

시대 흐름에 맞는 선호 학과 분석 및 예측

BigTS

이채은, 김부건, 이윤승, 임연주, 표성민

목차

- 수행 배경
- 수행 과정
- 결과
- 기대효과 및 활용방안

수행 배경

대학 졸업자 10명중 7명이 본인의 전공 후회

>> 다시 선택할 수 있다면
취업이 더 잘되는 학과 희망

직장인 10명중 7명, ‘전공 재선택 하고 파’

(자료 : 잡코리아x알바몬, 남녀 직장인 706명 조사)

Q. 다시 선택하고 싶은 전공 계열은 무엇입니까?



JOBKOREA x albamon

수행 배경



학과 트렌드 분석

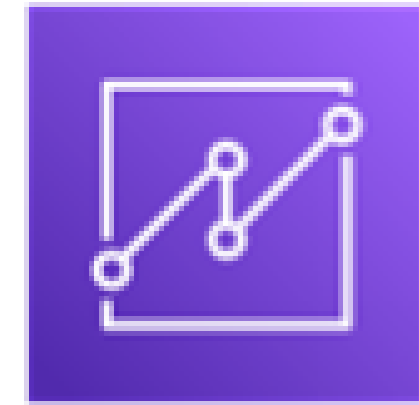
수행 과정 - 설계



교육통계서비스
2008-2023 학과별 데이터



Jupyter Notebook
Python



Amazon QuickSight

데이터 수집 및 정제

데이터 분석

정리 및 시각화

수행 과정 - 상세내용

데이터 수집 및 정제

1

2

3

4

5

6

7

8

9

10

11

12

13

14

6694

6695

6696

6697

6698

2023

기능대학

전문대학과정

예체능계열

디자인

산업디자인

C07010100010

산업디자인과

3

0

0

85

85

0

102

85

17

2023

기능대학

전문대학과정

예체능계열

디자인

시각디자인

C07010200002

시각디자인과

2

0

0

90

90

0

108

90

18

2023

기능대학

전문대학과정

예체능계열

디자인

패션디자인

C07010300011

패션디자인과

1

0

0

60

60

0

72

60

12

2023

기능대학

전문대학과정

예체능계열

디자인

패션디자인

C07010300063

패션디자인과(2년제)

