

R E P O R T

[대학별 기업규모의 따른 취업자 수]



수 강 과 목 :	빅데이터 처리
학 과 :	컴퓨터정보과
학 번 :	202144094
이 름 :	최정현
제 출 일 :	2023-11-27

< 목 차 >

1. 서론	2
1.1 배경	2
1.2 목적	2
2. 본론	3
2.1 데이터 수집 방법	3
2.2 데이터 전처리	3
2.3 데이터 시각화	
3. 결론	8
3.1 연구 요약	8
4. 참고문헌	8
5. github 주소	8

1.1 배경

현대 사회에서는 대학 졸업자들의 취업 환경이 더욱 다양해지고 있는데, 이는 다양한 기업 규모에서의 취업 기회가 확장되고 있다는 것을 의미합니다. 특히, 기술과 데이터의 중요성이 증대되면서 빅데이터 분석은 기업들의 채용 프로세스에서 핵심적인 역할을 하고 있습니다.

본 연구는 이러한 취업 동향의 변화 속에서 대학 졸업자들이 어떻게 기업 규모를 고려하며 취업 기회를 모색하는지에 대한 심층적인 이해를 목표로 합니다. 빅데이터 처리 기술을 활용하여 수집된 데이터를 토대로, 대학별로 기업규모에 따른 취업자 수의 패턴과 동향을 분석하여 그 결과를 정량적으로 제시할 것입니다.

1.2 목적

첫째, 대학 졸업자들이 선호하는 기업 규모를 파악하여, 학생들이 미래 취업을 고려할 때 어떤 기업을 주목해야 하는지에 대한 통찰력을 얻고자 합니다.

둘째, 우리는 빅데이터 분석을 통해 얻은 정보를 활용하여, 대학과 기업 간의 협력을 증진시키고 학문과 산업 간의 유기적인 연결을 지원하고자 합니다.

2.1 데이터 수집 방법

데이터 수집은 공공데이터 포털에서 제공되는 데이터를 활용했습니다. 해당 포털은 정부에서 제공하는 공공데이터를 효율적으로 검색하고 활용할 수 있는 플랫폼으로, 다양한 분야의 데이터를 수집할 수 있습니다.

수집된 데이터는 content.csv형식으로 제공되었으며, 각 행은 대학에 대한 정보를 담고 있습니다. 각 열은 학제, 지역, 분석 대상자 수, 그리고 다양한 기업 규모 범주에 대한 취업자 수입니다.

공공데이터포털 : <https://www.data.go.kr/data/15071025/fileData.do>

2.2 데이터 전처리

데이터 필터링 및 집계 : `data[data.sum(axis=1) > 0]`를 통해 취업자 수가 0인 행을 제거하고, 그 후 `groupby(['학제', '지역']).sum().sum(axis=1)`를 통해 필터링된 데이터를 그룹화하고 취업자 수를 통합하여 시리즈로 얻었습니다.

데이터 변환 과정 : 해당 과정에서 데이터를 재구조화하여 시각화 형태로 변환하고 있습니다. 데이터 수집, 전처리, 가공 모두 하나의 과정으로 볼 수 있습니다. 데이터가 과도하게 많아 세분화 되어있는 데이터들은 직접 CSV파일에서 데이터를 제거하여 시각적으로 보기 쉽게 변환하였습니다.

2.3 데이터 시각화

- 학제와 지역에 따른 취업자 수 통합 비율

기업규모에 따른 분석대상자 수를 파이 차트로 나타내어, 소규모, 중규모, 대규모 기업에서의 취업자 비율을 확인할 수 있습니다.

