



취업난

제목을 입력하십시오.
부제목도 입력하십시오.

취업? 난?!_4팀

Contents



01 / 주제 및 주제 선정 이유

02 / 데이터 및 모델 소개

- 로지스틱 회귀
- 랜덤 포레스트

03 / 기대효과

- 기대효과
- 향후 발전 가능성

주제 및 주제선정 이유

주제



최신뉴스

'청년 백수' 120만명 시대...취업해도 4명 중 1명꼴 단시간 근로

승고 2025-03-16 06:01

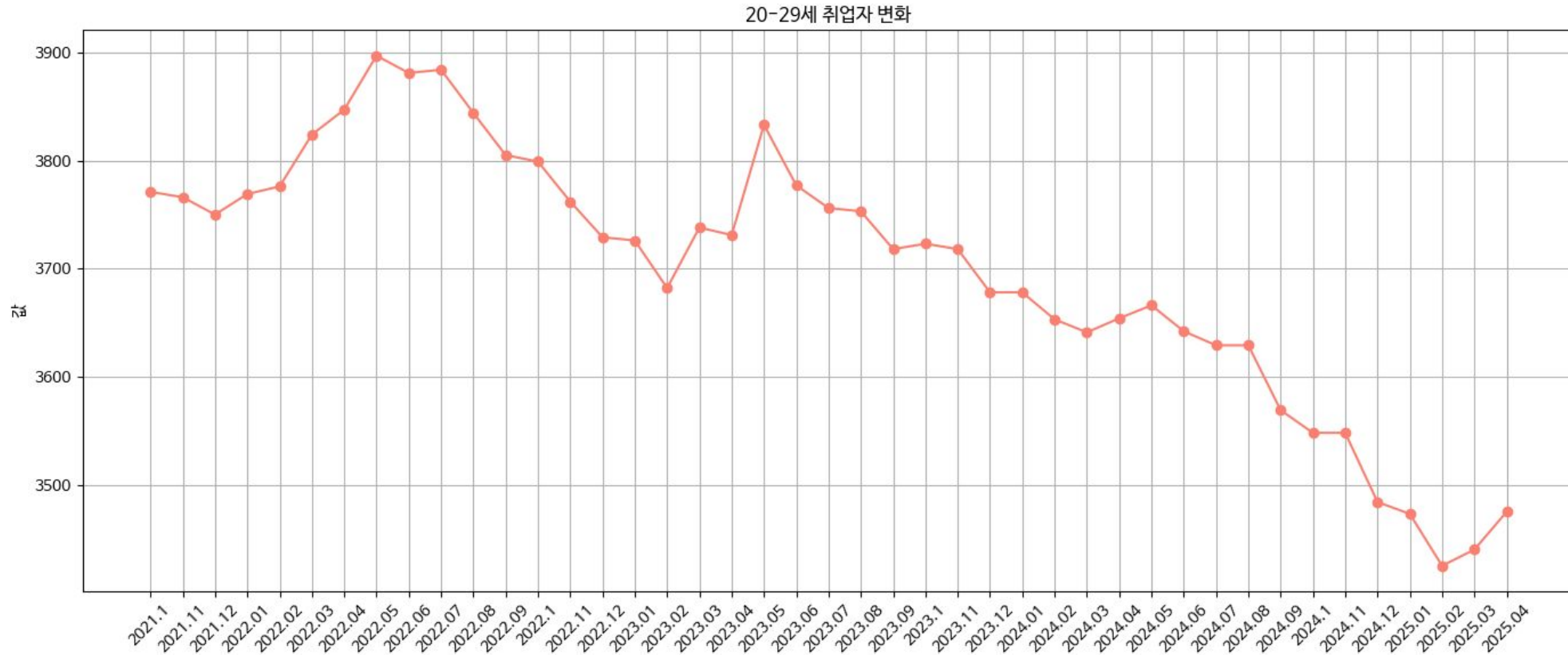


박재현 기자
+ 구독

쉬었음 50만명 '역대 최고'·실업자 27만명·취업준비자 43만명
어렵게 취업해도 전일제 아닌 '긱워커'...12만명은 "더 일하고 싶다"

기업들의 신규 채용에서 경력직 비중은 2009년 17.3%에서 2021년 37.6%로 두 배 이상 증가
출처 : 뉴스후플러스 (<https://www.newswhoplus.com>)

청년 고용률 현황



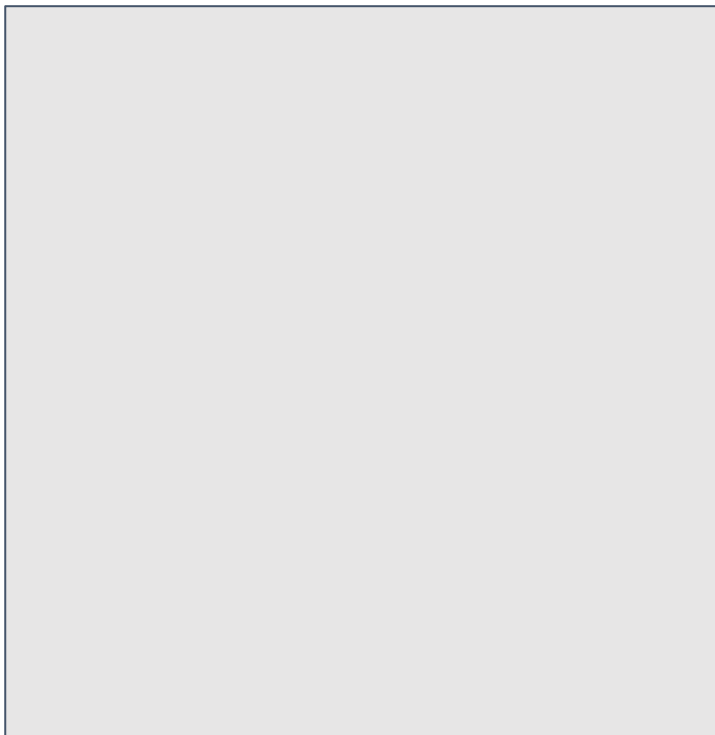
청년 취업난이 심화되고 있다.
2023년 하반기 이후 지속적인 하락세를 보인다

데이터 및 모델 소개

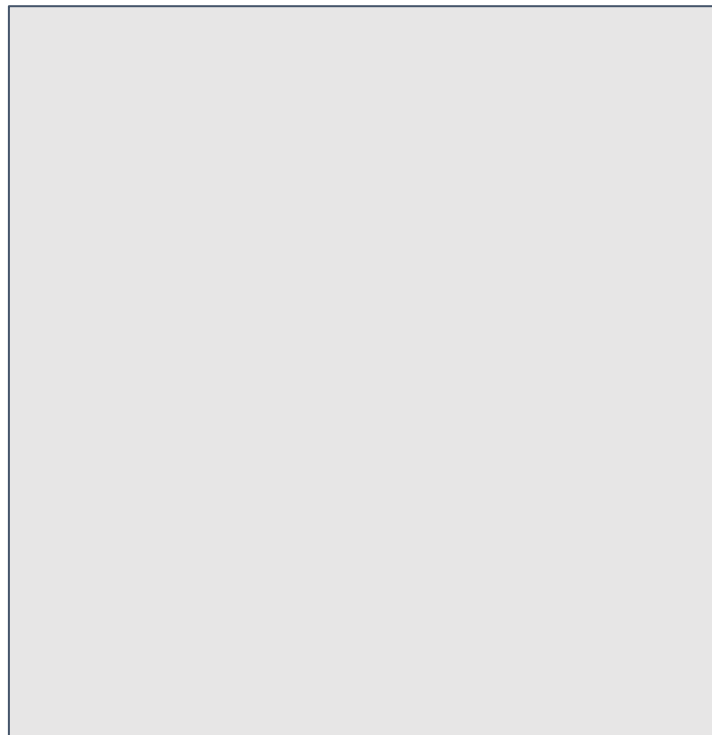
데이터 분석 및 시각화

분석: <- 데이터 전처리 끝나면 바로 진행

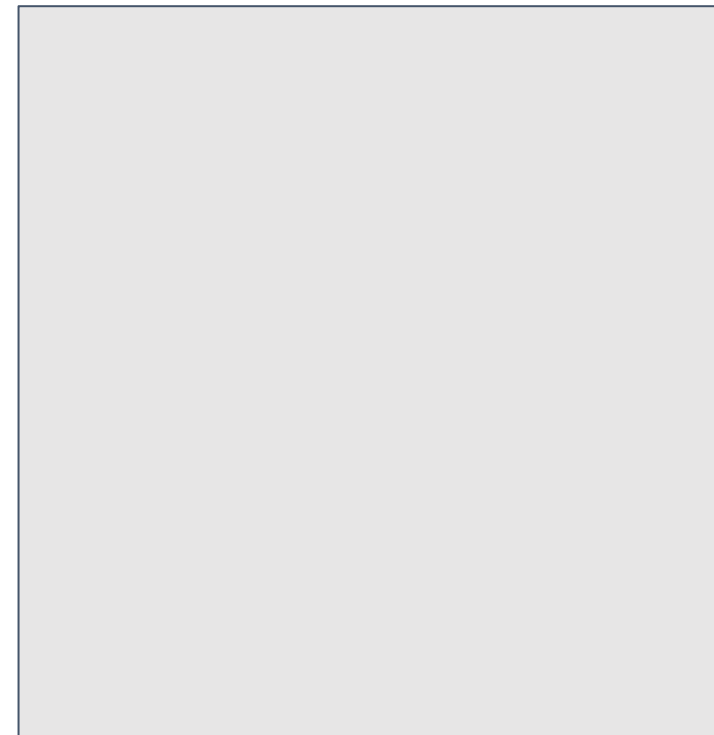
- 취업한 사람, 못한 사람 비교
- 취업한 사람들의 공통적인 특성
- 취업못한 사람들의 공통적인 특성