1

0

0

0

0

0

0

0

0

2023

기능대학

전문대학과정

예체능계열

디자인

패션디자인

C07010300080

스마트패션디자인과

0

0

0

60

60

0

72

60

12

필요없는 데이터 삭제, UTF-8형식 인코딩

수행 과정 - 상세내용

정제한 파일 불러오기

```
In [2]: df = pd.read_csv('2008-2023.csv')
```

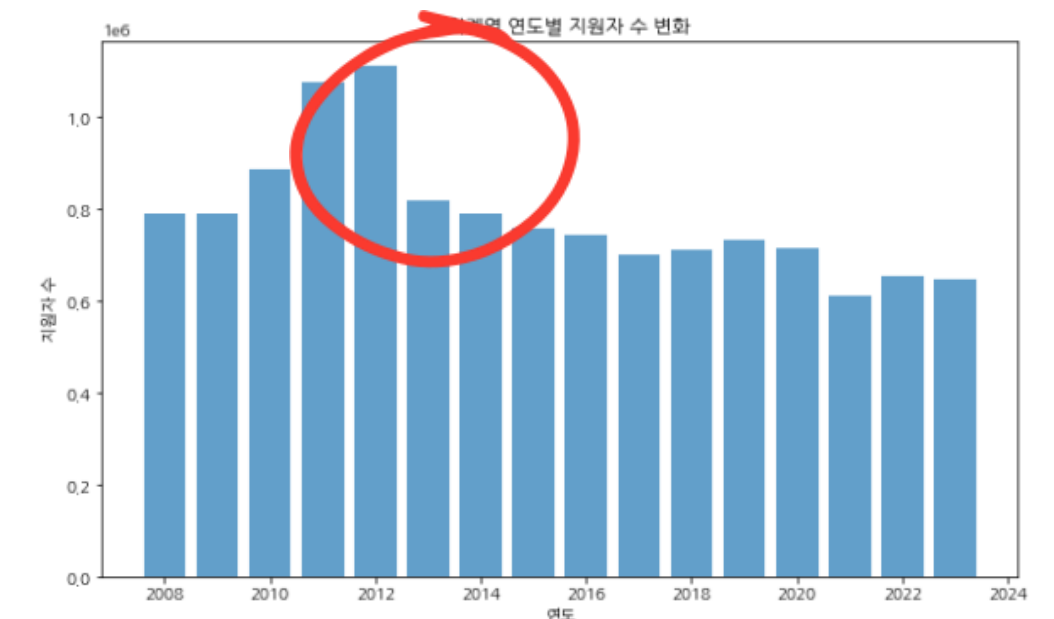
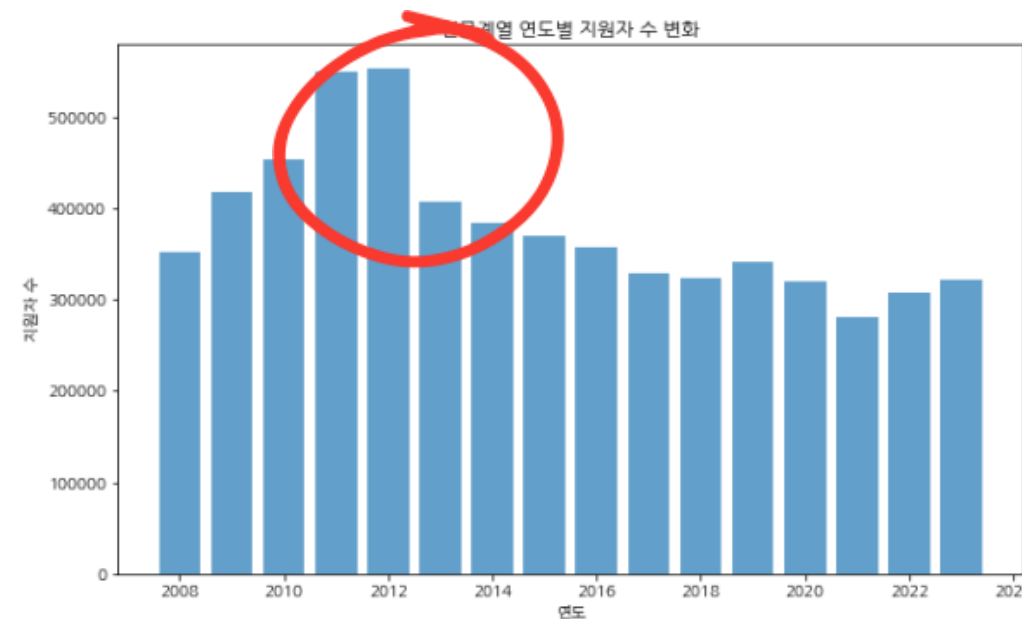
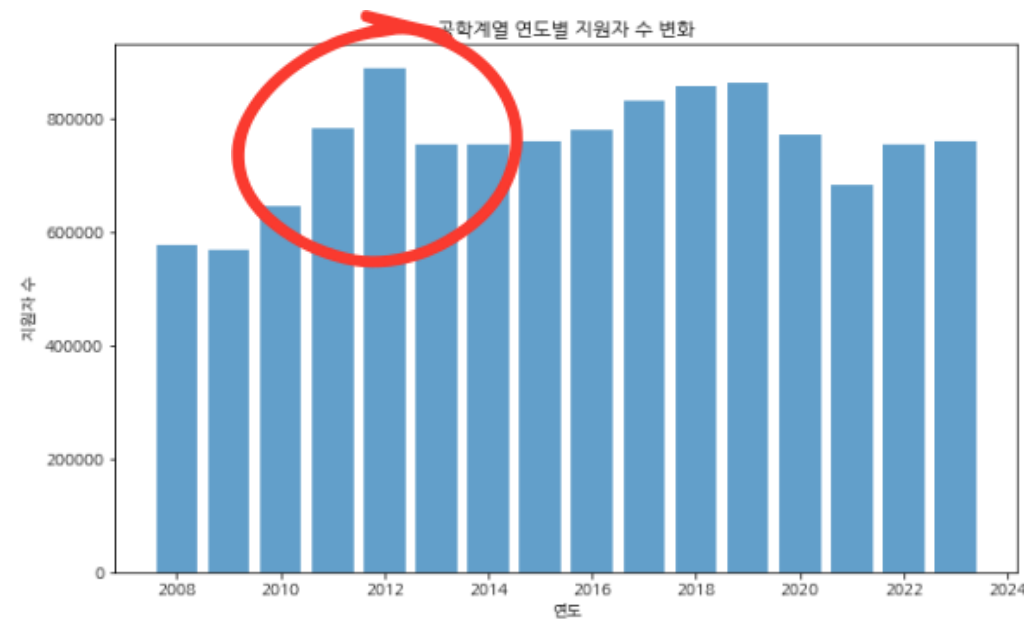
```
In [3]: df.head()
```

Out [3]:

	연도	대 계 열	중 계 열	소 계 열	학과 명	학 과 수 _ 전 체	지원자 _ 전체_ 계	지원 자_전 체_남	지원자 _ 전체_ 여	입학 자_전 체_계	입학 자_전 체_남	입학 자_전 체_여
0	2008	인문계열	언어·문화	언어학	언어학과	3	76	21	55	9	3	6
1	2008	인문계열	언어·문화	언어학	언어정보학과	1	204	71	133	34	12	22

