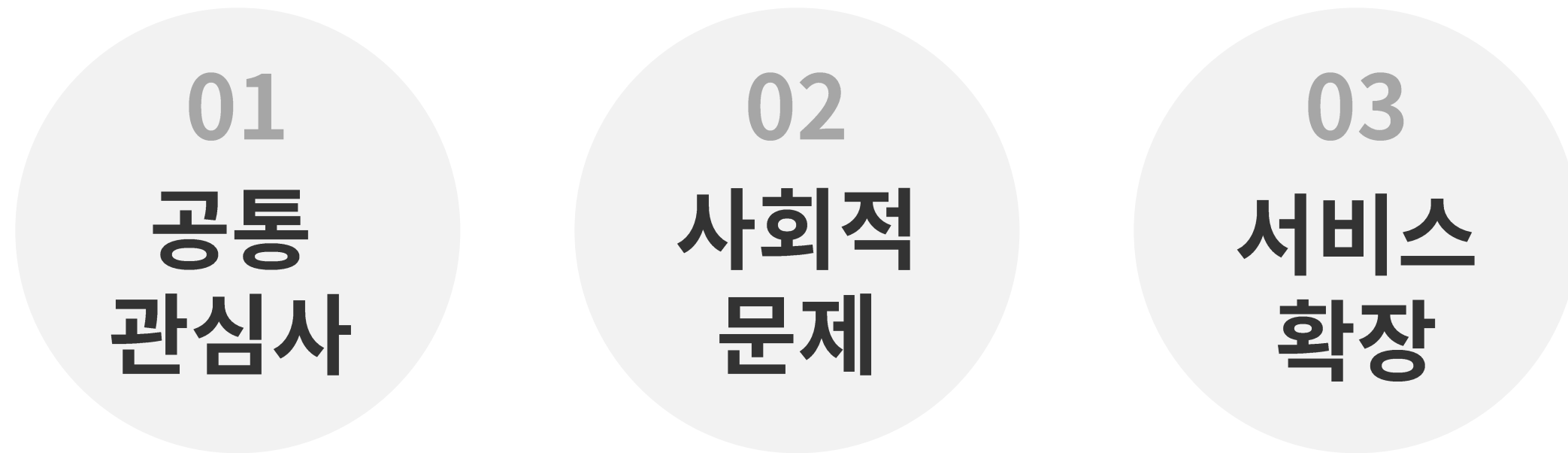
selenium / pandas / matplotlib / seaborn / sklearn

## 사용 데이터

---

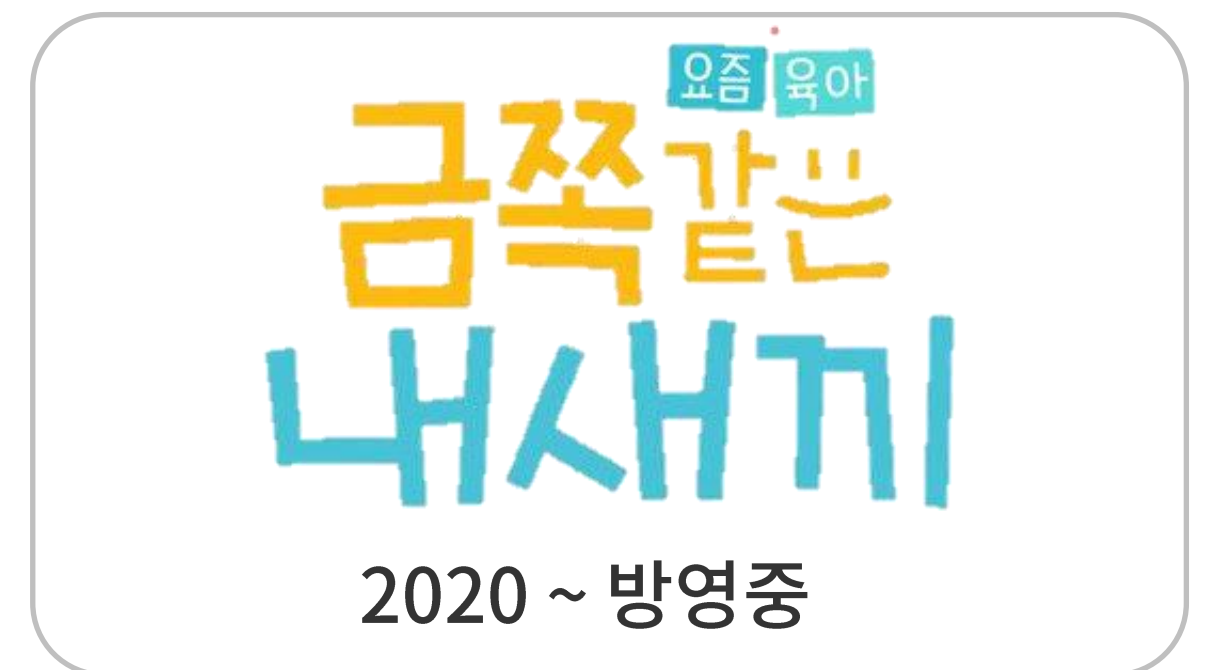
- 행복한 육아문화 정착을 위한 KICCE 육아정책 여론조사 - 육아 정책 연구소
- 네이버 카페(네이버 API), 네이버 뉴스(Selenium)
- 한국 아동·청소년 패널 조사(KCYPS2018) - 한국 청소년 정책 연구원
- 국민관심진료행위(검사/수술 등) 통계 - 건강보험심사평가원

# 01 서론



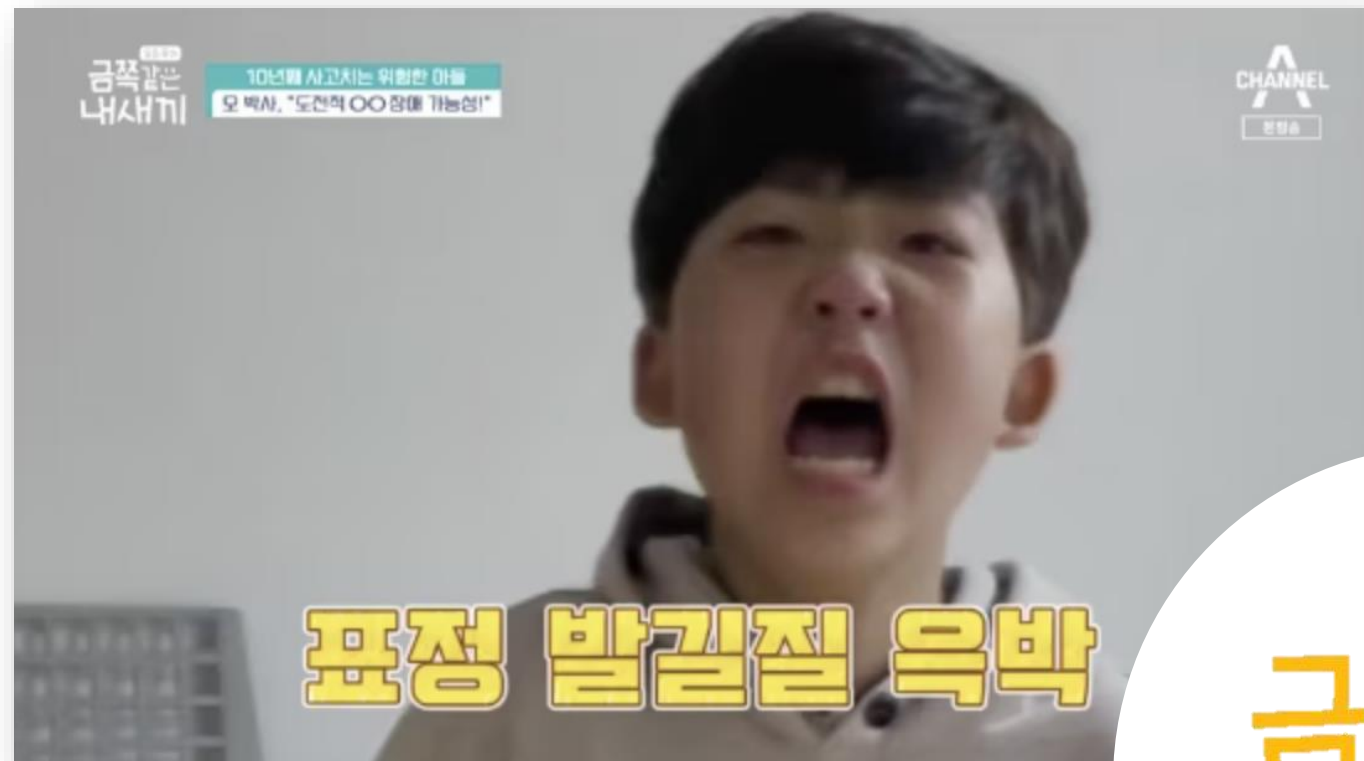
금쪽이는 예전부터 존재했다

**그러나** 관심의 정도는 다르다



# 01 서론 - 주제 선정 과정

분  
노  
조  
절  
장  
애



폭  
력  
성



자  
해



섭  
식  
장  
애



# 01 서론

---

Q. 금쪽이를 한마디로 정의해주세요



교사

선생님과 친구들을  
힘들게 하는 아이



미혼 성인

관심과 애정이  
필요한 아이



부모

이해할 수 없는 아이  
아픈 손가락

‘ 케 어 가 필 요 한 아 이 ’