- 취업한 사람, 못한 사람 비교



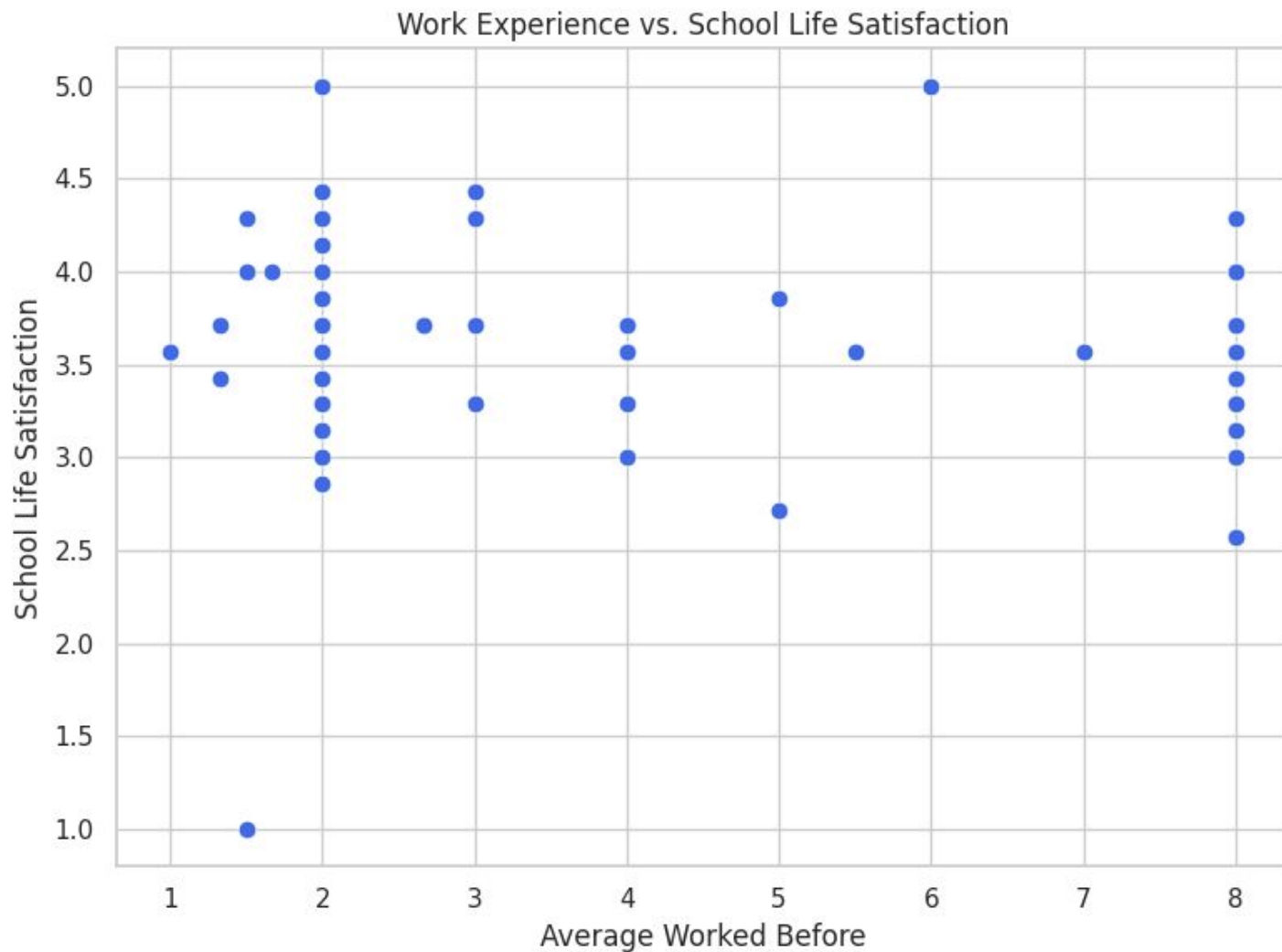
- 취업한 사람들의 공통적인 특성



- 취업못한 사람들의 공통적인 특성

데이터 분석 및 시각화

- 쓸지 안쓸지 모름
- 그냥 둘이 관계가 있을 거라고 생각했는데 관계가 없다.
- 사용한 자료 바탕으로



로지스틱 회귀

```
# 로지스틱 회귀
def fit_logistic_regression(self, **kwargs):
    self.lr_model = LogisticRegression(max_iter=200, **kwargs)
    self.lr_model.fit(self.X, self.y)
    importance = self.lr_model.coef_[0]
    return self.print_importance(importance, "Logistic Regression")
```

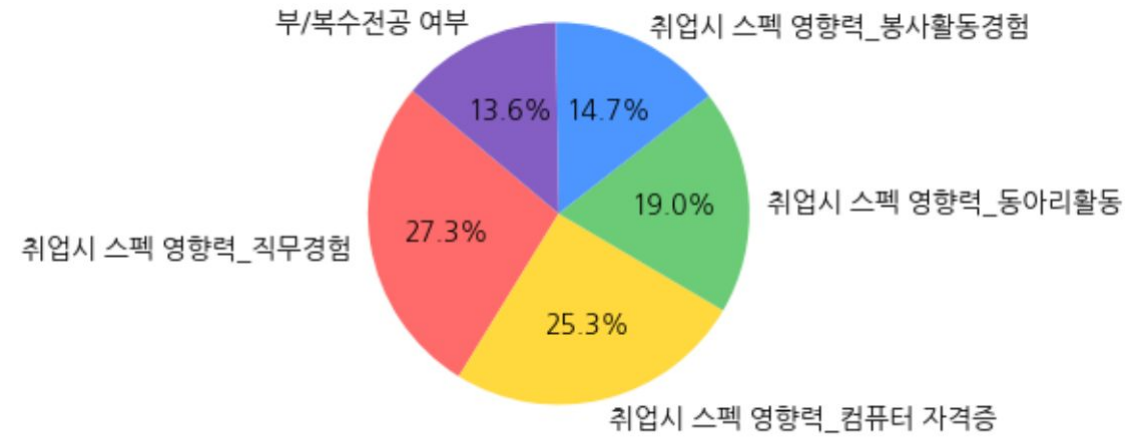
```
#로지스틱 회귀 하이퍼파라미터 튜닝
def tune_logistic_regression(self, param_grid=None, cv=5, scoring='accuracy'):
    if param_grid is None:
        param_grid = {
            'C': [0.01, 0.1, 1, 10],
            'penalty': ['l2'],
            'solver': ['lbfgs', 'liblinear'],
            'max_iter': [100, 200]
        }
    grid = GridSearchCV(LogisticRegression(), param_grid, cv=cv, scoring=scoring, n_jobs=-1)
    grid.fit(self.X, self.y)
    print("\n[Logistic Regression] Best Params:", grid.best_params_)
    print("[Logistic Regression] Best CV Score:", grid.best_score_)
    self.lr_model = grid.best_estimator_
    importance = self.lr_model.coef_[0]
    return self.print_importance(importance, "Logistic Regression (Tuned)")
```

```
[Logistic Regression] 중요도 TOP 5
y03c300      0.353341
y03c297      0.327131
y03c304      0.245633
y03c303      0.189937
y03a415      0.176283
dtype: float64
```

로지스틱 회귀의
회귀계수(coef) 값을
중요도로 판단

로지스틱 회귀

로지스틱 회귀 기반 중요도 TOP5



랜덤 포레스트

```
# 랜덤포레스트
def fit_random_forest(self, **kwargs):
    self.rf_model = RandomForestClassifier(n_estimators=100, random_state=42, **kwargs)
    self.rf_model.fit(self.X, self.y)
    importance = self.rf_model.feature_importances_
    return self.print_importance(importance, "Random Forest")
```

