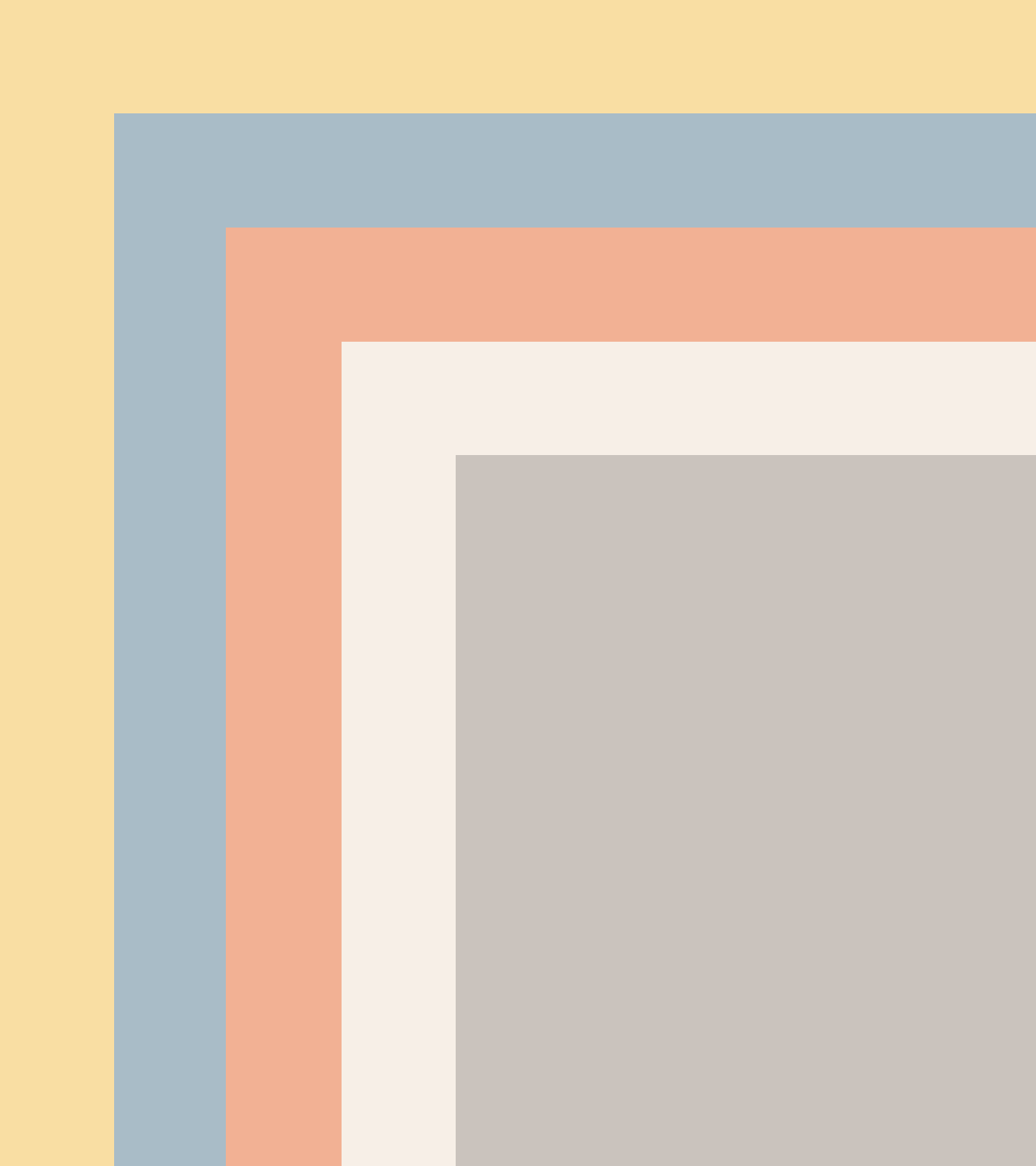




구직자를 위한 기업 트렌드 시각화 경진 대회

워크넷 오픈 API를 활용한 채용공고 분석

김형준





목차

1 서론

- 대회 개요
- 주제 선정 배경
- 분석 방향

2 분석 준비

- 데이터 수집
- 데이터 전처리
- 시각화 툴 선정

3 데이터 분석

- 전체 데이터 분석
- IT직종 세부 분석

4 마치며..