---

# 02

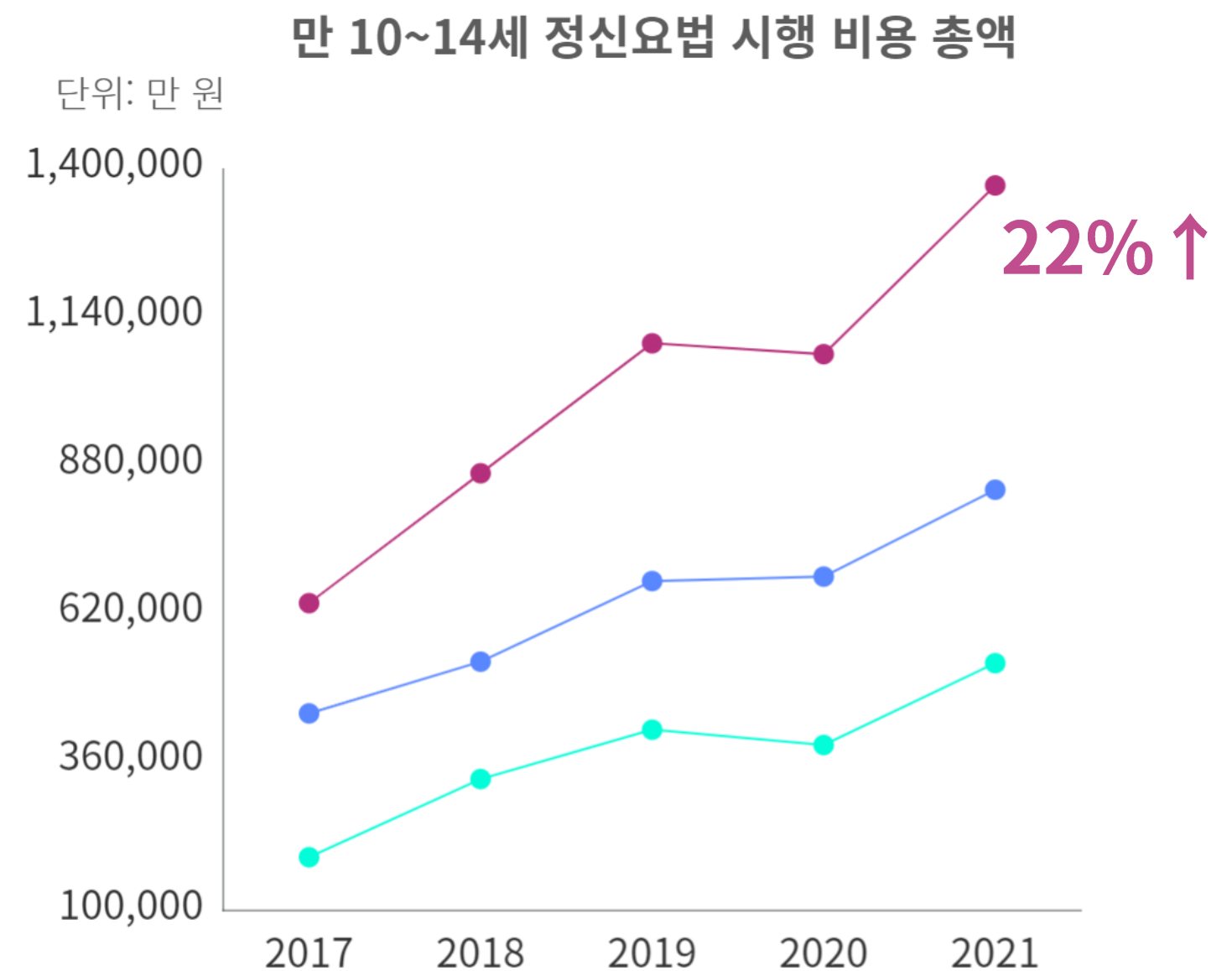
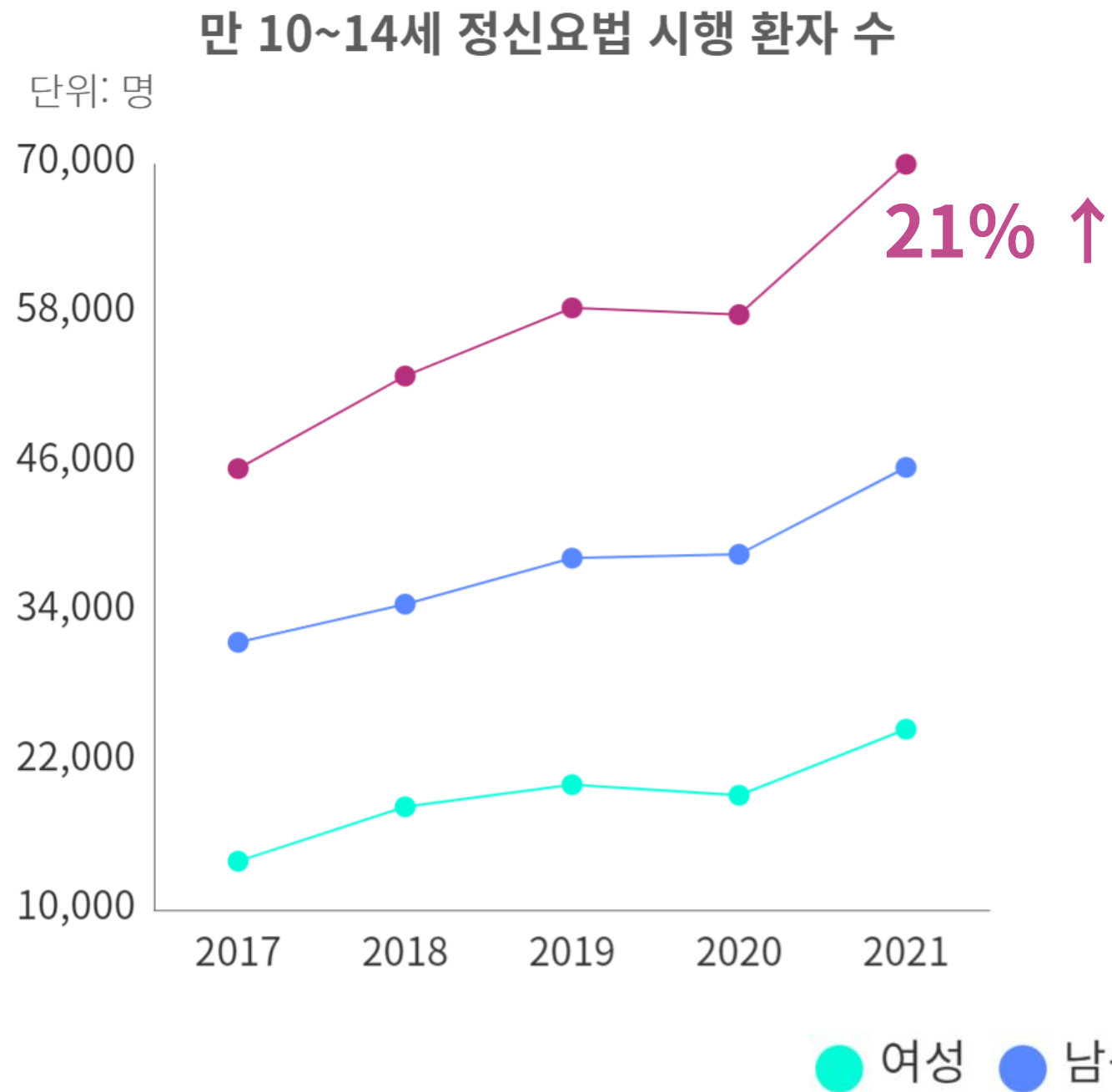
## 자료 수집 및 프로젝트 목표