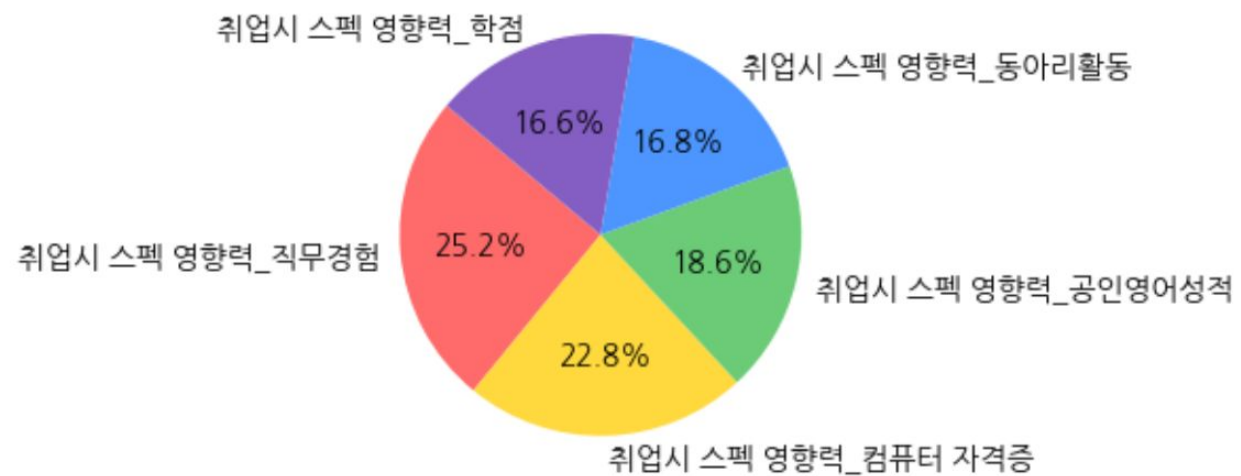
```
# 랜덤포레스트 하이퍼파라미터 튜닝
def tune_random_forest(self, param_grid=None, cv=5, scoring='accuracy'):
    if param_grid is None:
        param_grid = {
            'n_estimators': [100, 200],
            'max_depth': [None, 5, 10],
            'min_samples_split': [2, 5],
            'min_samples_leaf': [1, 2],
            'max_features': ['auto', 'sqrt']
        }
    grid = GridSearchCV(RandomForestClassifier(random_state=42), param_grid, cv=cv, scoring=scoring, n_jobs=-1)
    grid.fit(self.X, self.y)
    print("\n[Random Forest] Best Params:", grid.best_params_)
    print("[Random Forest] Best CV Score:", grid.best_score_)
    self.rf_model = grid.best_estimator_
    importance = self.rf_model.feature_importances_
    return self.print_importance(importance, "Random Forest (Tuned)")
```

```
[Random Forest] 중요도 TOP 5
y03c300    0.217122
y03c297    0.196390
y03c293    0.160610
y03c304    0.144600
y03c292    0.142649
dtype: float64
```

Random Forest 모델의
매개변수인
feature_importances_를
활용하여 중요도 측정

랜덤 포레스트

랜덤 포레스트 기반 중요도 TOP5



같은 전공 (계열) & 전공 내 취업자에서 활동 특이치

```
import pandas as pd
import numpy as np
from scipy.stats import ttest_ind
from statsmodels.stats.proportion import proportions_ztest

class HighEarningByMajorGroupAnalyzer:
    def __init__(self, df, salary_col, major_group_col, activity_cols, major_match_score_col, high_percentile=0.9):
        self.df = df
        self.salary_col = salary_col
        self.major_group_col = major_group_col
        self.activity_cols = activity_cols
        self.major_match_score_col = major_match_score_col # 일치도 점수 컬럼
        self.high_percentile = high_percentile
        self.results = []

    def _define_high_and_normal(self, sub):
        """고연봉자/일반자 집단 생성"""
        salary_z = (sub[self.salary_col] - sub[self.salary_col].mean()) / sub[self.salary_col].std()
        cutoff = salary_z.quantile(self.high_percentile)
        sub = sub.copy()
        sub['high_salary'] = (salary_z > cutoff).astype(int)
        high = sub[sub['high_salary'] == 1]
        normal = sub[sub['high_salary'] == 0]
        return high, normal

    def _compare_activities(self, high, normal):
        """각 활동별 고연봉자 vs 일반자 차이와 유의성 검증"""
        stats = {}
        for feat in self.activity_cols:
            try:
                mean_high = high[feat].mean()
                mean_normal = normal[feat].mean()
                diff = mean_high - mean_normal
                # 이진형 (0/1)
                if set(high[feat].dropna().unique()) <= {0,1} and set(normal[feat].dropna().unique()) <= {0,1}:
                    count = np.array([high[feat].sum(), normal[feat].sum()])
```

전공계열: 1.0 | 고연봉자:4 / 일반자:30

▶ 고연봉자에서만 유의하게 특이한 활동 (차이+유의성):

y03c303: 고연봉자평균=1.75, 일반평균=2.73, 차이=-0.98, p=0.085

y03c304: 고연봉자평균=1.75, 일반평균=2.73, 차이=-0.98, p=0.085

결과

사전 조사)

- 자신의 전공 분야로 취업하는 비율을 사전 조사.
- 활동 빈도수가 높은 활동을 '취준생이 생각하기에 중요한 활동'이라고 가정

데이터 분석)

1. 각 계열별 취업을 위한 활동 중요도 뽑아내기
2. 각 직무별 취업을 위한 활동 중요도 뽑아내기 => 빈도수 높은 feature 뽑기 , 로지스틱 회귀, 랜덤 포레스트
3. 각 전공 계열별 전공 내 취업자 내에서 특이치 잡아내기 => 전공 내 취업자만 다룸
4. 각 직무 별 같은 환경에서 특이치 잡아내기 => 전공 취업자와 비전공 취업자 따로 분석

[요약]

- 같은 환경 내에서 “고성과자”(고연봉자)를 특이치로 정의
- 활동 경험/스펙의 차이를 통계적, 머신러닝적으로 탐지
- “평균보다 연봉이 높은 사람 = 특정 활동 조합이 다르다”라는 인사이트 도출

결과

| | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | 6.0 | \ |
|---------|----------|----------|----------|----------|----------|----------|---|
| y03a415 | 4.751445 | 4.859873 | 4.902941 | 4.957055 | 4.979487 | 4.847826 | |
| y03c292 | 2.919355 | NaN | 2.953125 | 3.110390 | 2.721311 | 3.083333 | |
| y03c297 | NaN | 2.979381 | NaN | NaN | NaN | NaN | |
| y03c300 | 3.016129 | 3.170103 | 3.015625 | 3.298701 | 2.754098 | 3.250000 | |

| | 7.0 | 97.0 | 9090908.0 |
|---------|----------|------|-----------|
| y03a415 | 4.954545 | 5.0 | 5.0 |
| y03c292 | NaN | NaN | 5.0 |
| y03c297 | 2.744898 | 2.5 | NaN |
| y03c300 | 3.061224 | 2.5 | 5.0 |

[1.0] 상위 3개 활동:

| | |
|---------|----------|
| y03a415 | 4.751445 |
| y03c300 | 3.016129 |
| y03c292 | 2.919355 |

dtype: float64

[2.0] 상위 3개 활동:

| | |
|---------|----------|
| y03a415 | 4.859873 |
| y03c300 | 3.170103 |
| y03c297 | 2.979381 |

dtype: float64

[3.0] 상위 3개 활동:

| | |
|---------|----------|
| y03a415 | 4.902941 |
| y03c300 | 3.015625 |
| y03c292 | 2.953125 |

dtype: float64

[4.0] 상위 3개 활동:

| | |
|---------|----------|
| y03a415 | 4.957055 |
| y03c300 | 3.298701 |
| y03c292 | 3.110390 |

dtype: float64

[5.0] 상위 3개 활동:

| | |
|---------|----------|
| y03a415 | 4.979487 |
| y03c300 | 2.754098 |
| y03c292 | 2.721311 |

dtype: float64

[6.0] 상위 3개 활동:

| | |
|---------|----------|
| y03a415 | 4.847826 |
| y03c300 | 3.250000 |
| y03c292 | 3.083333 |

dtype: float64

[7.0] 상위 3개 활동:

| | |
|---------|----------|
| y03a415 | 4.954545 |
| y03c300 | 3.061224 |
| y03c297 | 2.744898 |