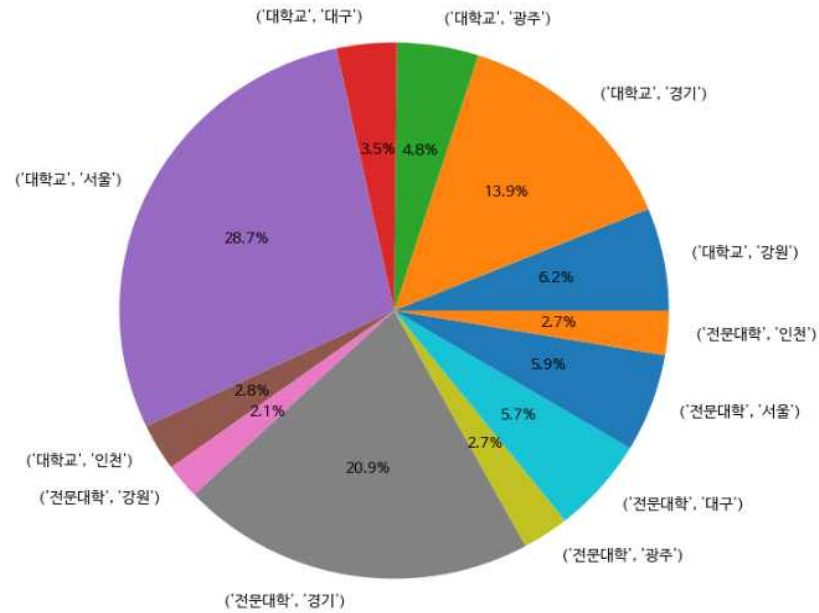
```
# Pie chart
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('/content/content.csv', encoding='euc-kr')

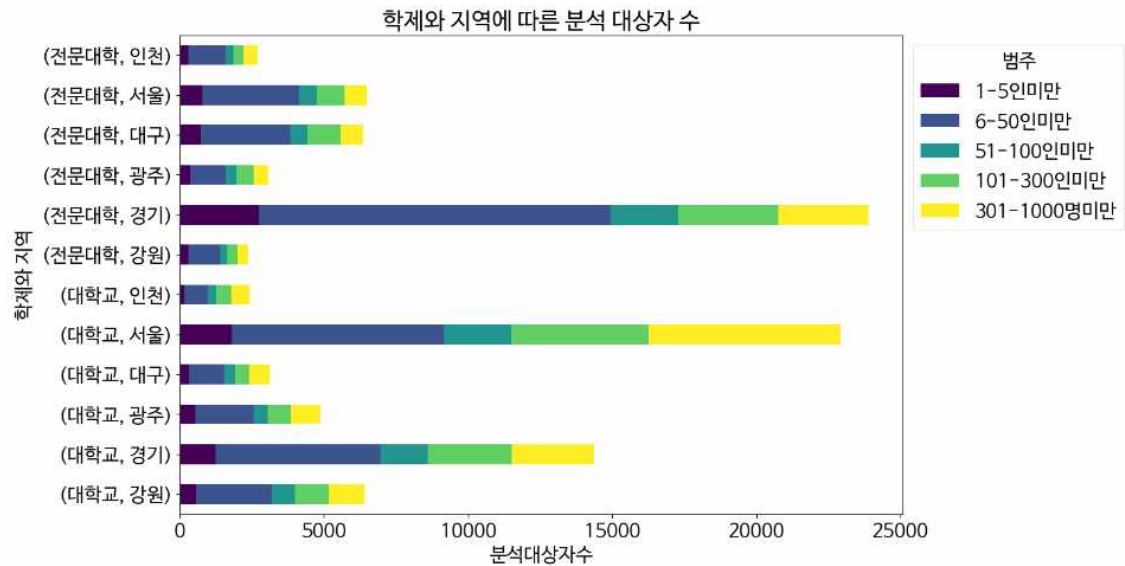
grouped_data = data[data.sum(axis=1) > 0].groupby(['학제', '지역']).sum().sum(axis=1)

plt.pie(grouped_data, labels=grouped_data.index, autopct='%.1f%%', textprops={'fontsize': 10})
plt.title('학제와 지역에 따른 취업자수 통합 비율')
plt.show()
```