수행 과정 - 상세내용

데이터 분석



In [32]: grouped2_df

Out [32]:

	연도	학과수_전체	입학자_전체_계	지원자_전체_계
0	2008	7727	342105	2481309
1	2009	8085	346715	2605032
2	2010	7992	357441	2976456
3	2011	8163	360748	3528390
4	2012	8648	372701	3753705
5	2013	8785	365041	3024571
6	2014	8787	362527	3000752

2012년 -> 2013년

3753705 -> 3024571

급격히 줄어든 지원자 수

수행 과정 - 상세내용

데이터 분석

In [10]: #한 연도의 모든 소계열 비율의 합을 1로 잡고 각 소계열의 비율 구하기

```
eng_df = grouped4_df
eng_df['학과수_비율'] = eng_df['학과수_전체'] / eng_df.groupby('연도')['학과수_전체'].transform('sum')
eng_df['입학자_비율'] = eng_df['입학자_전체_계'] / eng_df.groupby('연도')['입학자_전체_계'].transform('sum')
eng_df['지원자_비율'] = eng_df['지원자_전체_계'] / eng_df.groupby('연도')['지원자_전체_계'].transform('sum')
eng_df['소계열_경쟁률'] = eng_df['지원자_전체_계'] / eng_df['입학자_전체_계']

eng_df
```

Out [10]:

	연도	대계열	중계열	소계열	학과수_전체	입학자_전체_계	지원자_전체_계	학과수_비율	입학자_비율	지원자_비율	소계열_경쟁률
0	2008	공학계열	건축	건축·설비공학	111	3215	22610	0.014365	0.009398	0.009112	7.032659
1	2008	공학계열	건축	건축학	118	4484	30028	0.015271	0.013107	0.012102	6.696699
2	2008	공학계열	건축	조경학	16	504	2820	0.002071	0.001473	0.001136	5.595238
3	2008	공학계열	교통·운송	지상교통공학	4	146	950	0.000518	0.000427	0.000383	6.506849

연도별 학과수, 입학자 수, 지원자 수 비율 및 소계열별 경쟁률을 이용하여 분석 결정

수행 과정 - 상세내용

데이터 분석

```
In [12]: # 2008년 데이터만 선택
year1_2008_df = eng_df[eng_df['연도'] == 2008]

# 비율의 합 계산
sum_of_ratios = year1_2008_df[['학과수_비율', '입학자_비율', '지원자_비율']].sum()

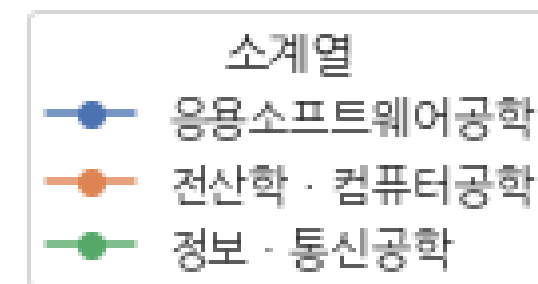
# 결과 출력
print(sum_of_ratios)
```

```
학과수_비율    1.0
입학자_비율    1.0
지원자_비율    1.0
dtype: float64
```