1

서론

서론

#대회 개요 #주제 선정 배경 #분석 방향



- 대한민국 최초의 **AI Competition** 플랫폼
- 대한민국의 Kaggle로서 수많은 빅데이터, AI 경진대회를 주최
- 데이터 분석에 관한 활발한 토론과 코드 공유, 리더보드를 활용한 실시간 순위를 확인 하는 등, 이용자로 하여금 **△△로의 실력을 발전** 시키기 위한 과제들을 제시한다.

서론


#대회 개요 #주제 선정 배경 #분석 방향

구직자를 위한 기업 트렌드 시각화 경진대회
고용정보원 | 기업 | 시각화

🏆 상금 : 총 1,000만원

📅 2021.12.06 ~ 2022.01.28 18:00 [+ Google Calendar](#)

👤 713명 🏠 마감



The diagram shows a central node labeled 'TREND' connected to four surrounding nodes: 'ELECTRON' (top left), 'COMPANY' (top right), 'SALARY' (bottom left), and 'LEAGUE' (bottom right). Below 'SALARY' is a node labeled 'SYSTEM'. A share icon is located at the bottom right of the diagram.

한국고용정보원

Korea Employment Information Service

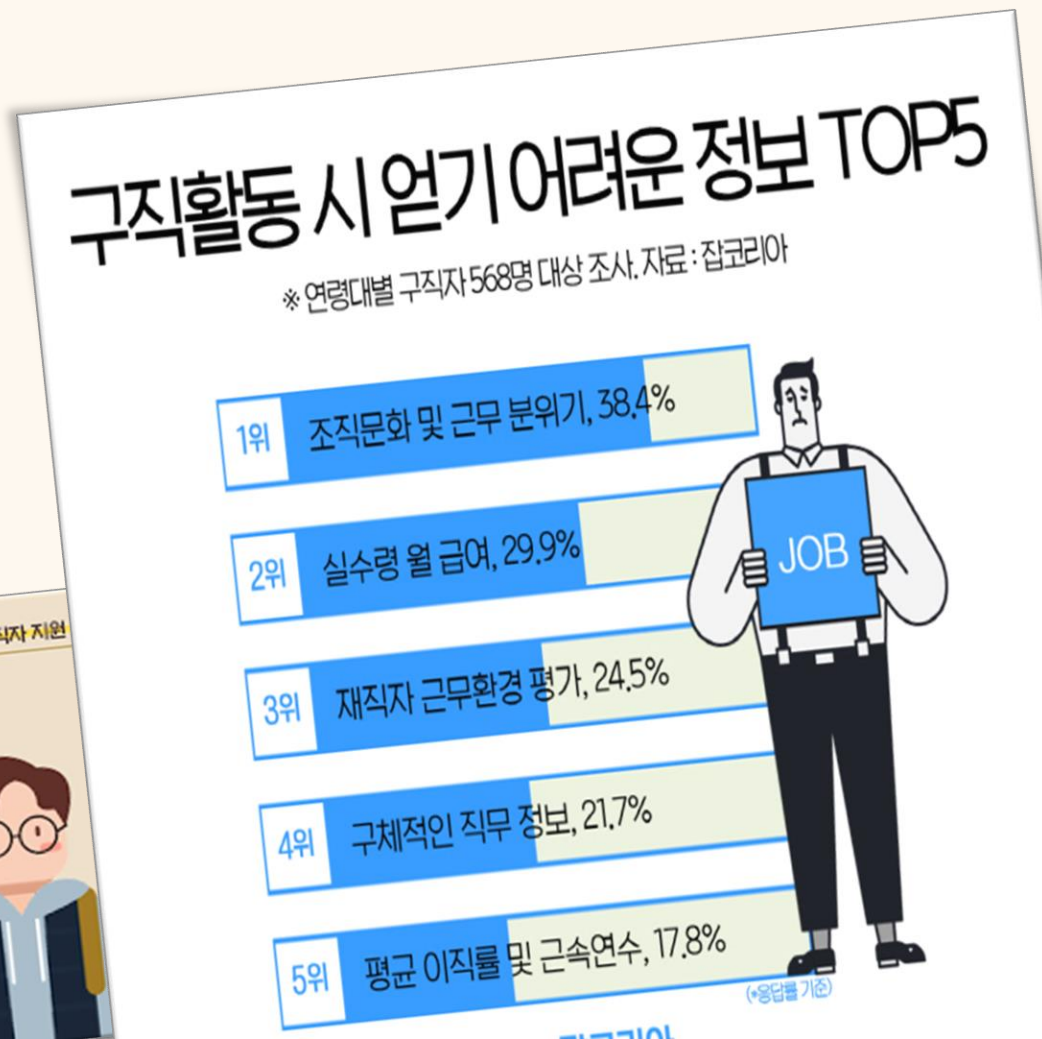
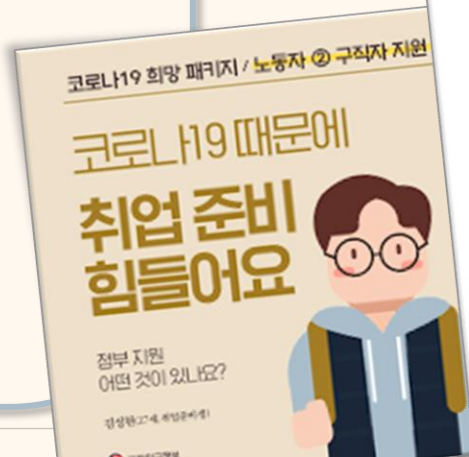