dtype: float64

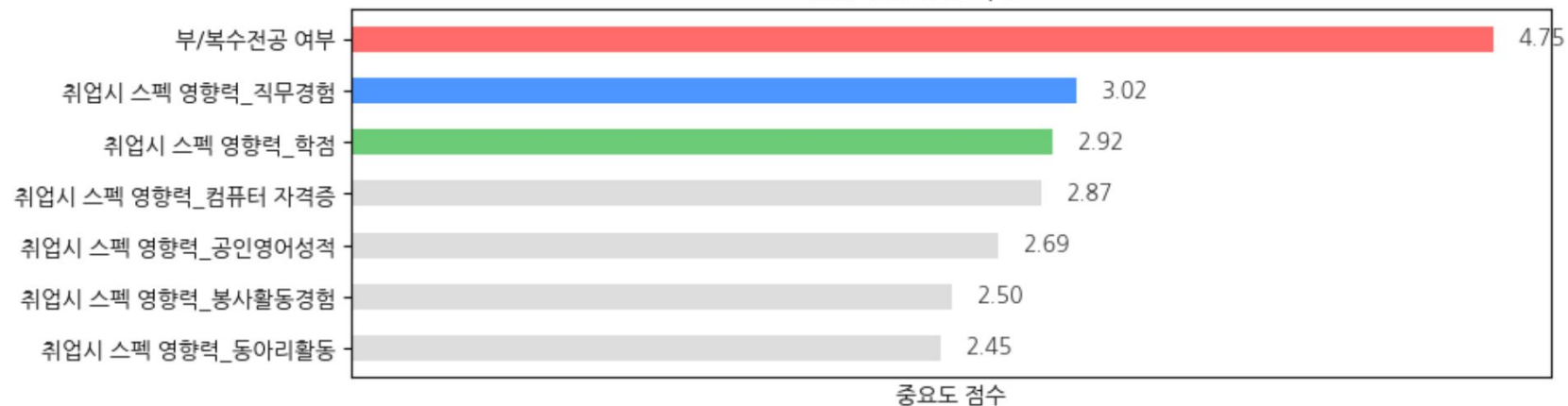
[97.0] 상위 3개 활동:

| | |
|---------|-----|
| y03a415 | 5.0 |
| y03c297 | 2.5 |
| y03c300 | 2.5 |

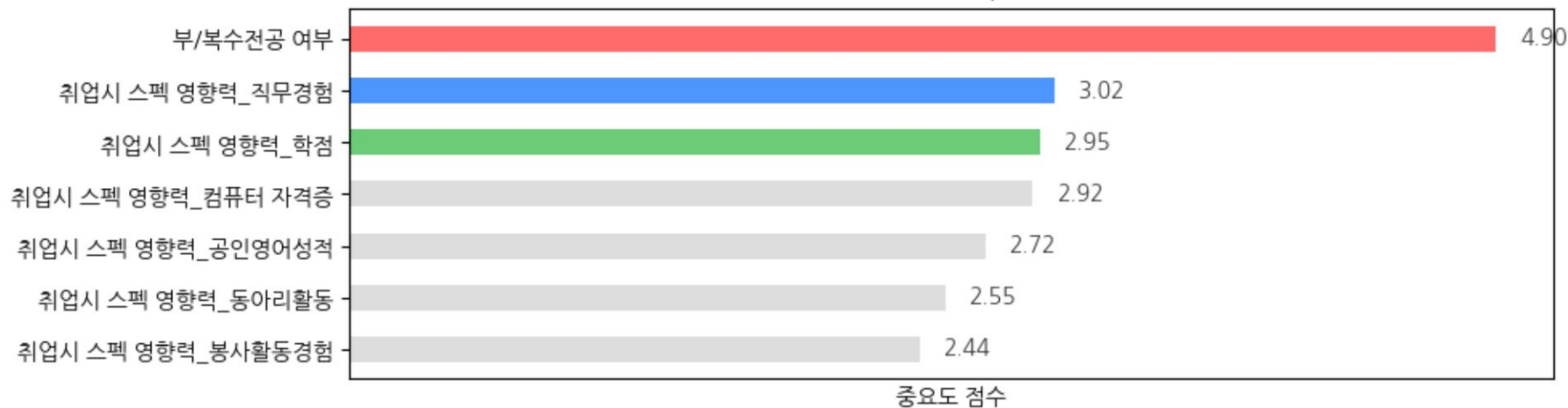
dtype: float64

데이터 분석 및 시각화

인문계열 취업 특성 TOP7

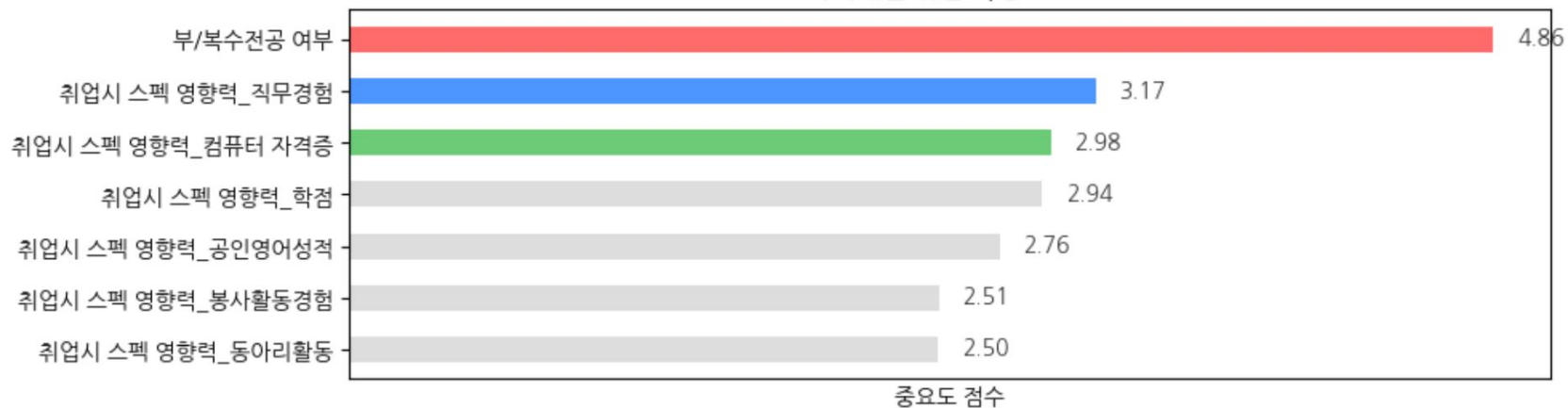


자연계열 취업 특성 TOP7

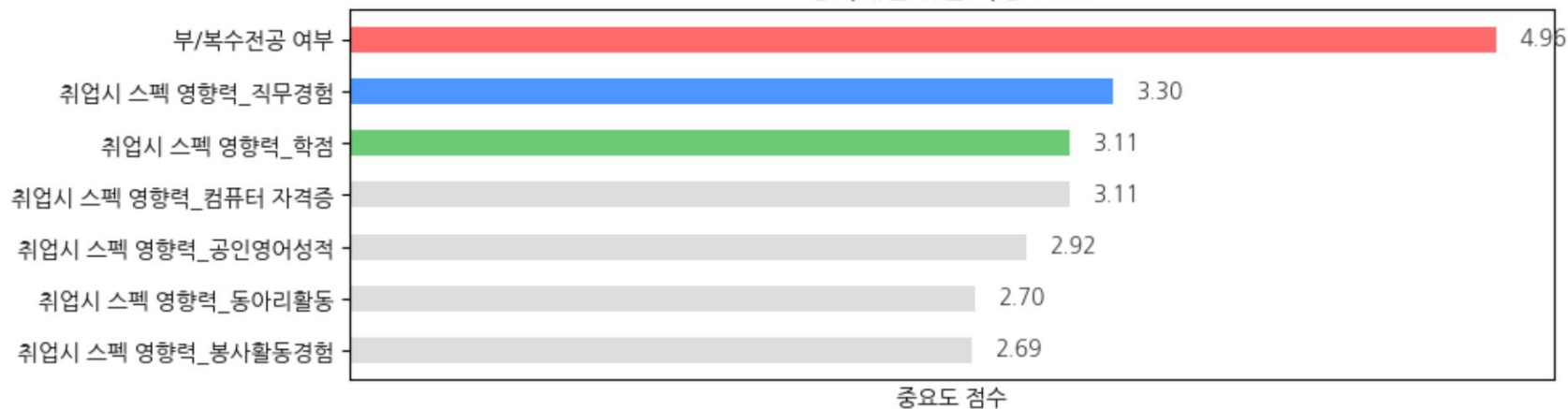


데이터 분석 및 시각화

사회계열 취업 특성 TOP7

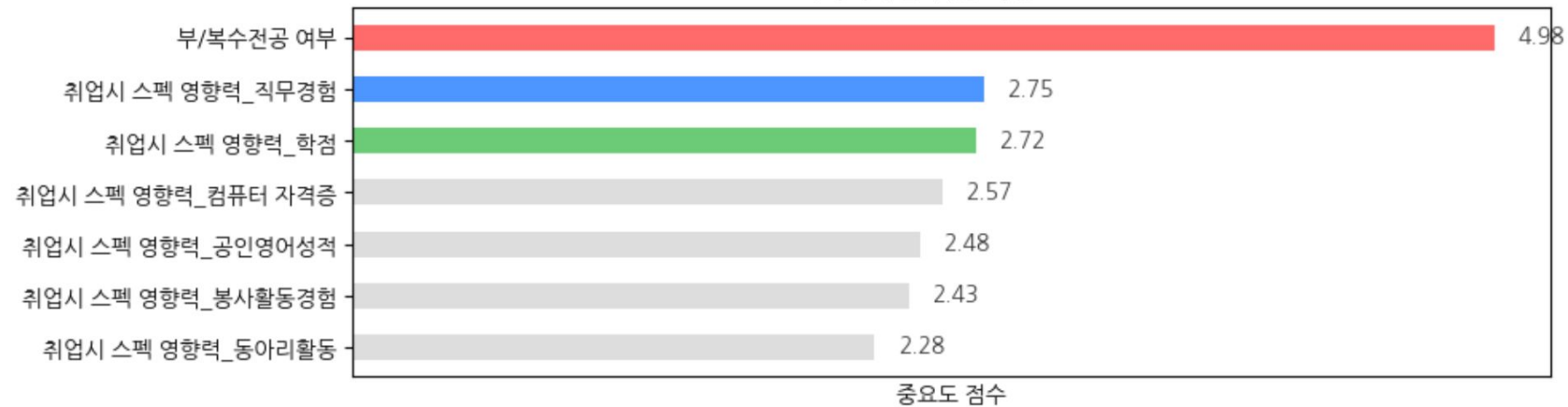


공학계열 취업 특성 TOP7

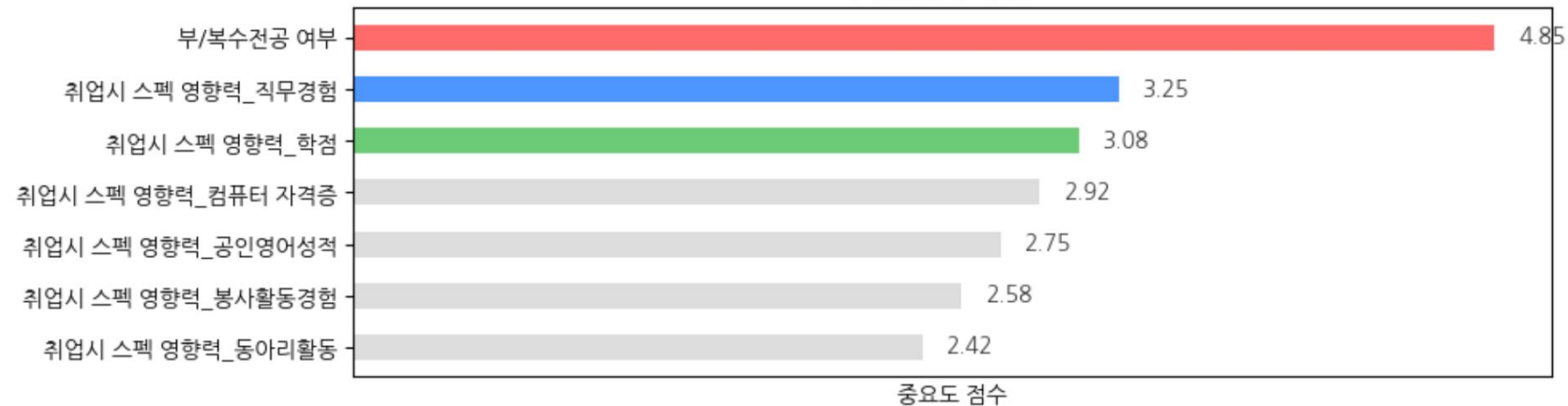


데이터 분석 및 시각화

의약학계열 취업 특성 TOP7

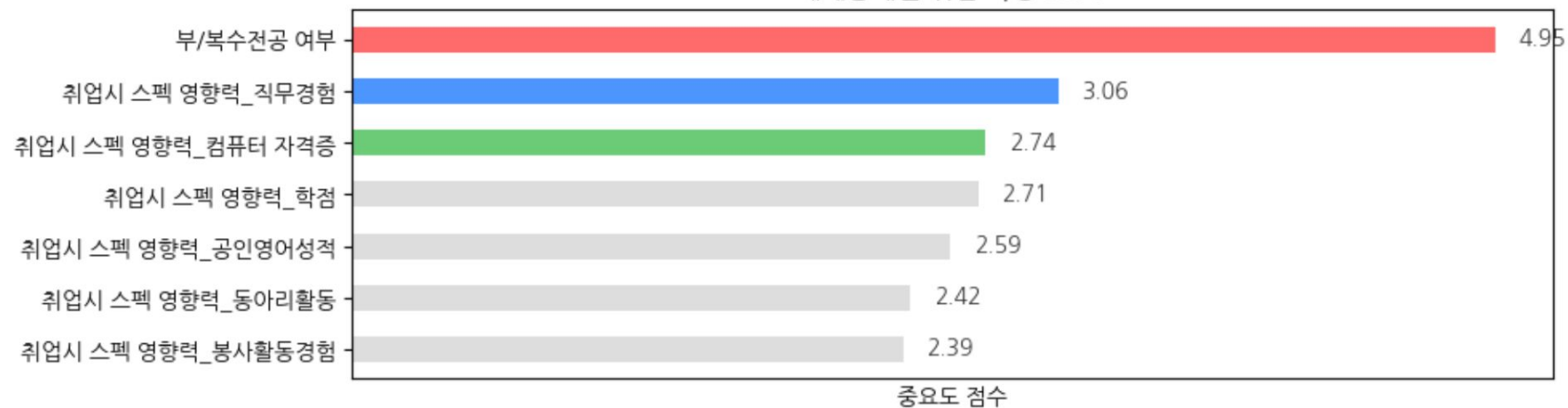


교육계열 취업 특성 TOP7

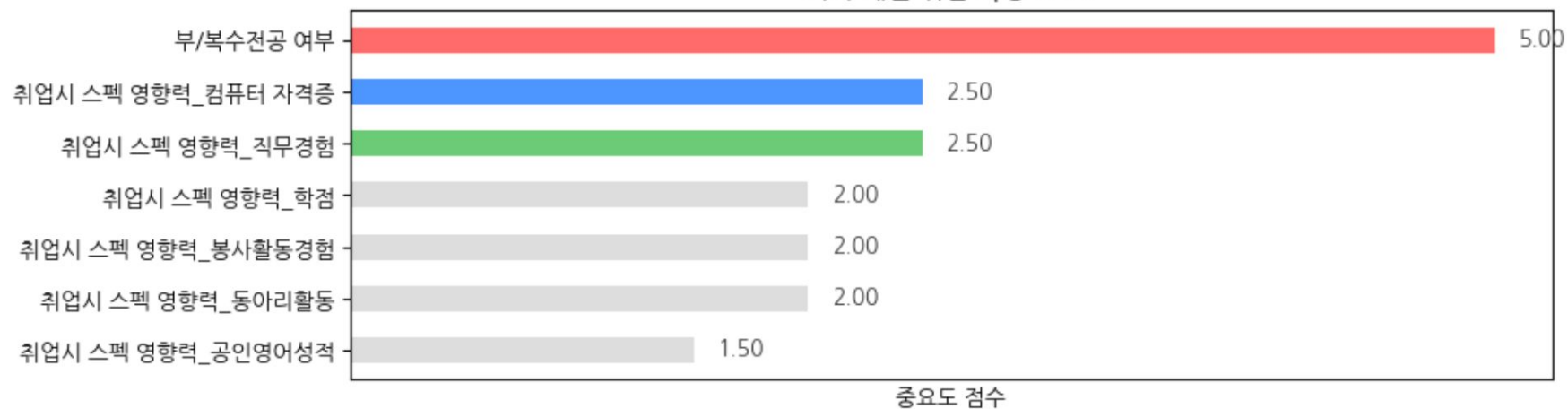


데이터 분석 및 시각화

예체능계열 취업 특성 TOP7



기타 계열 취업 특성 TOP7



기대효과

향후 발전 가능성



기대효과



개인 맞춤형 취업 전략 수립

데이터 기반으로 자신의 취업 성공 확률과 부족한 준비 요소를 객관적으로 파악할 수 있어 효율적인 취업 준비가 가능하다



정확한 자가 진단 및 동기 부여

구체적인 수치와 맞춤형 피드백을 제공해 사용자가 현재 위치와 앞으로 준비 방향을 인식할 수 있다



정책 및 교육 프로그램 개선

대학·정부·교육기관이 청년들의 실제 요구와 현황을 반영해 효과적인 지원책을 마련할 수 있도록 도울 수 있다.