수행 과정 - 상세내용

데이터 분석

컴퓨터통신계열 시각화 예시



```
import matplotlib.pyplot as plt
import seaborn as sns

# 한글 폰트 설정
plt.rcParams['font.family'] = 'NanumGothic'

# 그래프 그리기
fig, axes = plt.subplots(nrows=4, ncols=1, figsize=(8, 17))

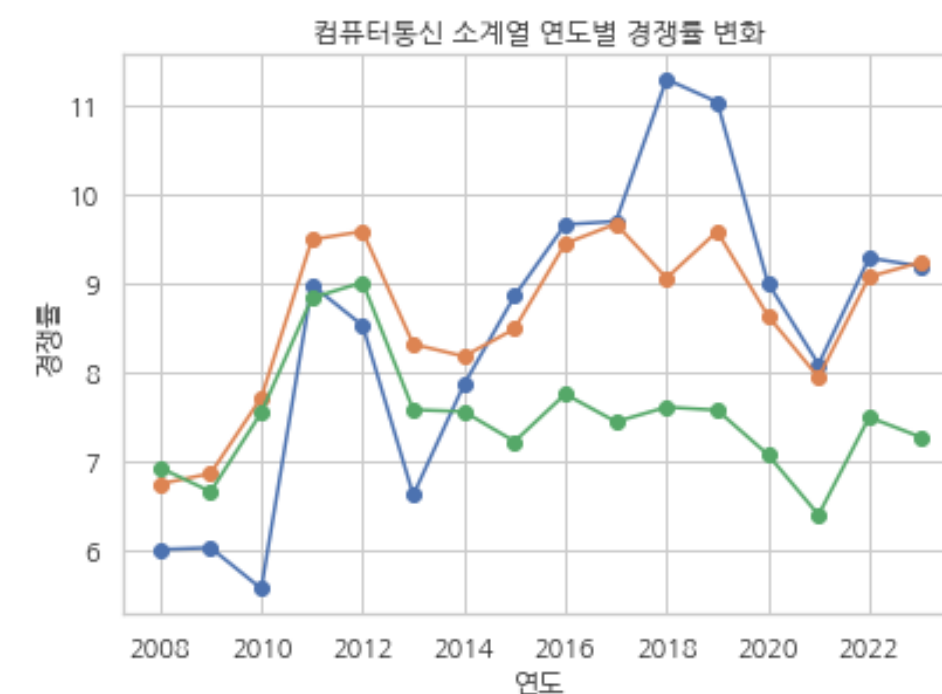
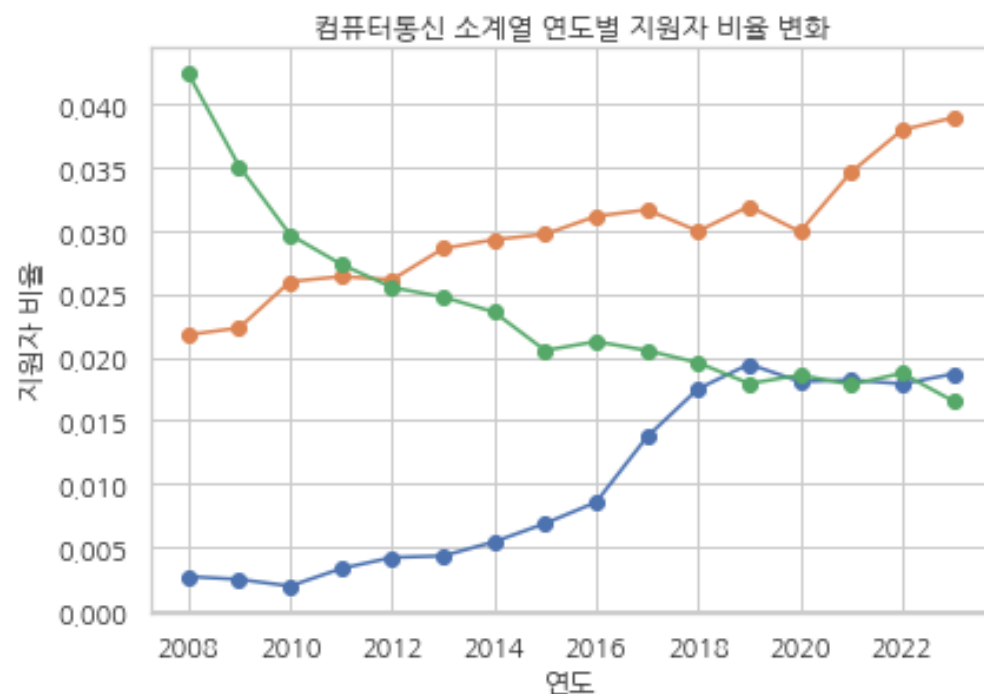
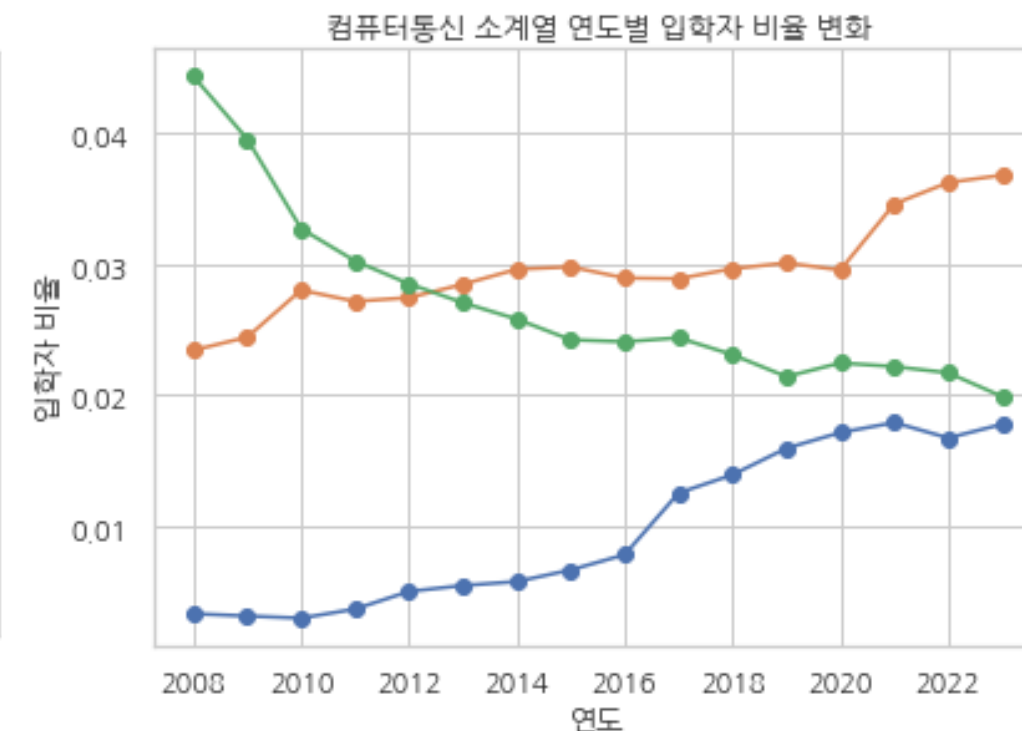
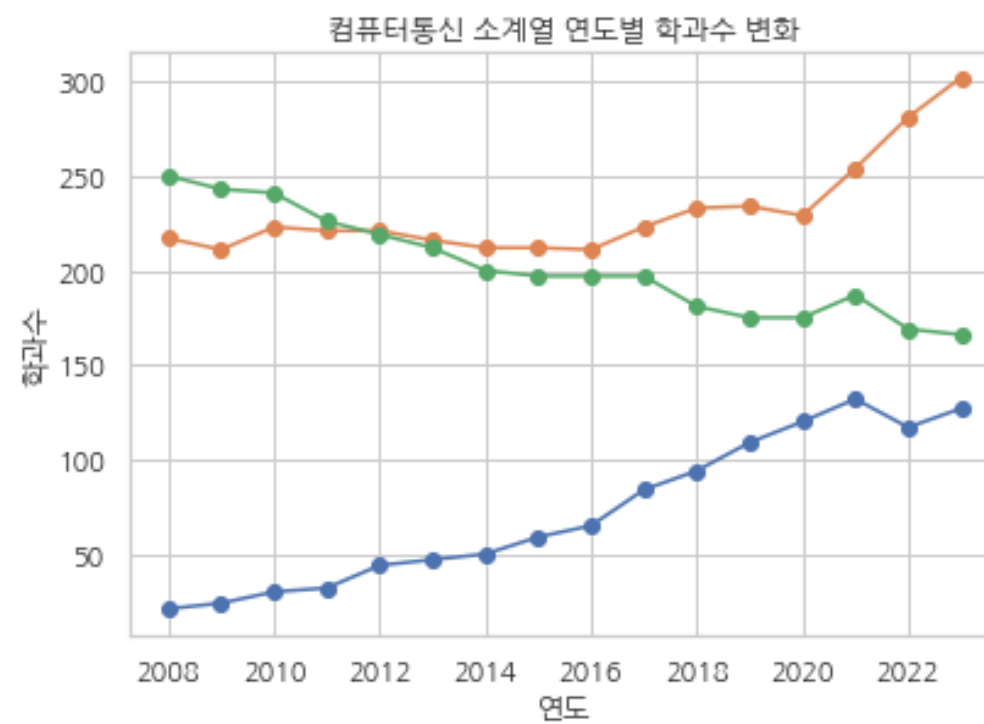
# 대계열별로 그래프 그리기
for idx, (비율_종류, ylabel) in enumerate([('학과수_전체', '학과수'), ('입학자_비율', '입학자 비율'), ('지원자_비율', '지원자 비율'), ('경쟁률', '경쟁률')]):
    ax = axes[idx]
    for 소계열, 그룹 in eng11_df.groupby('소계열'):
        ax.plot(그룹['연도'], 그룹[비율_종류], marker='o', label=소계열)

# 그래프에 제목과 축 레이블 추가
ax.set_title(f'컴퓨터통신 소계열 연도별 {ylabel} 변화')
ax.set_xlabel('연도')
ax.set_ylabel(ylabel)

# 범례 추가
ax.legend(title='소계열', bbox_to_anchor=(1, 1), loc='upper left')

# 그래프 간 간격 조절
plt.tight_layout()

# 그래프 표시
plt.show()
```