- 통계로 알아보는 ‘금쪽이’
  - 시행되고 있는 서비스
  - 프로젝트 목표 설정
-



## 02 자료 수집 및 프로젝트 목표

### ✓ 통계로 알아보는 금쪽이 현황

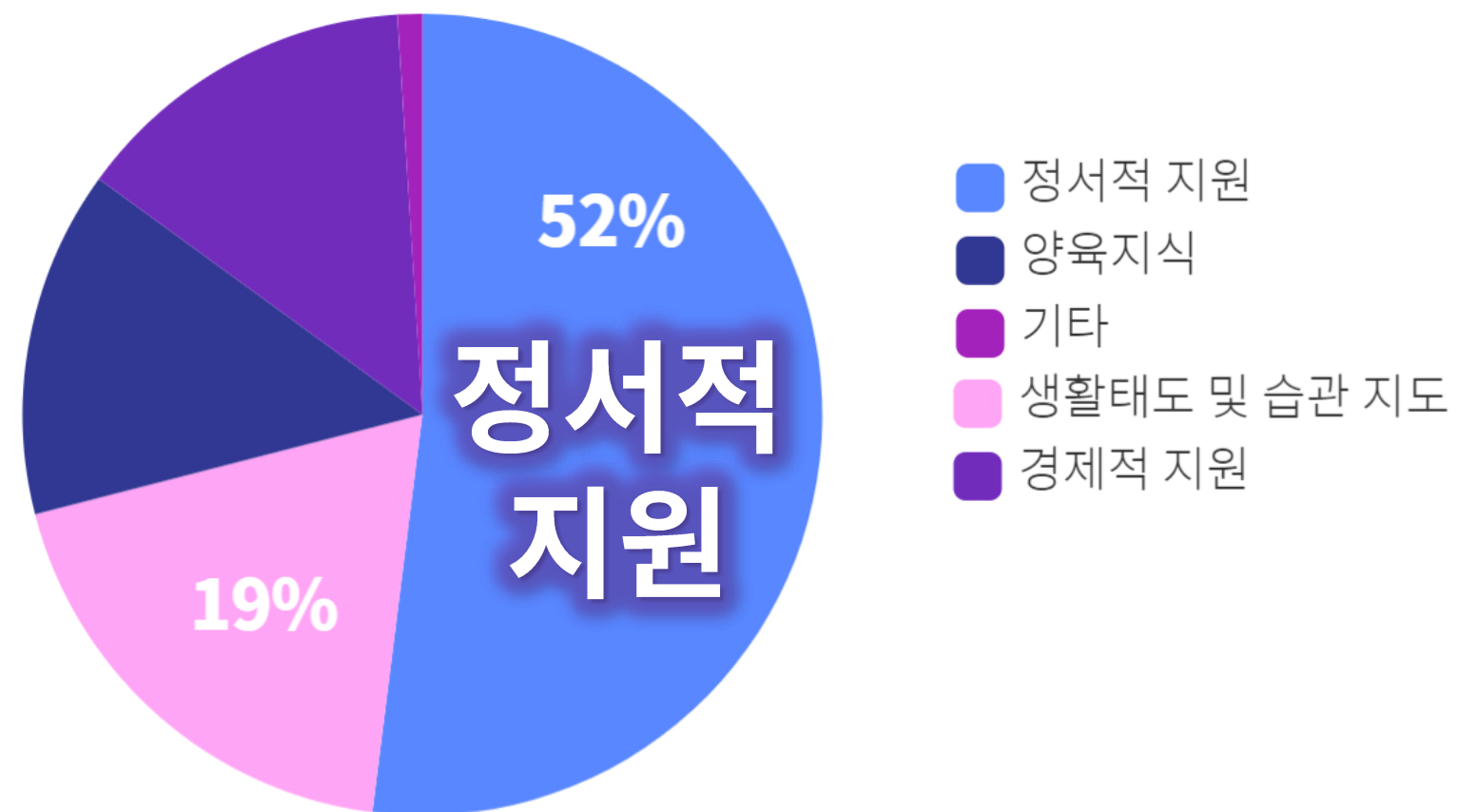


출처: 건강보험심사평가원

## 02 자료 수집 및 프로젝트 목표

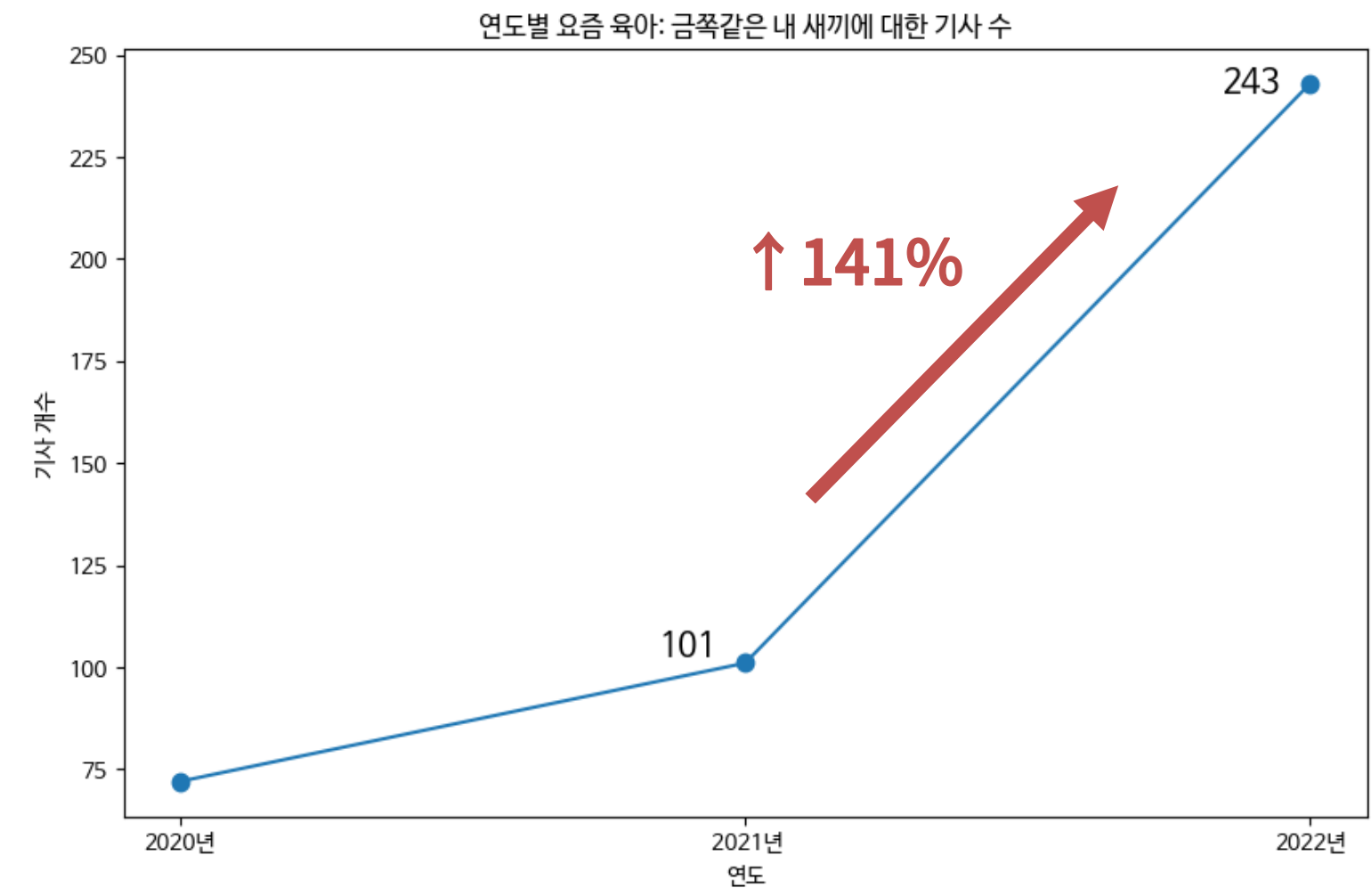
✓ 통계로 알아보는 부모의 마음

좋은 부모란?  
(좋은 부모 역할에서 가장 중요한 것)



출처: 육아 정책 연구소, 2020

✓ 통계로 알아보는 ‘금쪽이’ 관심도



출처 : 스포츠조선, 2020-2022

## 02 자료 수집 및 프로젝트 목표



그로잉맘



킨더 메이트

출시일	2015년	2022년
대상	만 0 ~ 9세	만 0 ~ 9세
장점	믿을 수 있는 전문가 소개	기질, 적성, 부모양육 검사 + 솔루션
단점	유료, 상담 보다는 연결	부모의 입장에서만 관찰

**만 10세 이후를 위한 서비스는 없을까?**

# 질풍노도의 시기

몹시 빠르게 부는 바람과 무섭게 소용돌이치는 물결



출처: 한겨레신문

2차 성징  
예민함 증가

자아의식  
발현

관계에 따른  
정서 변화

작은 변화에도 관심을 기울여  
지속적으로 관찰할 필요가 있다.