- 구직자들을 위해 기업 트렌드 정보를 빅데이터 시각화로 구현하고, 정보전달의 효율성을 기대할 수 있는 아이디어 및 인사이트 발굴
- 기업 트렌드를 확인할 수 있는 비정형&정형 데이터 발굴
- 구직자가 필요로 하는 기업 트렌드 아이디어 및 인사이트 발굴

서론

#대회 개요 #주제 선정 배경 #분석 방향

- **포스트 코로나** 시대 취업시장 또한 청년 구직자에게 점점 힘든 상황이 되고 있다.
- 다양한 취업 플랫폼에서 많은 취업 정보가 올라오고 있지만 **이용자의 접근이 쉽지 않다.**
- 이러한 상황에서 구직자의 취업을 위해 정부와 플랫폼에서 **다양한 지원**을 하고 있다.
- 이러한 지원과 다양한 채용정보들을 이용자가 **접근하기 쉽고**, 시각화를 통해 **직관적**으로 받아 들이기 위해 데이터를 **가공&시각화**를 한다.



서론

#대회 개요 #주제 선정 배경 #분석 방향

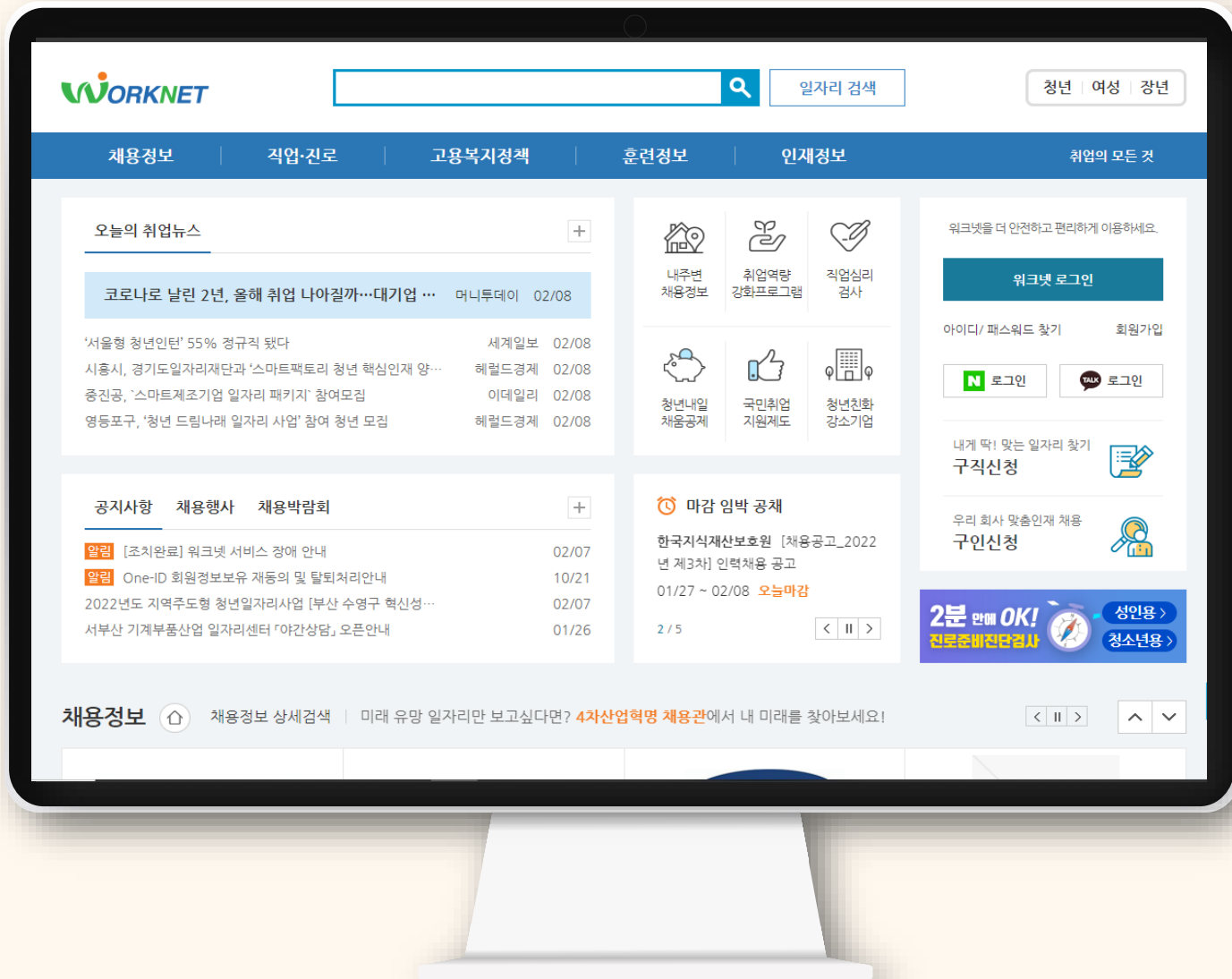


2

분석 준비

분석 준비

#데이터 수집 #데이터 전처리 #시각화 톨 선정



- 워크넷은 한국고용정보원에서 운영하는 국내 최대 취업 플랫폼
- 많은 취업정보가 매일 업로드 되고 다양한 직군의 정보가 업데이트
- 오픈 API를 통해 채용공고들을 XML로 제공
- BeautifulSoup 패키지를 이용해 데이터를 받아와 분석을 진행
- API외 별도 직종정리 데이터 수집

분석 준비

#데이터 수집 #데이터 전처리 #시각화 툴 선정

채용공고 DATA 수집

COLUMN	TYPE
회사명	Object
채용제목	Object
임금형태	Object
급여	Object
근무지역	Object
근무형태	Object
최소학력	Object
경력	Object
등록일자	Object
직종코드	Float

※ 데이터 수집 기준일 2022년 01월 19일

DATA 1차 전처리

- 급여 데이터 전처리
 - ✓ 시급, 일급 데이터 삭제
 - ✓ 월급 -> 연봉 환산
- 직종 코드 타입 변환 & 매칭
 - ✓ Float -> int
 - ✓ 직종정보 데이터와 매칭
- 등록일자 전처리
 - ✓ Object -> datetime
 - ✓ 년도/월별 그룹화
- 지역별 그룹화
 - ✓ 주요 도시별 그룹화