수행 과정 - 상세내용

데이터 분석

```
# 대계열, 중계열, 소계열별로 추세선 기울기 계산
```

```
for 대계열, 대계열그룹 in eng_df.groupby('대계열'):
    for 중계열, 중계열그룹 in 대계열그룹.groupby('중계열'):
        for 소계열, 그룹 in 중계열그룹.groupby('소계열'):
            x = np.array(그룹['연도']).reshape(-1, 1)
```

```
# 학과수비율 기울기 계산
```

```
y_학과수 = np.array(그룹['학과수_비율'])
model_학과수 = LinearRegression().fit(x, y_학과수)
slope_학과수 = model_학과수.coef_[0]
```

```
# 입학자비율 기울기 계산
```

```
y_입학자 = np.array(그룹['입학자_비율'])
model_입학자 = LinearRegression().fit(x, y_입학자)
slope_입학자 = model_입학자.coef_[0]
```

```
# 지원자비율 기울기 계산
```

```
y_지원자 = np.array(그룹['지원자_비율'])
model_지원자 = LinearRegression().fit(x, y_지원자)
slope_지원자 = model_지원자.coef_[0]
```

```
# 결과 저장
```

```
slope_dict['대계열'].append(대계열)
slope_dict['중계열'].append(중계열)
slope_dict['소계열'].append(소계열)
slope_dict['학과수비율기울기'].append(slope_학과수)
slope_dict['입학자비율기울기'].append(slope_입학자)
slope_dict['지원자비율기울기'].append(slope_지원자)
```