## 02 자료 수집 및 프로젝트 목표

---

### 프로젝트 목표

#### 분석목표

- 텍스트 마이닝을 통해 ‘금쪽이’에 대한 수요 조사
- 데이터 시각화를 통해 ‘금쪽이’의 필요성 조사
- 머신러닝을 통해 아동의 정서문제를 예측

#### 프로젝트 목표

- 아동의 상태를 추적 관찰함과 동시에 현재의 상태를 보고
- 아동뿐만 아니라 보호자에 대한 분석도 진행
- 추가 확장 또는 연결할 수 있는 서비스 발견

---

03

# 텍스트 마이닝 정보 추출

- 텍스트 마이닝
- 정보 추출

· 텍스트 분류

· 텍스트 군집

· 텍스트 요약

· 텍스트 생성

---

# 03 텍스트 마이닝과 정보 추출

네이버 카페

네이버 뉴스

검색 기간

2020.05 ~ 2023.06

선정 이유

다수의 맘카페

객관적인 정보 추출 가능

검색 키워드

우리 아들/딸, 정신과, ADHD, 사춘기, 금쪽이

금쪽같은 내새끼

선정 기준

고민글을 작성할 때 쓰는 빈출 단어

스포츠 조선(예능 프로그램 요약)

약 6만 5000개의 문장 수집



## 03 텍스트 마이닝과 정보 추출



혹시  
갑자기



# 03 텍스트 마이닝과 정보 추출



도움 걱정 불안 고민



초진 보험 실비

---

# 04

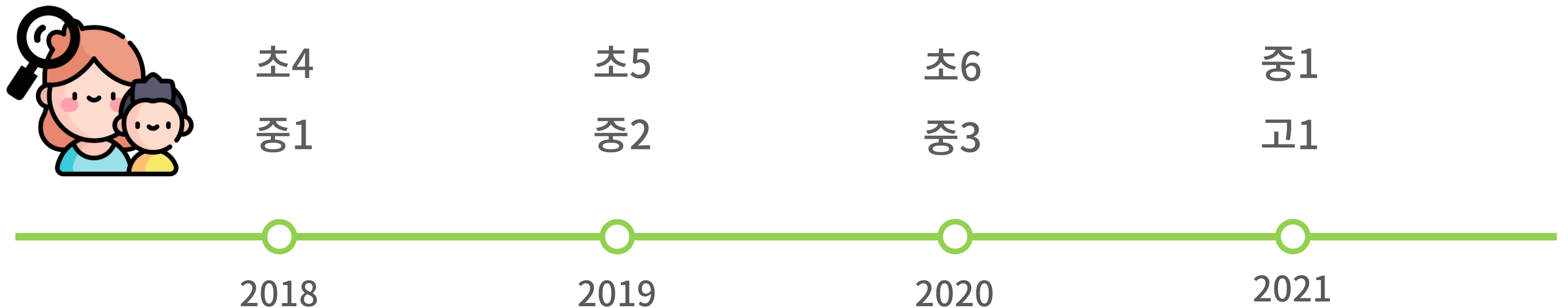
## 머신러닝을 이용한 예측

- 데이터 설명
  - 분석 과정
  - 분석 결과
-

## 05 머신 러닝을 이용한 예측

### 한국 아동·청소년 패널 조사(KCYPS2018)

- 패널 데이터(Panel Data) : 동일한 개체를 시간에 따라 **반복적으로 관측**한 자료
  - 아동과 보호자에게 각각 설문을 진행
  - 설문지의 항목들은 전문가들에 의해 짜여진 진단도구



## 05 머신 러닝을 이용한 예측

---

### 데이터 전처리 (Data Preprocessing)

**STEP 1**

필요한  
변수 선정

**STEP 2**

결측치  
처리

**STEP 3**

변수 변환  
점수화

# 05 머신 러닝을 이용한 예측

## STEP 1 필요한 변수 선정

조사년도

성별

거주지

월 평균 가구소득

월 평균 사교육비

일상 생활

자기관리

여가생활

학업 성취 만족도

학교 생활 만족도

개인 발달

삶의 만족도/행복감

자아인식

끈기

협동심

정서문제

발달 환경

스마트폰 의존도

부모 관계

학생 관계

교사 관계

# 05 머신 러닝을 이용한 예측

## STEP 2 결측치 처리

Id	X1	X2	X3	...
1_18				
2_18				
3_18				
⋮				
1_19				
2_19		NA		
3_19				
⋮				
1_20				
2_20				
3_20				
⋮				

- 사교육 여부 2(아니오) → 월 평균 사교육비 0
- 월 평균 가구소득 무응답자(9999) → 삭제
- 과반 수 이상의 결측치인 행(ID)는 '무응답자' 라고 정의 → 삭제
- 그 외의 결측치 → 삭제

20788 → 15720

각 년도마다 3930건

# 05 머신 러닝을 이용한 예측

## STEP 3 변수 변환 / 점수화

### 1. 방향성을 통일하기 위해 반대 되는 의미는 역코딩 진행

(예제) 스마트폰 의존도 설문 일부

스마트폰의 지나친 사용으로 학교성적이 떨어진다
수시로 스마트폰을 사용하다가 지적을 받은 적이 있다
스마트폰을 너무 자주 또는 오래한다고 가족이나 친구들로부터 불평을 들은 적이 있다
스마트폰을 사용하느라 지금 하고 있는 일(공부)에 집중이 안 된 적이 있다
스마트폰 사용이 지금 하고 있는 일(공부)에 방해가 되지 않는다
스마트폰을 사용하지 못하면 온 세상을 잃은 것 같은 생각이 든다
가족이나 친구들과 함께 있는 것보다 스마트폰을 사용하고 있는 것이 더 즐겁다
스마트폰을 사용할 수 없게 된다면 견디기 힘들 것이다
스마트폰이 없으면 안절부절 못하고 초조해진다
스마트폰이 없어도 불안하지 않다