※ 시급, 일급 데이터의 경우 단순 아라바이트의 비중이 많아 일괄적으로 삭제

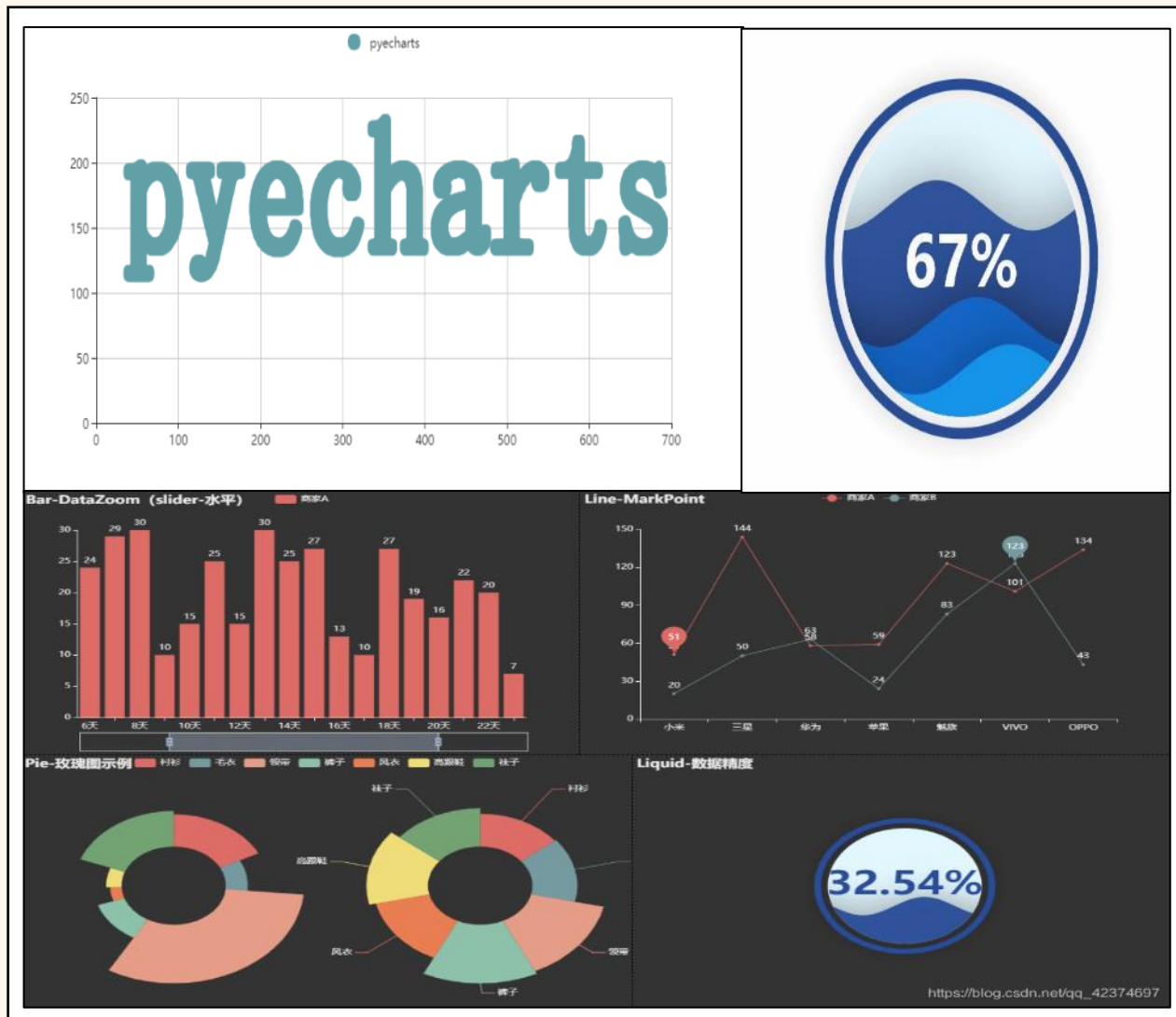
DATA 2차 전처리

- 채용제목 키워드 토큰화
 - ✓ IT기업 그룹화
 - ✓ Object -> List
 - ✓ 형태소 분리
 - ✓ 특수문자&불용어 삭제
 - ✓ 특정 value 미만 단어 삭제
 - ✓ 불필요한 단어 삭제