KHODA

감사합니다

취업? 난?!_4팀

타이틀을 작성해주세요

분석 결과 기반이 아님

분석 결과 기반으로 바뀌어야 하지만 대충 틀은 짜둠

예상질문)

질문 1. 취업 컨설팅이 더 좋지 않느냐

질문 2. 취업에 무엇이 도움이 되었는지 설문한

결과가 있는데 왜 굳이 이게 필요한지

질문 3.

주제

타이틀에 맞는 세부내용을
핵심만 간략하게
설명해주세요



01

키워드를
외력해 주세요
키워드를 설명하는 내용을
작성해주세요

02

키워드를
외력해 주세요
키워드를 설명하는 내용을
작성해주세요

03

키워드를
외력해 주세요
키워드를 설명하는 내용을
작성해주세요
키워드를 설명하는 내용을
작성해주세요

타이틀을 작성해주세요

타이틀에 맞는 세부내용을 핵심만 간략하게
설명해주세요

핵심 키워드를 작성해주세요

- 핵심 키워드를 설명하는 세부 내용을 작성해주세요.
- 핵심 키워드를 설명하는 내용을 작성해주세요.

핵심 키워드를 작성해주세요

- 핵심 키워드를 설명하는 세부 내용을 작성해주세요.
- 핵심 키워드를 설명하는 내용을 작성해주세요.

Keyword

핵심 키워드를 작성해주세요

- 핵심 키워드를 설명하는 세부 내용을 작성해주세요.
- 핵심 키워드를 설명하는 내용을 작성해주세요.

핵심 키워드를 작성해주세요

- 핵심 키워드를 설명하는 세부 내용을 작성해주세요.
- 핵심 키워드를 설명하는 내용을 작성해주세요.

타이틀을 작성해주세요

타이틀에 맞는 세부내용을 핵심만 간략하게
설명해주세요

키워드 입력

키워드를 설명하는 내용을 입력해주세요
세부 내용을 입력해주세요

- 키워드에 맞는 내용을 입력해주세요
- 키워드에 맞는 내용을 입력해주세요
- 키워드에 맞는 내용을 입력해주세요

키워드 입력

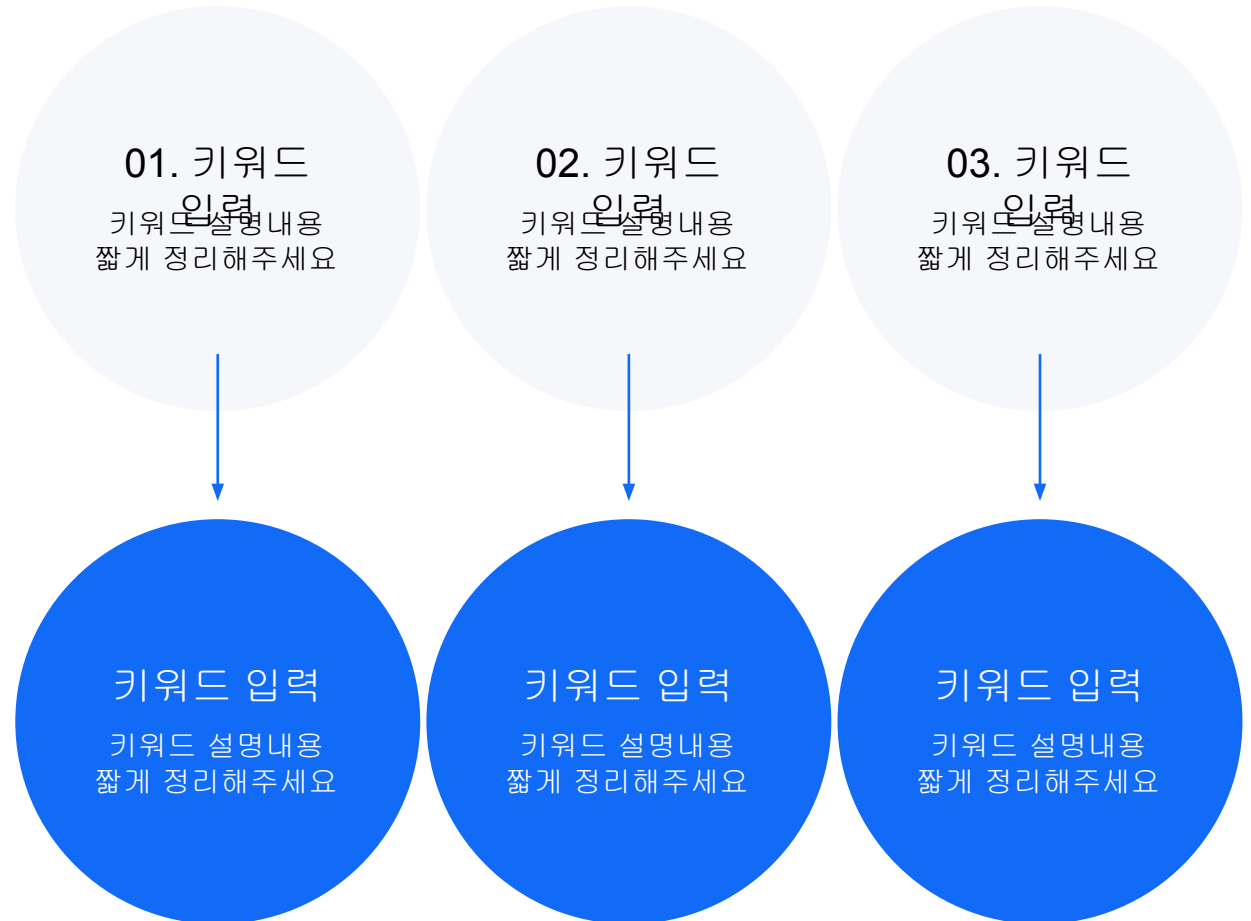
키워드를 설명하는 내용을 입력해주세요
세부 내용을 입력해주세요

- 키워드에 맞는 내용을 입력해주세요
- 키워드에 맞는 내용을 입력해주세요
- 키워드에 맞는 내용을 입력해주세요

타이틀을 작성해주세요

타이틀에 맞는 세부내용을
간략하게 설명해주세요

3가지로 뽑을 수 있는 포인트만
원 도형에 잘 넣어서
설명해주세요.



타이틀을 작성해주세요

타이틀에 맞는 세부내용을 핵심만 간략하게 설명해주세요



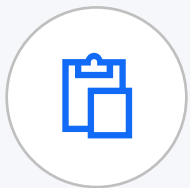
키워드를 입력해주세요

내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. **중요 내용을 칼라 포인트 해주세요.** 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요.



키워드를 입력해주세요

내용을 상세하게 입력해주세요. **중요 내용을 칼라 포인트 해주세요.** 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요.



키워드를 입력해주세요

중요 내용을 칼라 포인트 해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요. 내용을 상세하게 입력해주세요.

타이틀을 작성해주세요

타이틀에 맞는 세부내용을
핵심만 간략하게
설명해주세요

내용을 입력해주세요

“ 키워드를 설명하는 내용을 입력해주세요
세부 내용을 입력해주세요 ”