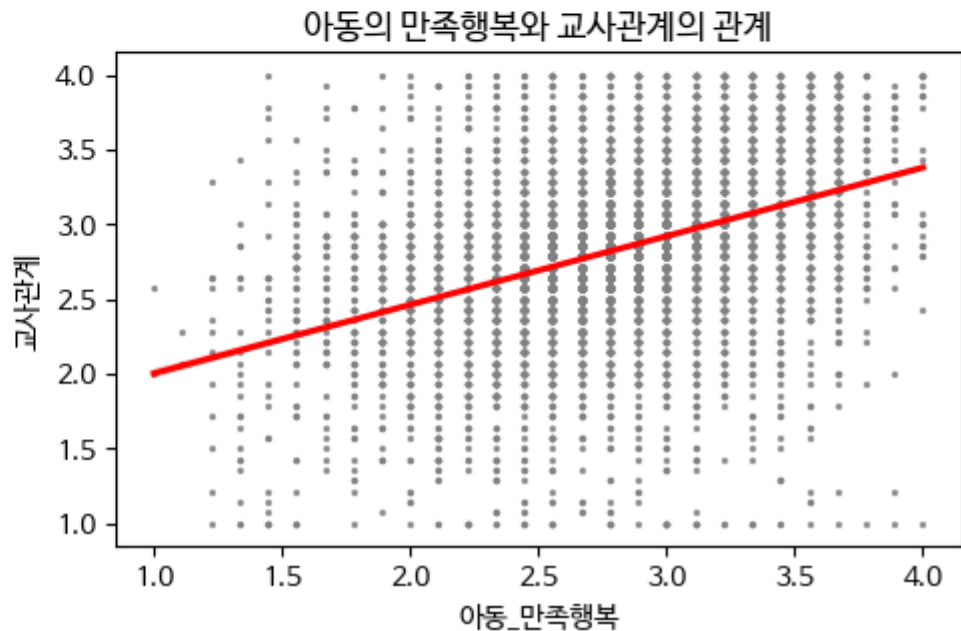
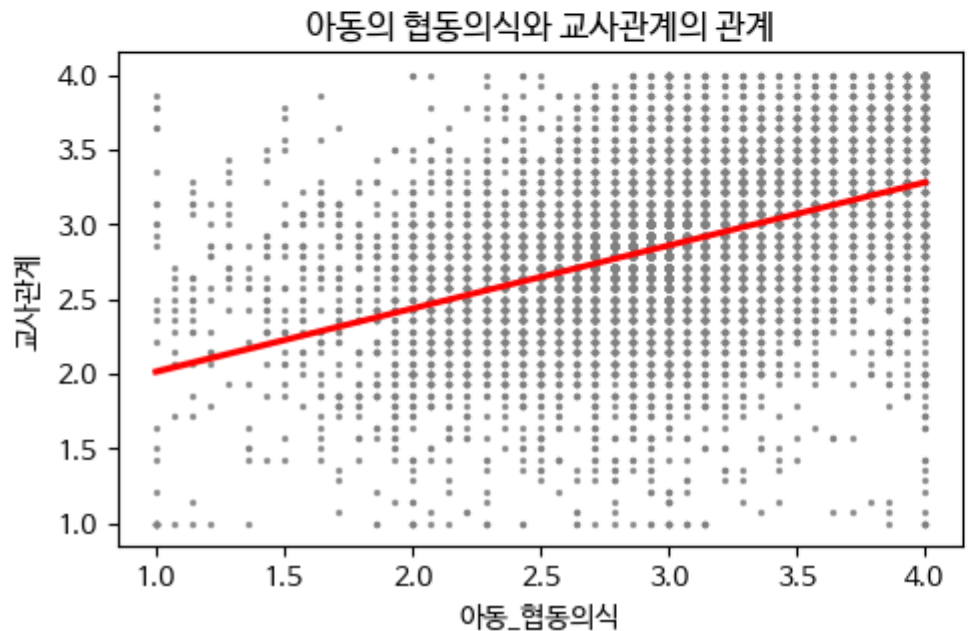
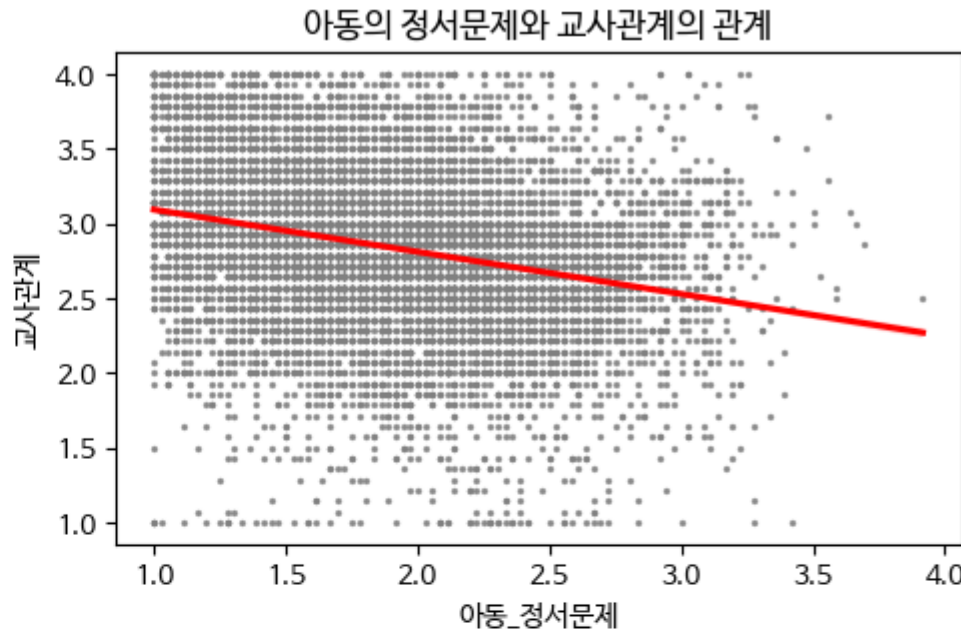
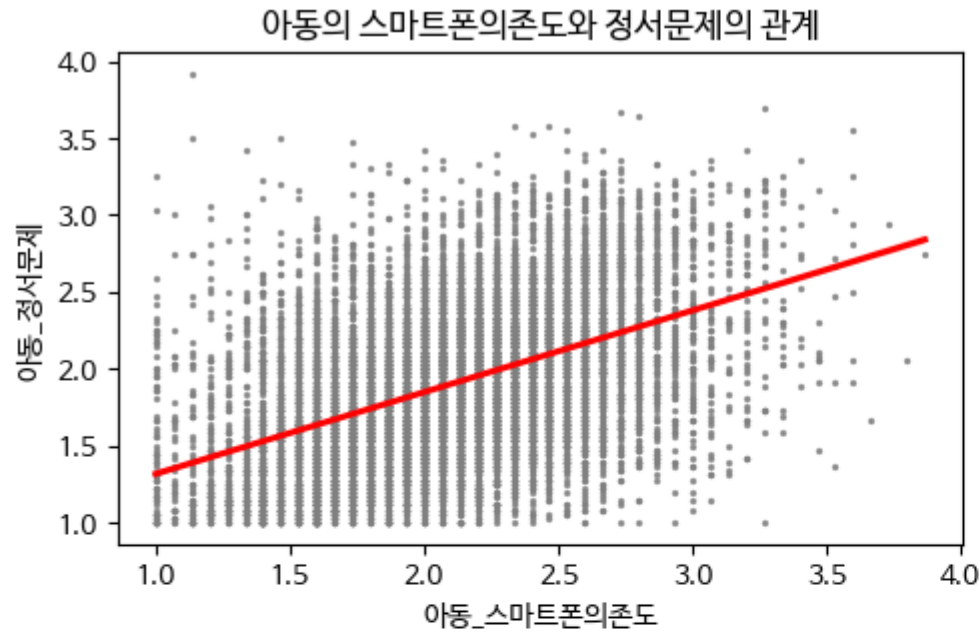
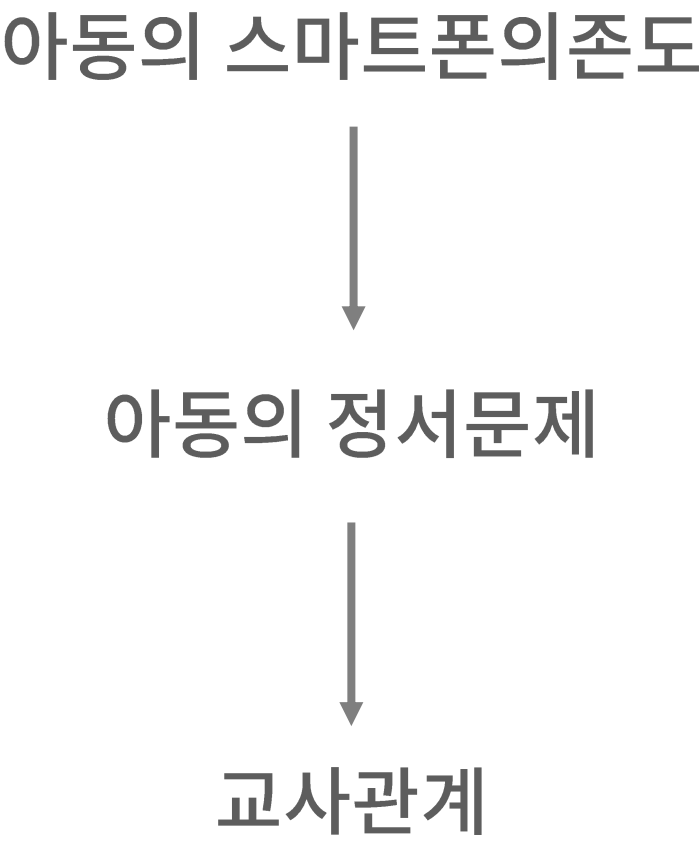
최종 데이터