slope_df

	대계열	중계열	소계열	학과수비율기울기	입학자비율기울기	지원자비율기울기
0	공학계열	건축	건축·설비공학	-0.000137	2.310339e-06	-0.000027
1	공학계열	건축	건축학	-0.000100	8.346446e-05	0.000243
2	공학계열	건축	조경학	0.000057	-1.249324e-05	-0.000017
3	공학계열	교통·운송	지상교통공학	0.000075	9.087218e-05	0.000042
4	공학계열	교통·운송	항공학	0.000361	2.841390e-04	0.000119

수행 과정 - 상세내용

데이터 분석



**Amazon
QuickSight**

Amazon에서 제공하는 BI도구

- > 간단한 시각화
- > 풍부한 대시보드
- > 외부 데이터베이스 저장

수행 과정 - 상세내용

데이터 분석

데이터

데이터 세트 100%

SPIICE eng_df.csv

필드 검색

+ 계산된 필드

대계열

소계열

소계열_경쟁률

연도

입학자_비율

입학자_전체_계

중계열

지원자_비율

지원자_전체_계

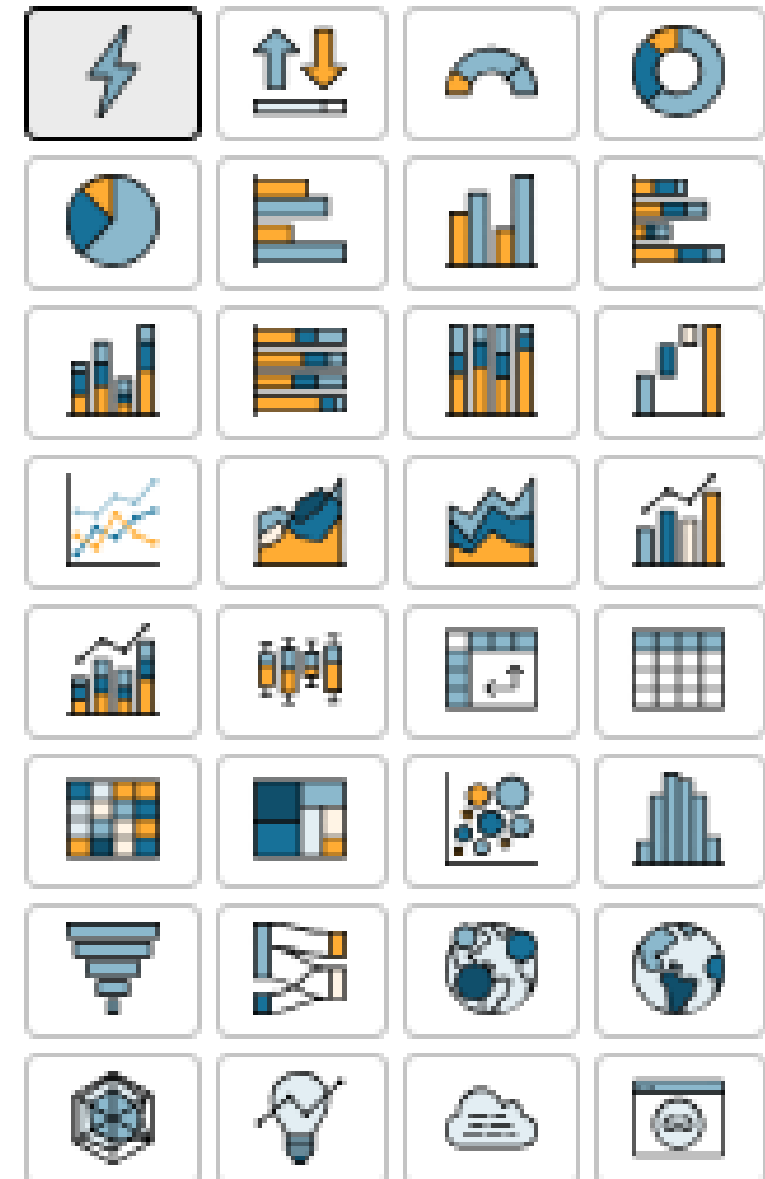
학과수_비율

학과수_전체

시각적 객체

+ 추가

시각적 객체 유형 변경



결과

기울기 높은 소계열 TOP3

학과수 비율 상승세

대계열	중계열	소계열	학과수비율기울기
공학계열	컴퓨터·통신	응용소프트웨어공학	0.00085
공학계열	기계·금속	기계공학	0.00047
자연계열	생활과학	식품영양학	0.00039