※ Konlpy패키지의 Komoran를 이용해 실행

분석 준비

#데이터 수집 #데이터 전처리 #시각화 툴 선정



- **Pyecharts**는 반응형 시각화를 지원하는 라이브러리
- 별도의 상용 툴을 이용하지 않고 **대시보드처럼 사용 가능**
- 이용자로 하여금 차트를 조절해 원하는 데이터에 접근 가능
- **다양한 커스터마이징**을 통해 디자인 가능

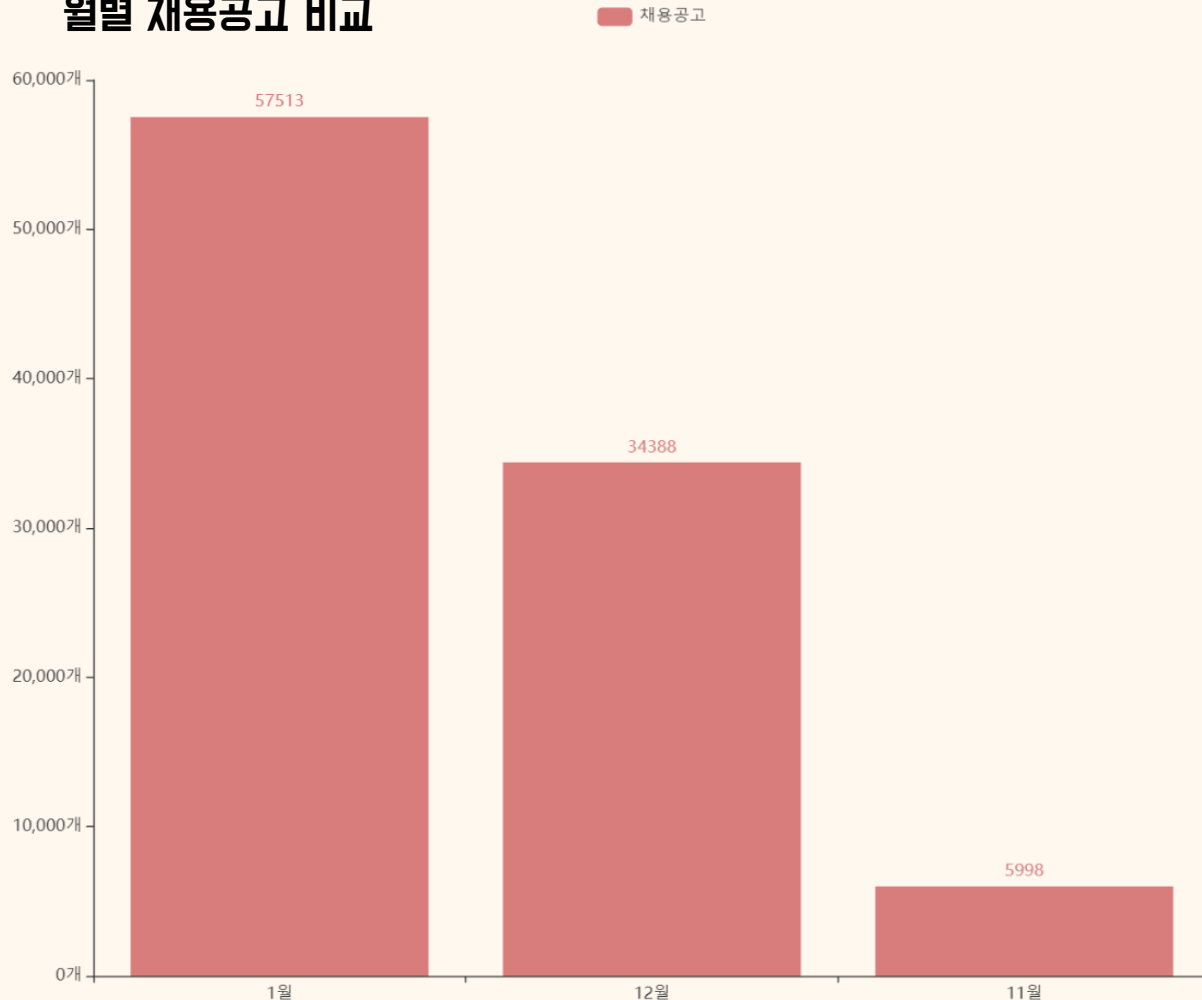
3

데이터 분석