Image here

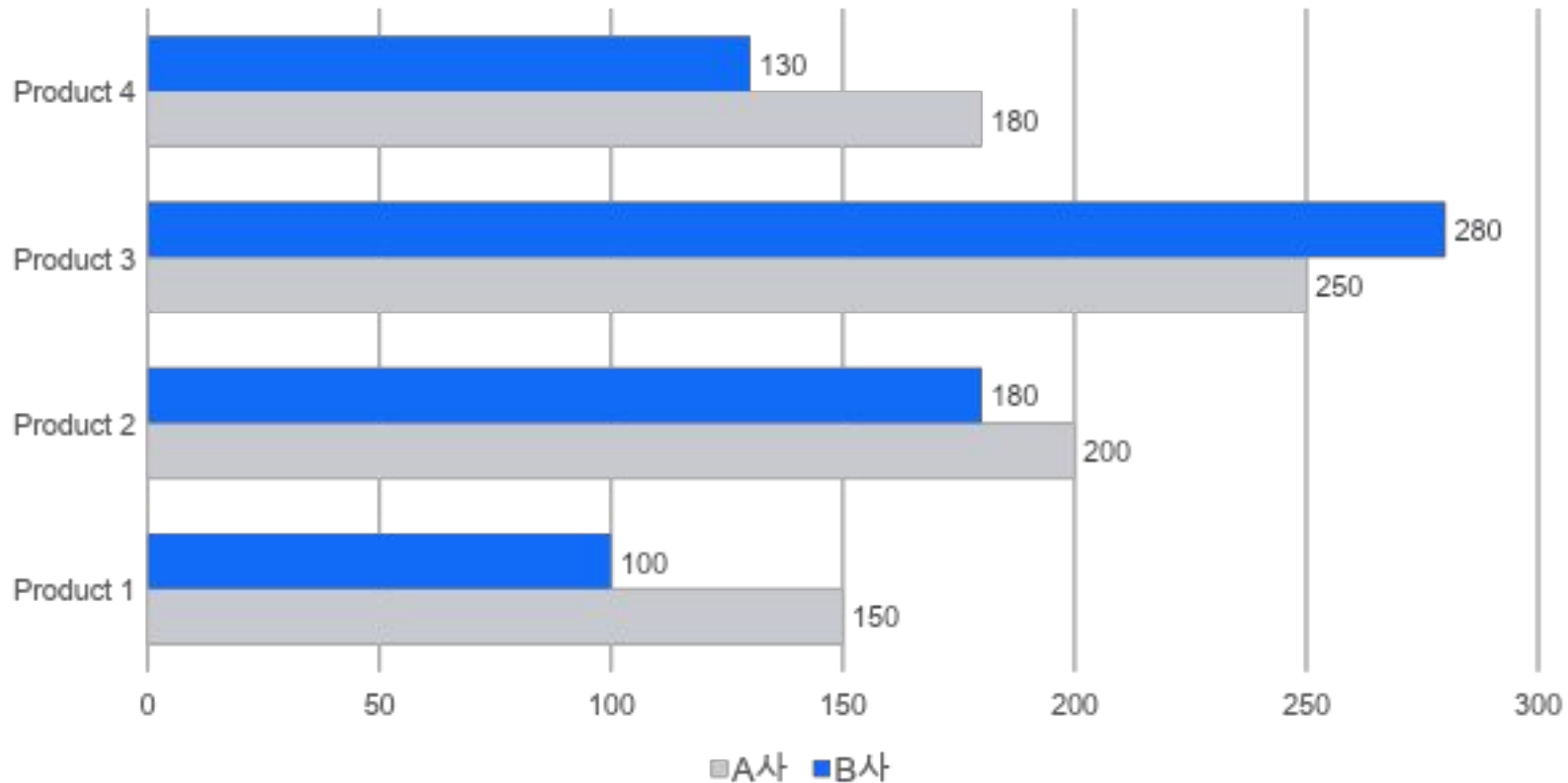
내용을 입력해주세요

“ 키워드를 설명하는 내용을 입력해주세요
세부 내용을 입력해주세요 ”