입학자 비율 상승세

대계열	중계열	소계열	입학자비율기울기
의약계열	간호	간호학	0,00137
공학계열	컴퓨터·통신	응용소프트웨어공학	0,00120
공학계열	컴퓨터·통신	전산학·컴퓨터공학	0,00069

지원자 비율 상승세

대계열	중계열	소계열	지원자비율기울기
의약계열	간호	간호학	0,00149
의약계열	의료	의학	0,00141
공학계열	컴퓨터·통신	응용소프트웨어공학	0,00140

기울기 낮은 소계열 TOP3

학과수 비율 상승세

대계열	중계열	소계열	학과수비율기울기
인문계열	언어·문학	영미어·문학	-0.00053
사회계열	법률	법학	-0.00077
공학계열	컴퓨터·통신	정보·통신공학	-0.00091

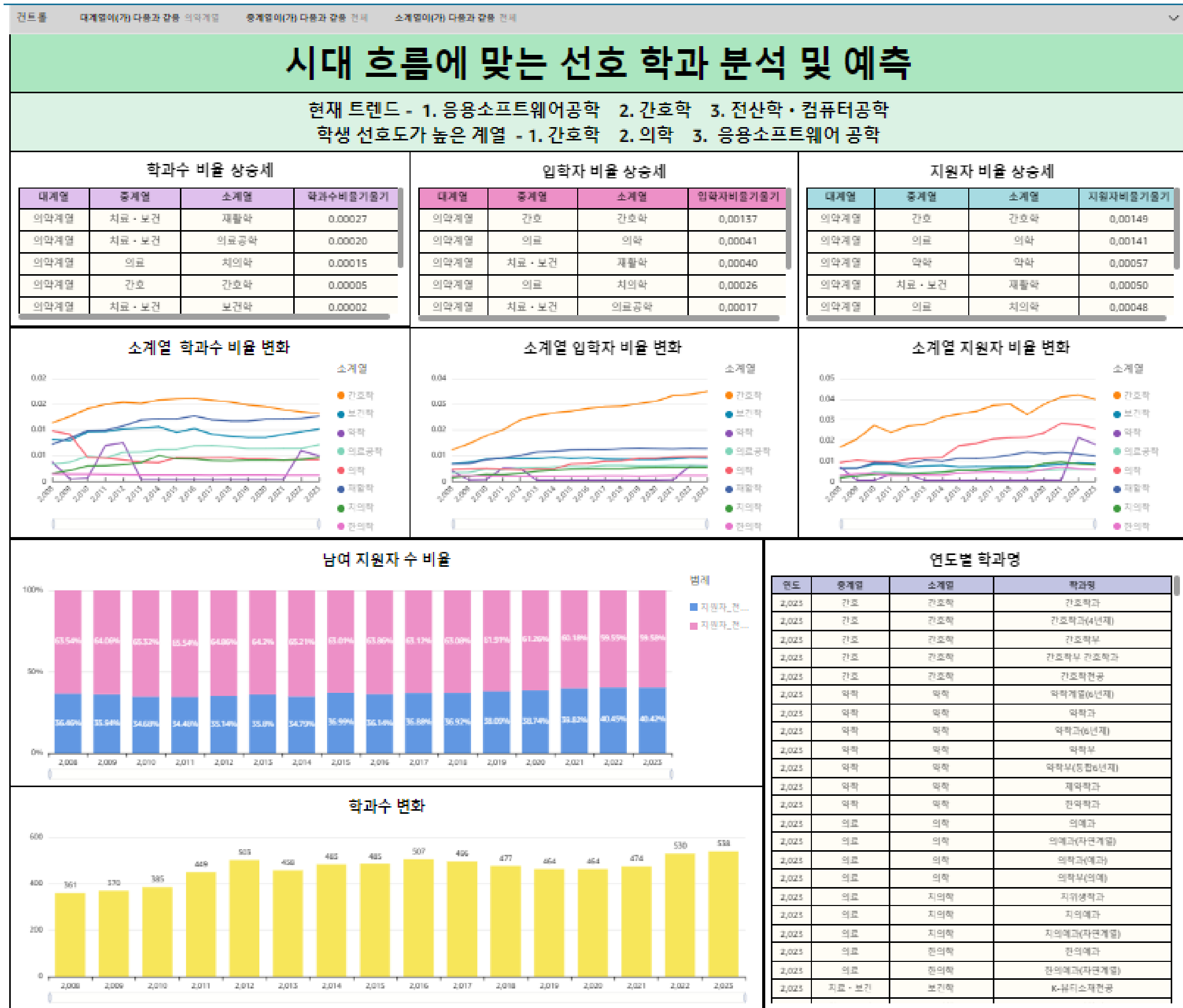
입학자 비율 상승세

대계열	중계열	소계열	입학자비율기울기
자연계열	화학·물리·생명과학	응용화학공학	-0,00081
인문계열	언어·문학	영미어·문학	-0,00067
사회계열	법률	법학	-0,00086
공학계열	컴퓨터·통신	정보·통신공학	-0,00126