15720 × 27

### 2. 각 지표마다의 설문 점수를 평균으로 대체



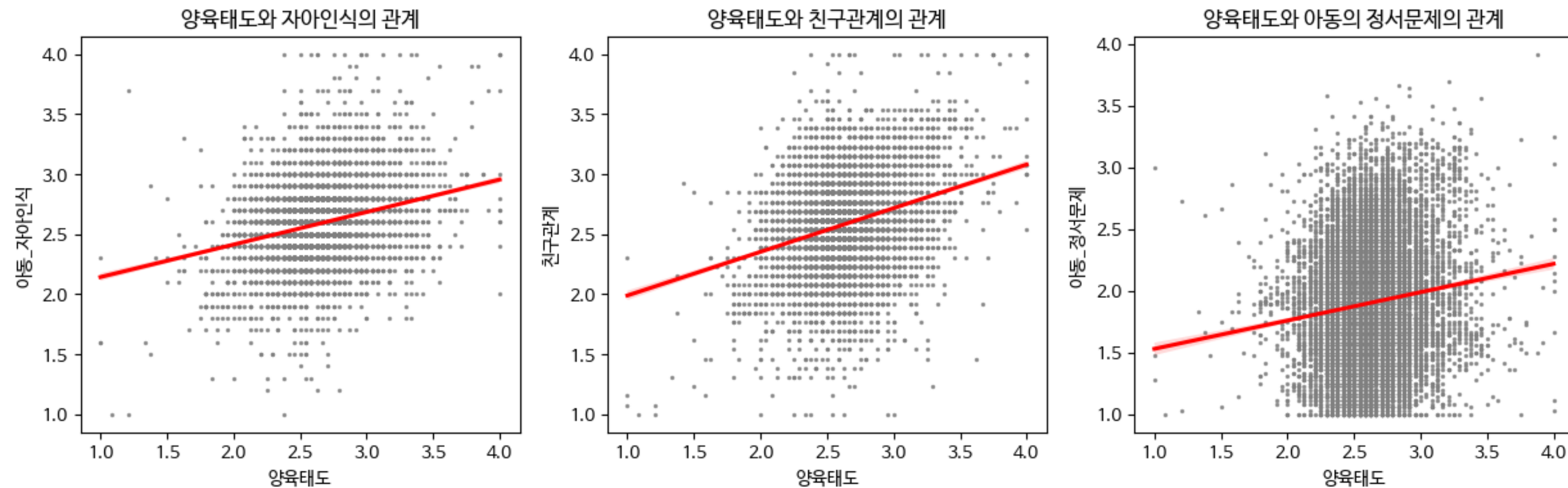
시각화 1 변수간의 관계





## 시각화 2 변수간의 관계

양육태도 : 따스함/거부/자율성지지/강요/구조제공/비일관성. 아동의 관점으로 조사한 항목

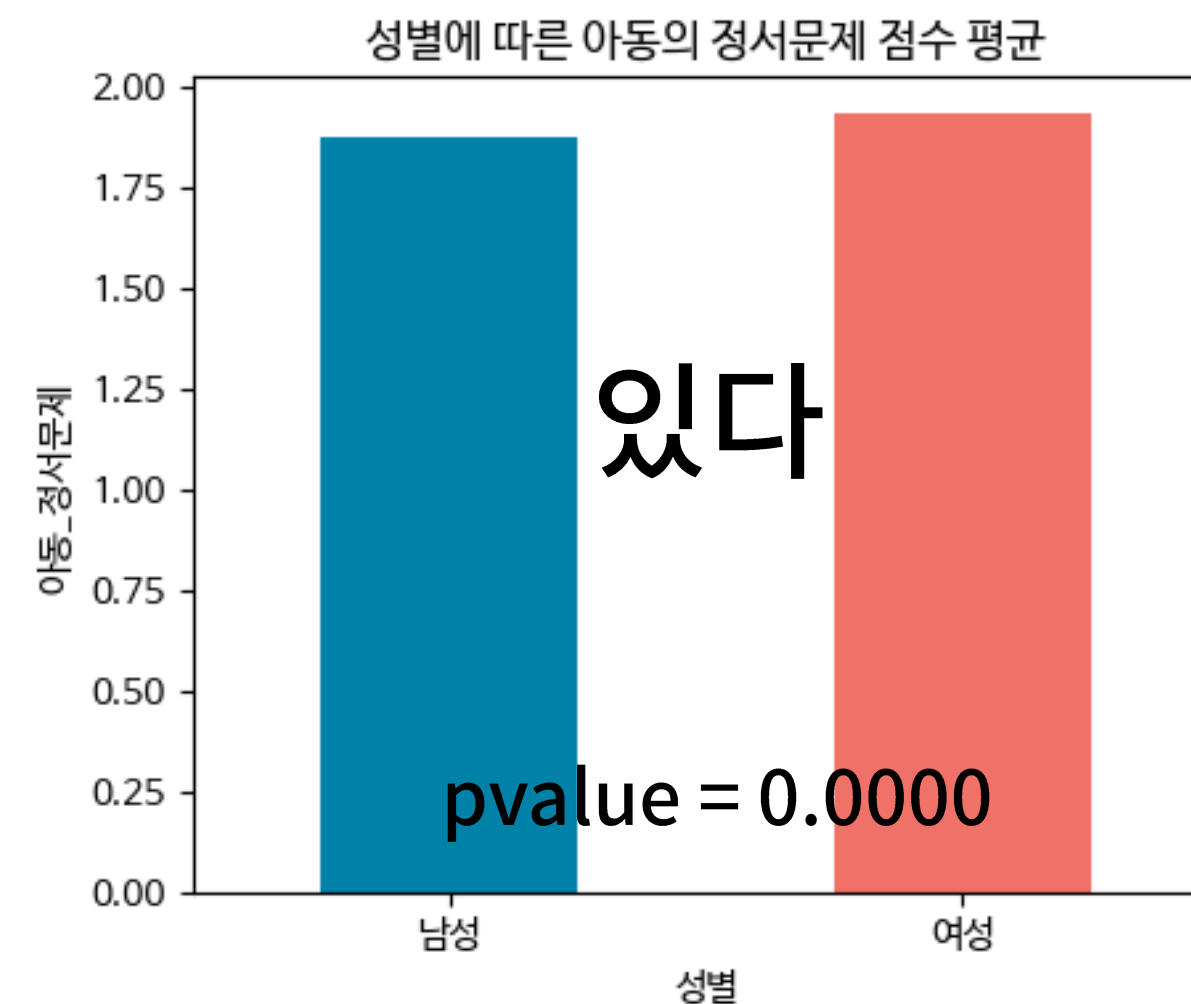
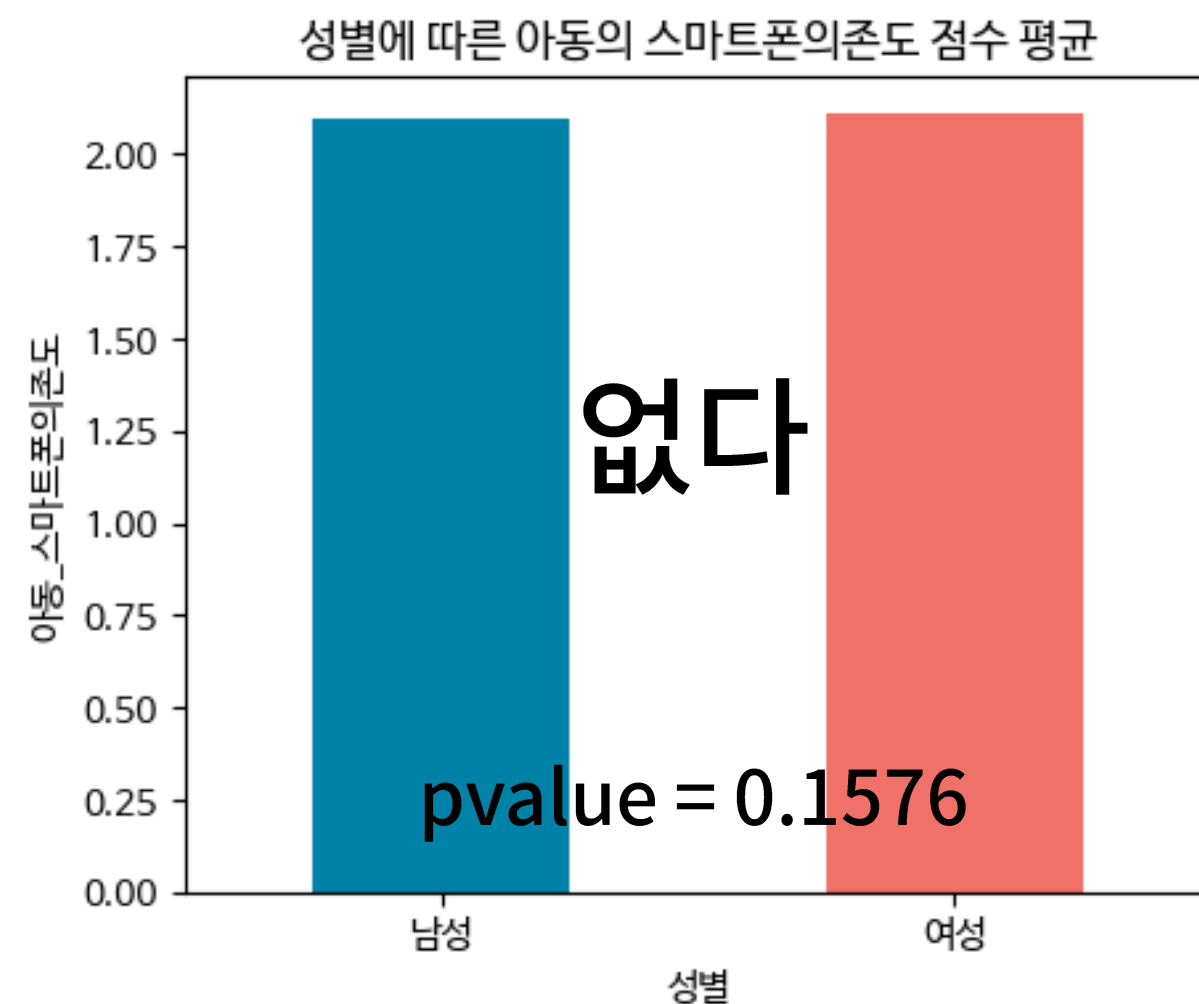