학제와 지역에 따른 취업자수 통합 비율



• 학제 및 지역별 분석대상자 수



각 학제 및 지역에 대한 분석대상자 수를 막대 그래프로 시각화하여 어떤 학제나 지역에서 취업자 수가 많은지 비교할 수 있습니다.

```
# Pandas (matplotlib) bar
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('/content/content.csv', encoding='euc-kr')

grouped_data = data.groupby(['학제', '지역']).sum()[['1-5인미만', '6-50인미만', '51-100인미만', '101-300인미만', '301-1000명미만']]

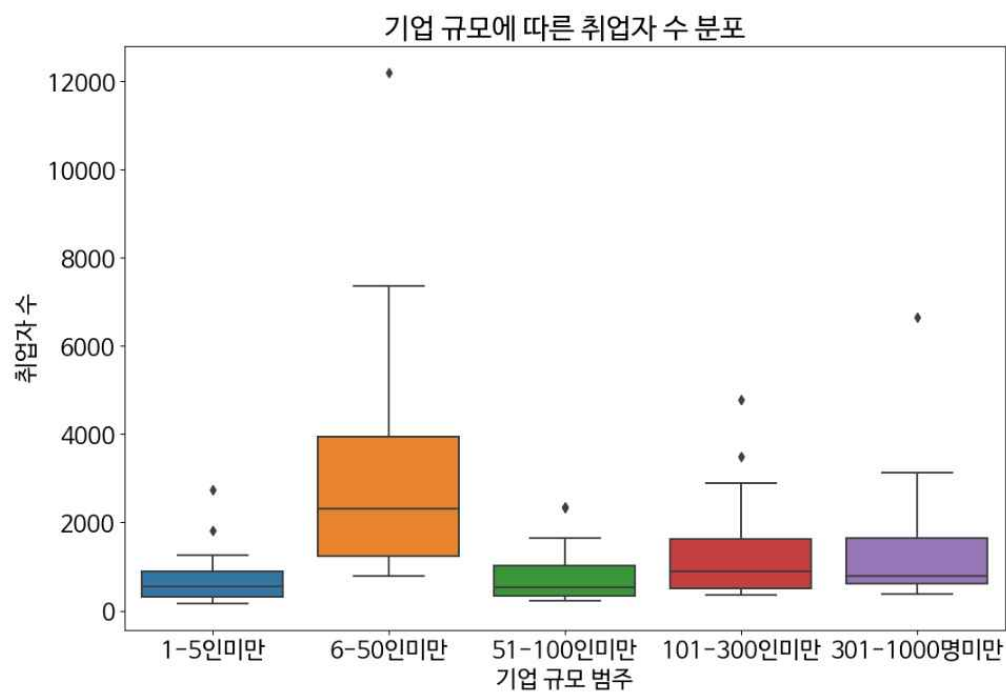
grouped_data.plot(kind='barh', stacked=True, figsize=(12, 8), colormap='viridis')

plt.title('학제와 지역에 따른 분석 대상자 수')
plt.xlabel('분석대상자수')
plt.ylabel('학제와 지역')

plt.legend(title='범주', bbox_to_anchor=(1, 1))

plt.show()
```

- 기업 규모에 따른 취업자 수 분포



소규모, 중규모, 대규모 기업 중 어느 범주에서 취업자 수가 가장 많은지 확인할 수 있고 박스 플롯으로 통해 각 기업 규모 범주에서의 취업자 수 분포를 박스 플롯으로 나타낼 수 있습니다. 이는 중앙값, 사분위수, 이상치 등을 확인할 수 있어 분포의 특징을 더 자세히 파악할 수 있습니다.

```
# 취업자 수의 주요 분포는 어떻게 되나요?
import seaborn as sns

plt.figure(figsize=(12, 8))
sns.boxplot(data=data[['1-5인미만', '6-50인미만', '51-100인미만', '101-300인미만', '301-1000명미만']])
plt.title('기업 규모에 따른 취업자 수 분포')
plt.xlabel('기업 규모 범주')
plt.ylabel('취업자 수')
plt.show()
```