지원자 비율 상승세

대계열	중계열	소계열	지원자비율기울기
사회계열	법률	법학	-0,00083
사회계열	사회과학	교양사회과학	-0,00126
공학계열	컴퓨터·통신	정보·통신공학	-0,00132
사회계열	경영·경제	경영학	-0,00173

결과



컨트롤

대계열이(가) 다음과 같음

중계열이(가) 다음과 같음

소계열이(가) 다음과 같음

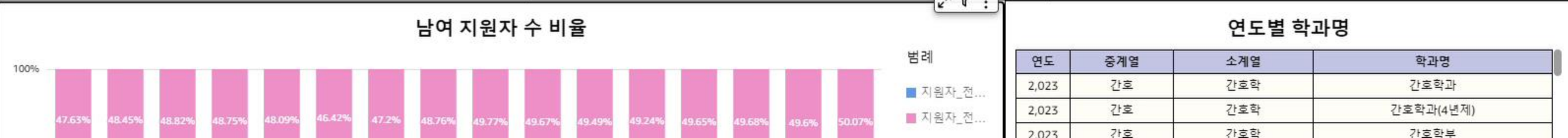
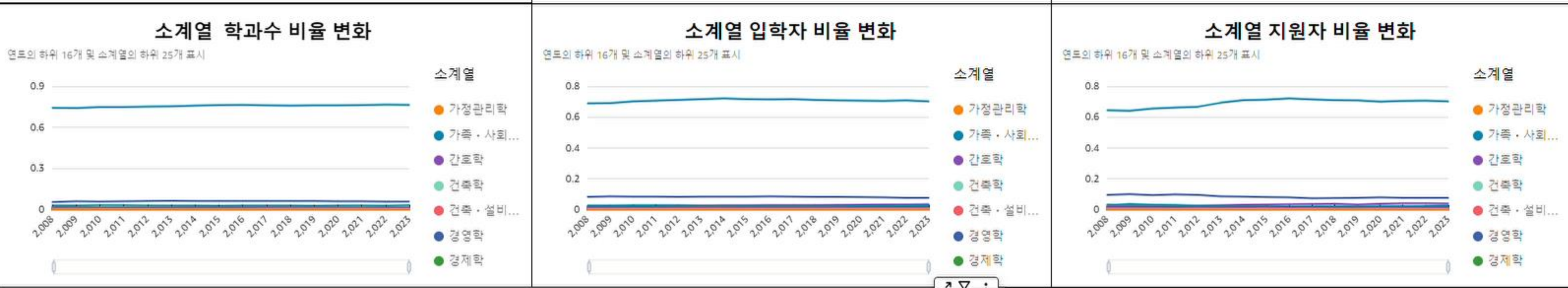
전체

전체

전체

현재 트렌드 - 1. 응용소프트웨어공학 2. 간호학 3. 전산학 · 컴퓨터공학
학생 선호도가 높은 계열 - 1. 간호학 2. 의학 3. 응용소프트웨어 공학

학과수 비율 상승세				입학자 비율 상승세				지원자 비율 상승세			
대계열	중계열	소계열	학과수비율기울기	대계열	중계열	소계열	입학자비율기울기	대계열	중계열	소계열	지원자비율기울기
공학계열	컴퓨터 · 통신	응용소프트웨어공학	0.00085	의약계열	간호	간호학	0,00137	의약계열	간호	간호학	0,00149
공학계열	기계 · 금속	기계공학	0.00047	공학계열	컴퓨터 · 통신	응용소프트웨어공학	0,00120	의약계열	의료	의학	0,00141
자연계열	생활과학	식품영양학	0.00039	공학계열	컴퓨터 · 통신	전산학 · 컴퓨터공학	0,00069	공학계열	컴퓨터 · 통신	응용소프트웨어공학	0,00140
공학계열	정밀 · 에너지	에너지공학	0.00039	공학계열	기타	교양공학	0,00065	공학계열	컴퓨터 · 통신	전산학 · 컴퓨터공학	0,00094
사회계열	사회과학	행정학	0.00036	공학계열	화공	화학공학	0,00043	공학계열	기타	교양공학	0,00077



기대효과 및 활용방안

- 해당 분야의 교육 자원 확대
- 학생들에게 최신 지식과 기술을 제공
- 학과를 개선하거나 새로운 학과를 개발

-> 학생들의 학습 만족도를 향상시키고 학생들이 학과를 선택하는 데 도움을 줄 것

감사합니다!