보호자의 양육태도가 ‘금쪽이’를 만들 수도 있다.

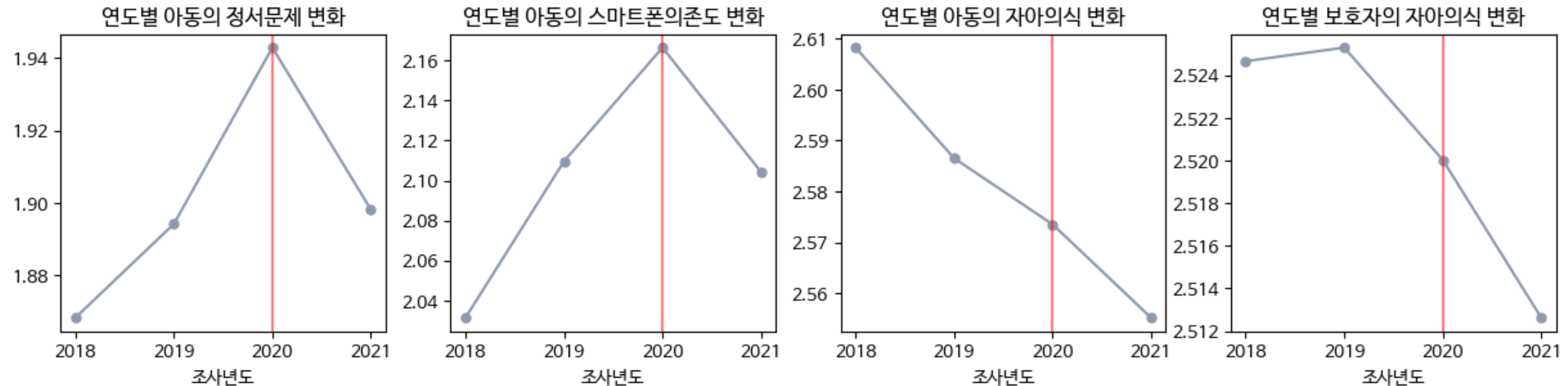
## 시각화 3 성별간의 차이

H0: 성별 스마트폰의존도(정서문제)에 차이가 없다.

H1: 성별 스마트폰의존도(정서문제)에 차이가 있다.



## 시각화 3 연도별 차이



코로나를 제외해도 이전부터 정서문제, 스마트폰 의존도는 계속 높아지고 있는 상태  
아동뿐만 아니라 보호자 또한 자아의식이 낮아지고 있는 상태

## 모델 선정 : 성능 기준 RMSE

1. Lazypredict를 이용한 AutoML 실행 → **[Error] 리소스 용량 초과**
2. PyCaret 을 이용한 AutoML 실행 → Top 5 모델 추출
3. Lazypredict에 Top 5 모델만 실행 → 성능의 비슷한 정도를 확인