- 지역별 기업 규모에 따른 취업자 수

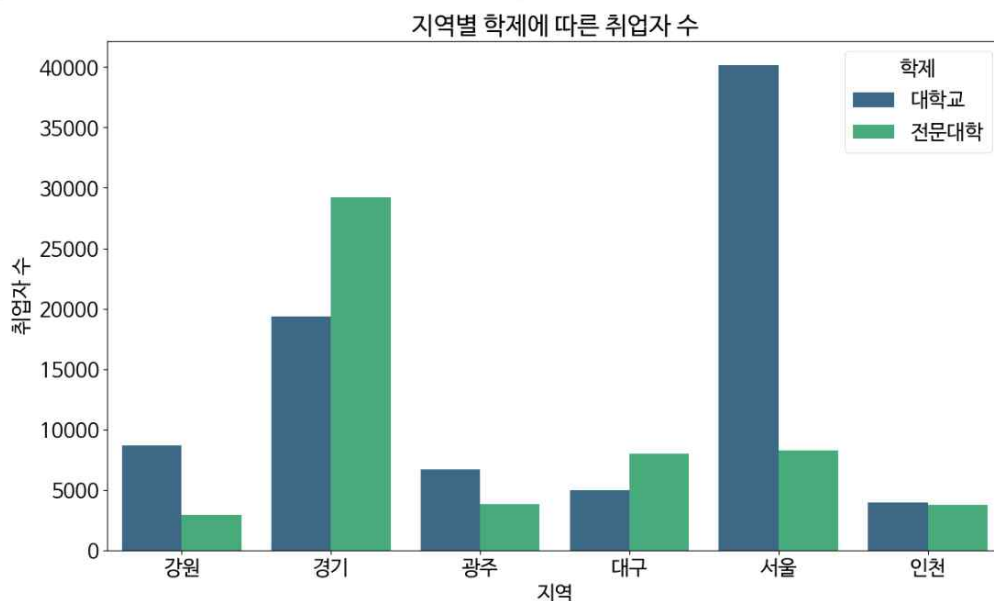


특정 지역에서는 특정 기업 규모 범위에서의 취업자 수가 두드러지게 높은지 확인할 수 있습니다. 히트맵을 통해 지역별로 기업 규모에 따른 취업자 수의 패턴을 시각적으로 확인할 수 있습니다. 각 행은 지역을 나타내며, 각 열은 기업 규모 범주를 나타냅니다. 색상의 진하 기는 해당 지역에서 해당 기업 규모 범주의 취업자 수를 나타냅니다.

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

plt.figure(figsize=(12, 8))
heatmap_data = info.pivot_table(index='지역', values=['1-5인미만', '6-50인미만', '51-100인미만', '101-300인미만', '301-1000명미만'], aggfunc='sum')
sns.heatmap(heatmap_data.T, cmap='viridis', annot=True, fmt=".0f", linewidths=.5)
plt.title('지역별 기업 규모에 따른 취업자 수')
plt.show()
```

- 지역별 학제에 따른 취업자 수



이 그래프를 통해 각 지역에서의 대학교와 전문대학의 학제별 취업자 수를 비교할 수 있습니다. 시각적으로 구분되어 있는 그룹화된 막대 그래프는 지역별로 어떤 학제가 취업자를 보유하고 있는지를 잘 보여줍니다.

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

# 지역별 학제에 따른 취업자 수 그룹화
grouped_data = info.groupby(['지역', '학제']).sum().reset_index()

# 그룹화된 막대 그래프
plt.figure(figsize=(14, 8))
sns.barplot(x='지역', y='분석대상자수', hue='학제', data=grouped_data, palette='viridis')
plt.title('지역별 학제에 따른 취업자 수')
plt.xlabel('지역')
plt.ylabel('취업자 수')
plt.legend(title='학제', bbox_to_anchor=(1, 1))
plt.show()
```


3. 결론

대학별 기업규모에 따른 취업자 수를 분석하고 시각화함으로써, 대학 졸업생들의 취업 동향에 대한 통찰을 얻고자 하였습니다. 주어진 데이터셋을 활용하여 다양한 학제와 지역에 속하는 대학들의 기업규모에 따른 취업자 수를 효과적으로 시각화하였습니다.

3.1 연구 요약

본 연구에서는 먼저 공공데이터 포털에서 제공되는 데이터를 수집하고, 해당 데이터를 적절하게 전처리하여 분석에 활용하였습니다. 데이터 수집 이후, 필요한 부분만을 추출하고 학제 및 지역에 따른 취업자 수를 통합하였습니다. 이를 통해 취업 동향의 다양성을 보다 명확하게 이해할 수 있었습니다 또한, 시각화를 통해 얻은 결과를 토대로 대학의 학제 및 지역별로 어떤 기업규모에서 취업이 주로 이루어지고 있는지를 파악할 수 있었습니다.

4. 참고 문헌

- 공공데이터포털 : <https://www.data.go.kr/data/15071025/fileData.do>
- 인하공업전문대학 : 교안 참고

5. github 주소

https://github.com/choi-jeonghyun/202144094_ChoiJeongHyun/tree/main