Image here

타이틀을 작성해주세요

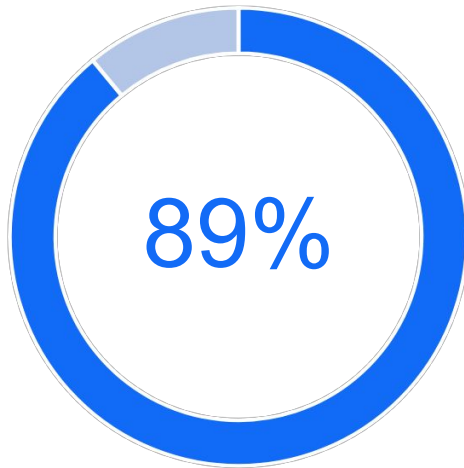
타이틀에 맞는 세부내용을 핵심만 간략하게
설명해주세요



타이틀을 작성해주세요

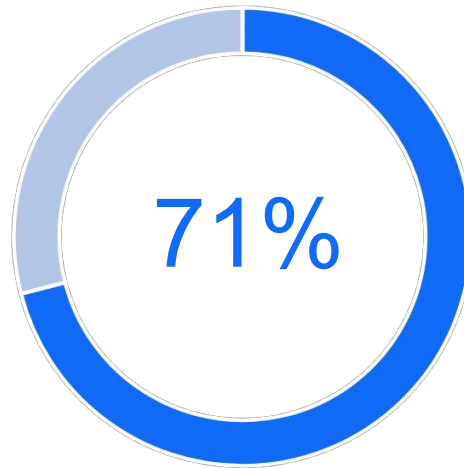
타이틀에 맞는 세부내용을 핵심만 간략하게
설명해주세요

키워드 입력



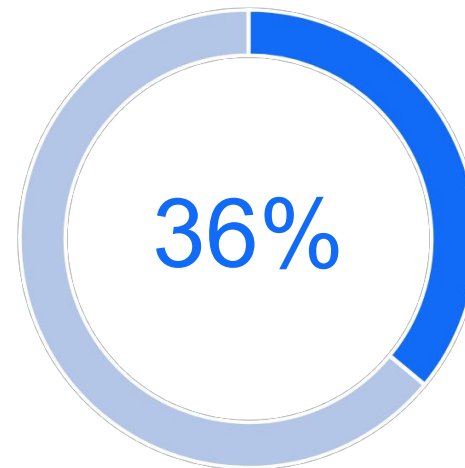
세부내용 작성해주세요
내용 작성해주세요

키워드 입력



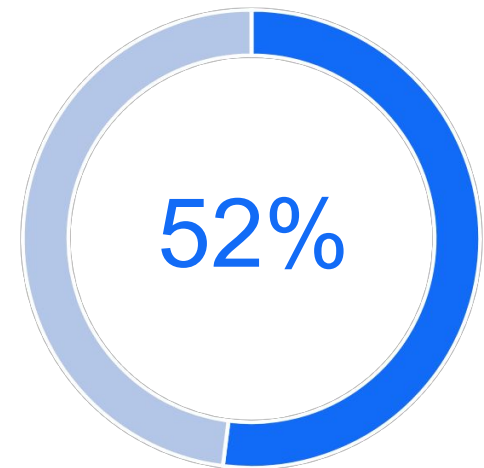
세부내용 작성해주세요
내용 작성해주세요

키워드 입력



세부내용 작성해주세요
내용 작성해주세요

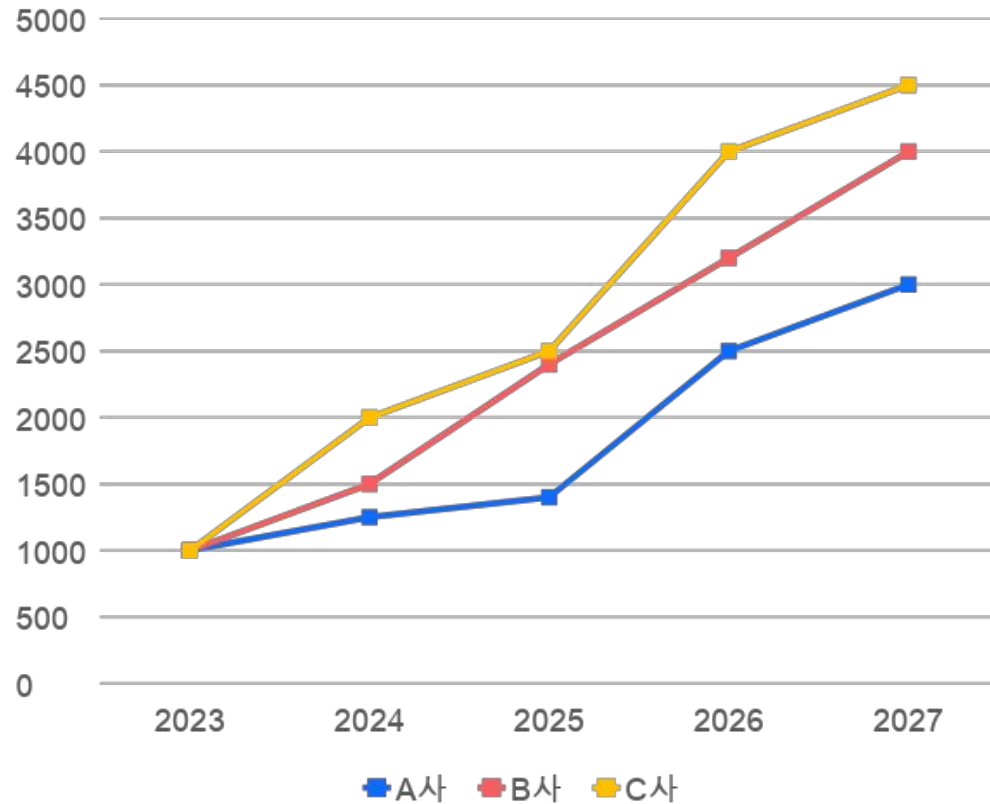
키워드 입력



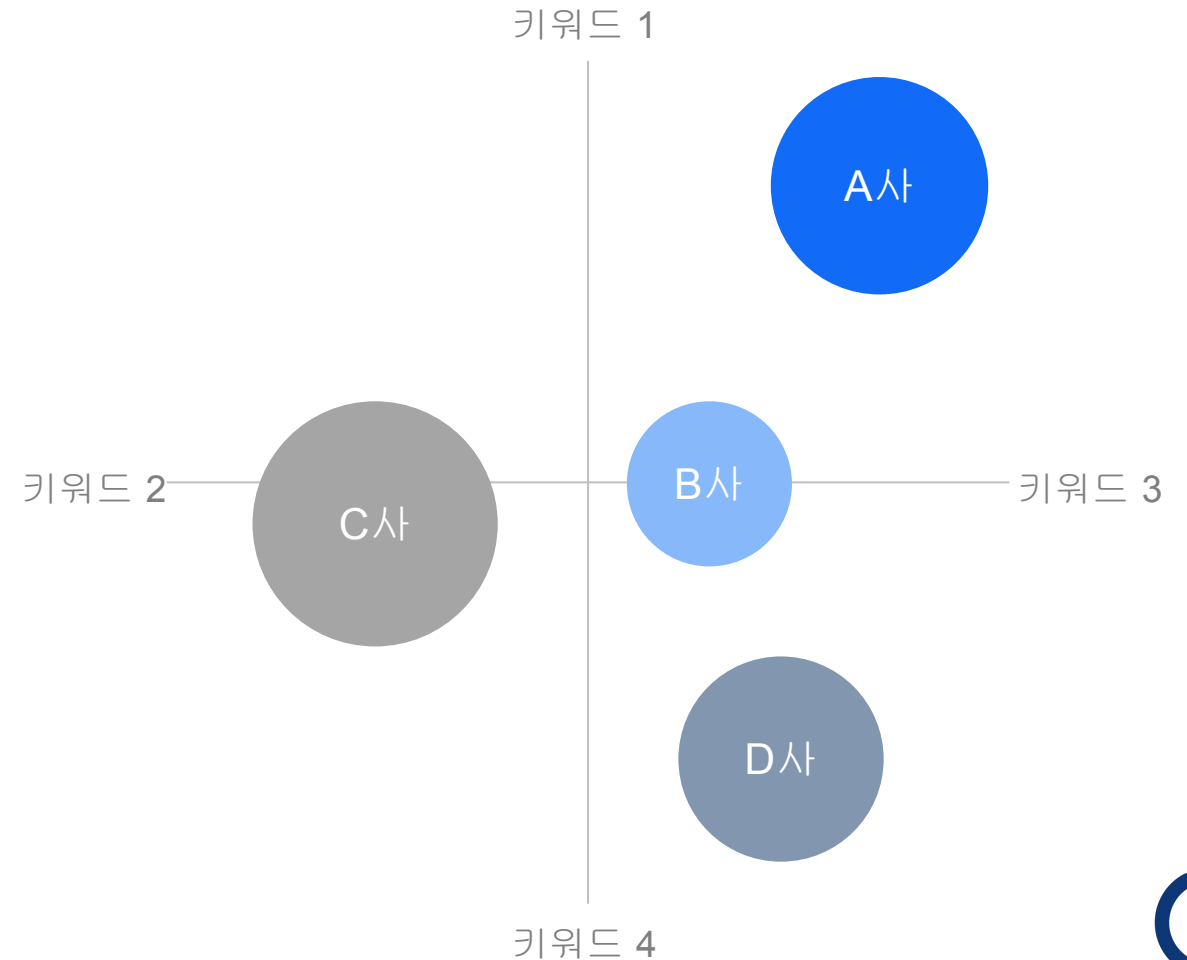
세부내용 작성해주세요
내용 작성해주세요

타이틀을 작성해주세요

내용을 입력해주세요



내용을 입력해주세요



타이틀을 작성해주세요

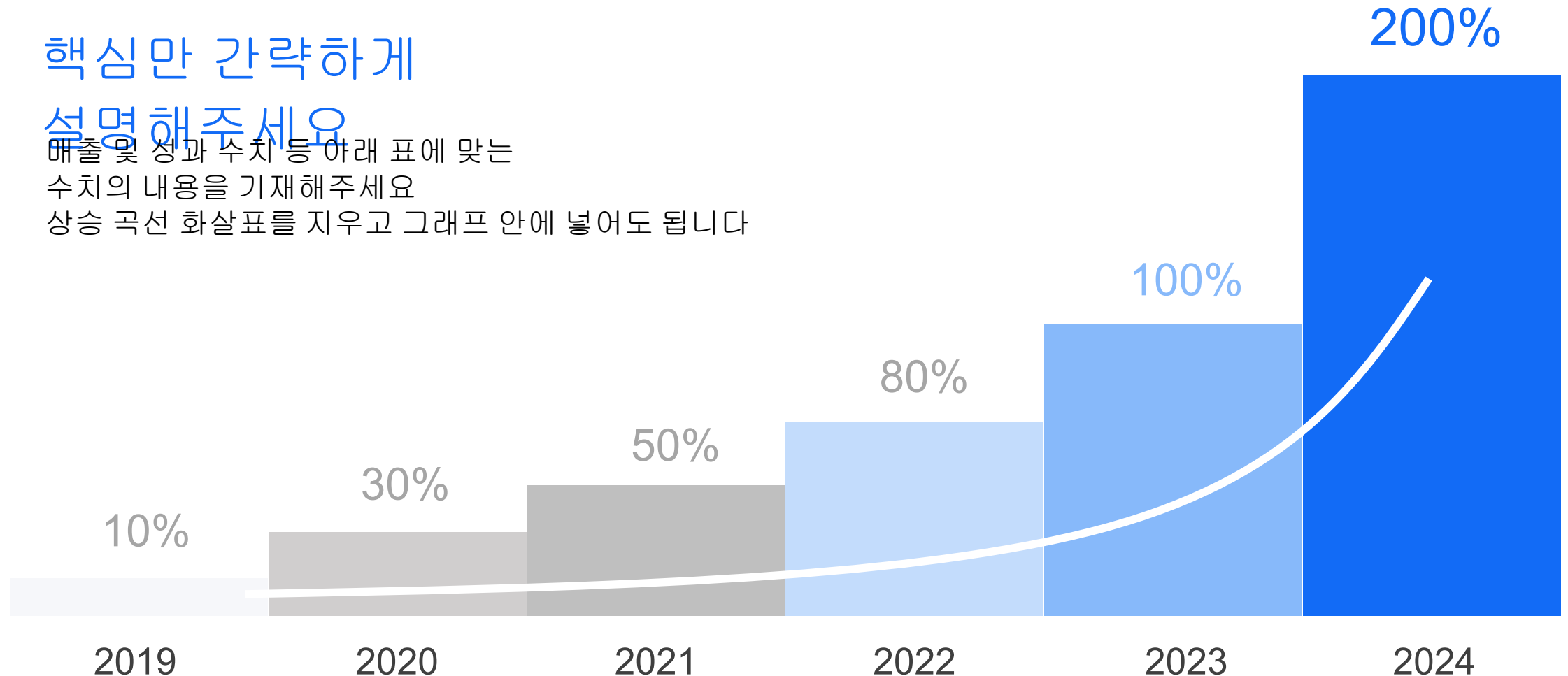
타이틀에 맞는 세부내용을

핵심만 간략하게

설명해주세요

매출 및 성과 수치 등 아래 표에 맞는
수치의 내용을 기재해주세요

상승 곡선 화살표를 지우고 그래프 안에 넣어도 됩니다



타이틀을 작성해주세요

타이틀에 맞는 세부내용을 핵심만 간략하게
설명해주세요



히스토리 내용을
입력해주세요
내용을 입력해주세요
2024



히스토리 내용을
입력해주세요
내용을 입력해주세요
2026



히스토리 내용을
입력해주세요
내용을 입력해주세요
2028

2023

히스토리 내용을
입력해주세요
내용을 입력해주세요



2025

히스토리 내용을
입력해주세요
내용을 입력해주세요



2027

히스토리 내용을
입력해주세요
내용을 입력해주세요



타이틀을 작성해주세요

타이틀에 맞는 세부내용을 핵심만 간략하게
설명해주세요



JASON KIM
CEO

- 그동안의 이력을 잘 정리해서 나열해주세요.



EMILY KIM
COO

- 그동안의 이력을 잘 정리해서 나열해주세요.



JASON KIM
CMO

- 그동안의 이력을 잘 정리해서 나열해주세요.



EMILY KIM
CTO

- 그동안의 이력을 잘 정리해서 나열해주세요.

타이틀을 작성해주세요

타이틀에 맞는 세부내용을 핵심만 간략하게
설명해주세요