Extra Trees

0.3373

LightGBM

0.3388

RandomForest

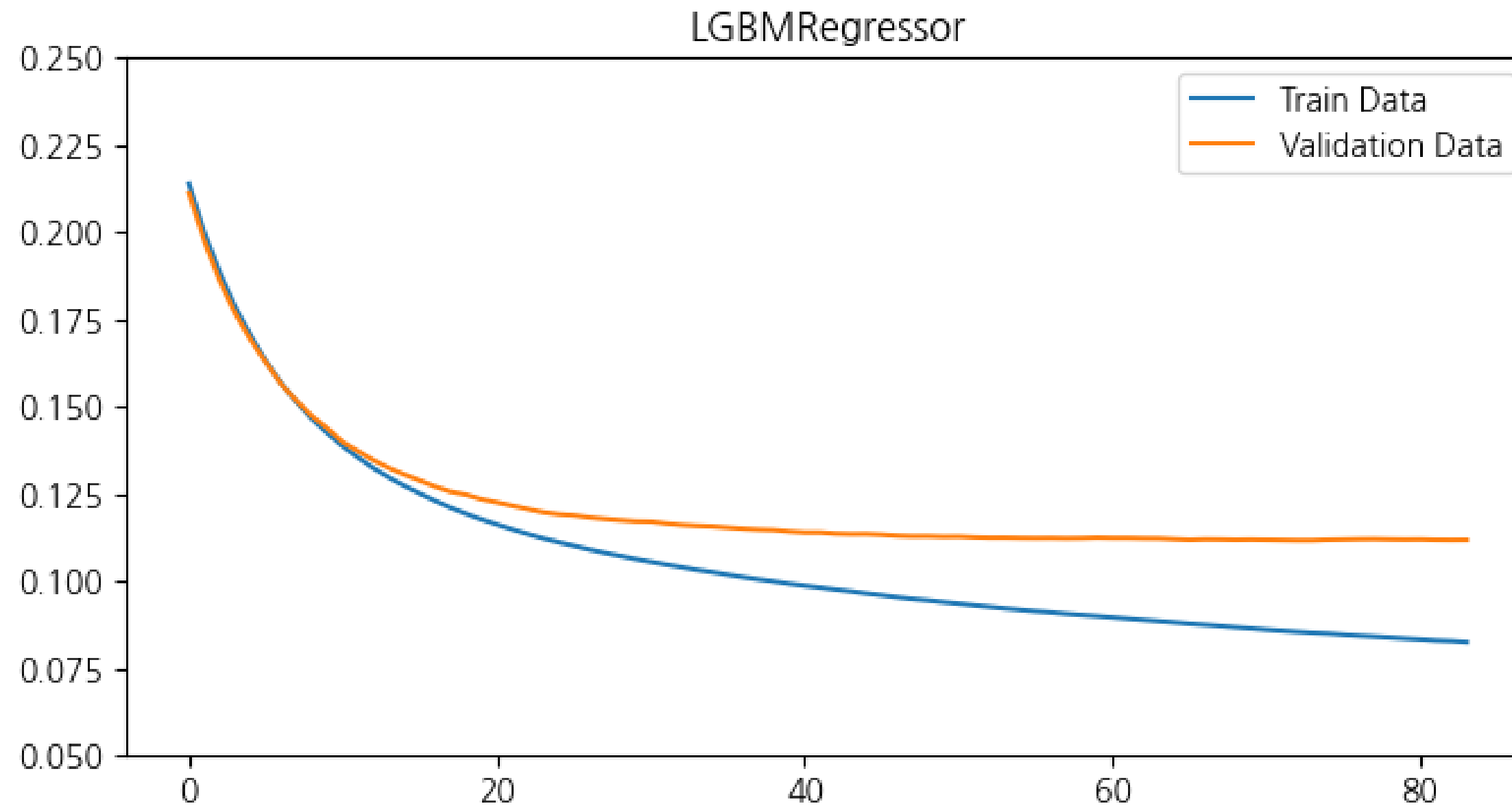
0.3409

✓ 분석 결과

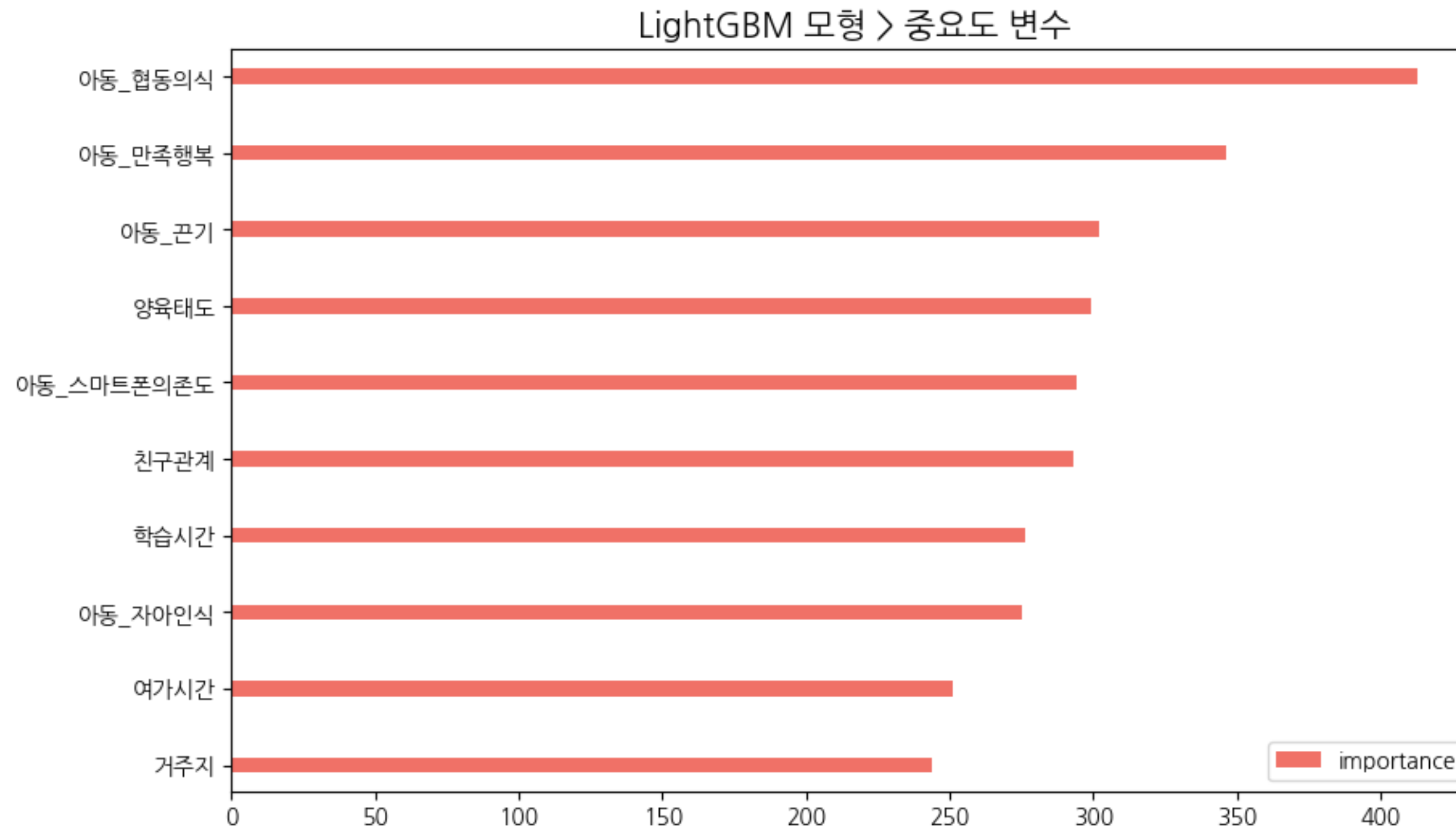
LightGBM > Random Forest > Extra Trees

Model	Train RMSE	Test RMSE
Extra Trees	0.1809	0.3355
LightGBM	0.2919	0.3337
RandomForest	0.1807	0.3353

## ✓ 분석 결과



## ✓ 분석 결과 - 중요도 변수



---

# 05 서비스 시뮬레이션

- 서비스 소개
- 시뮬레이션



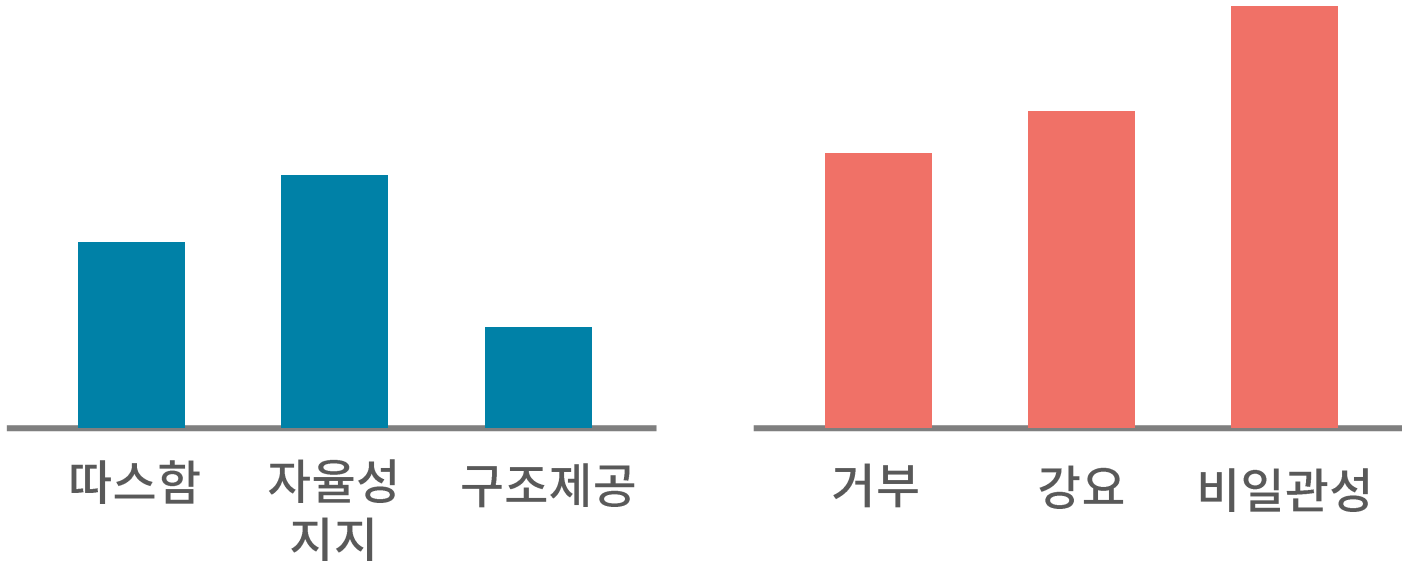
✓ 서비스 소개

Golden  
Care

대상	만 10세 ~ 16세
기능	아이의 정서 상태 예측
차별점	<ul style="list-style-type: none"><li>• 아동의 관점에서 상태를 파악</li><li>• 아동과 보호자의 관계 파악</li><li>• 항목을 세분화 하여 원인 파악</li><li>• 이를 종합한 상담 솔루션 제공</li><li>• 데이터 축적으로 지속적인 관리 가능</li></ul>

# 05 서비스 시뮬레이션

✓ 시뮬레이션



---

# 06

## 결론

- 앞으로의 과제
  - 결론
-

## ✓ 앞으로의 과제

- 현재 서비스는 ‘설문조사’의 형태로 아동이 느끼기에는 딱딱하고 지루한 형태일 수 있음  
이를 재미있게 풀어낼 수 있는 서비스를 생성해야 함.(ex. 이미지, 음성)
- 이 서비스는 ‘금쪽이’가 될 수 있는 아동을 사전에 예방하고, 지속적으로 관리하기 위함.  
따라서 전문가의 의견을 포함한 문항 연구와 솔루션 연구가 필요함.

## ✓ 결론

- 이론에서 배운 데이터의 형식과 달리 여러 개체가 반복으로 측정된 데이터를 사용해보며  
머신러닝의 우수성을 확인함
- 모든 사람들이 공감할 수 있는 문제를 데이터(정형, 비정형) 다각도로 살펴보고,  
눈으로 볼 수 없는 영향력을 확인함

---

**THANK YOU**

---