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

월별 채용공고 비교

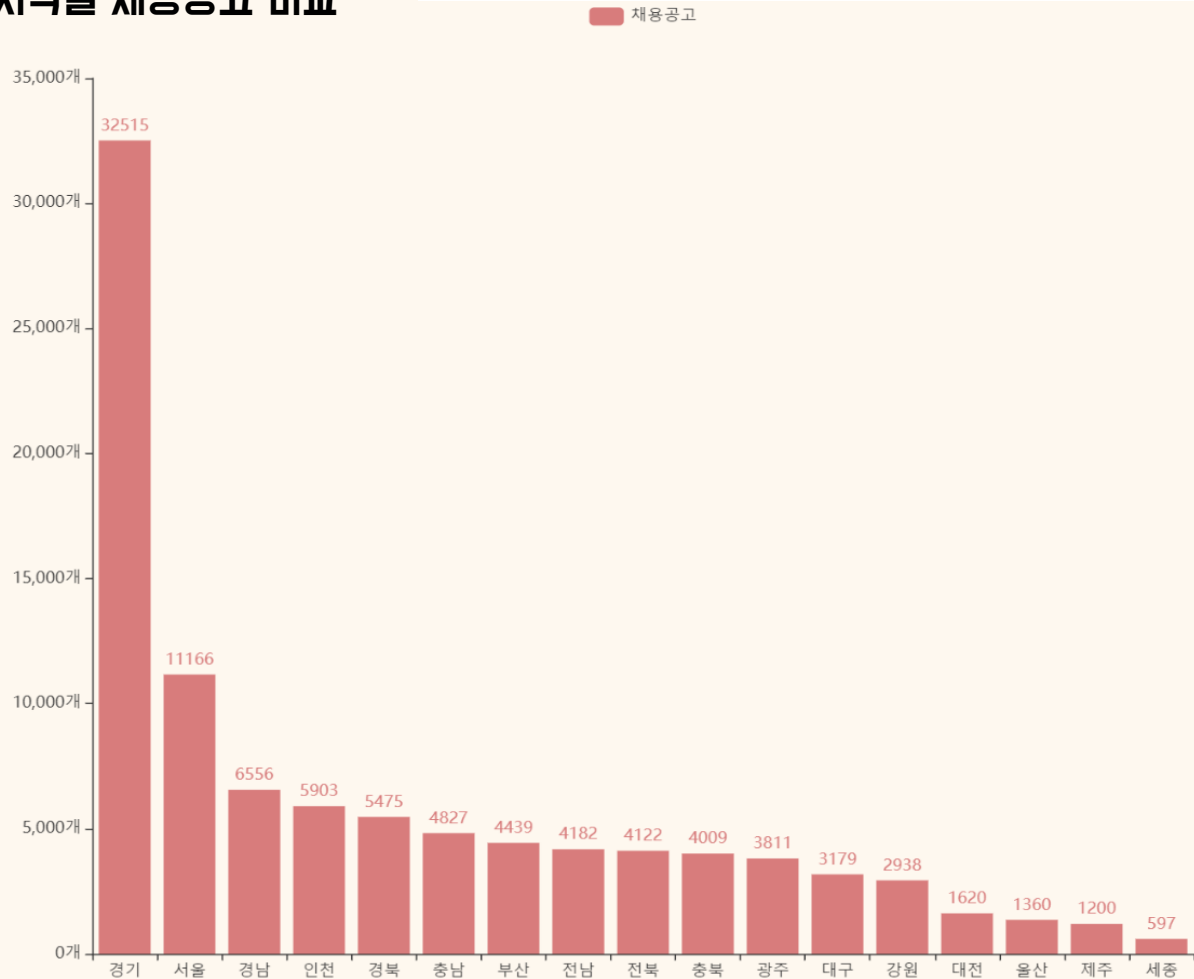


- 전처리가 후 총 채용공고
- 97,899개
- 1월 > 12월 > 11월
- 11월의 경우 데이터 수집 기준일
1월19일 당시 많은 채용공고가
채용마감 되었다.

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

지역별 채용공고 비교

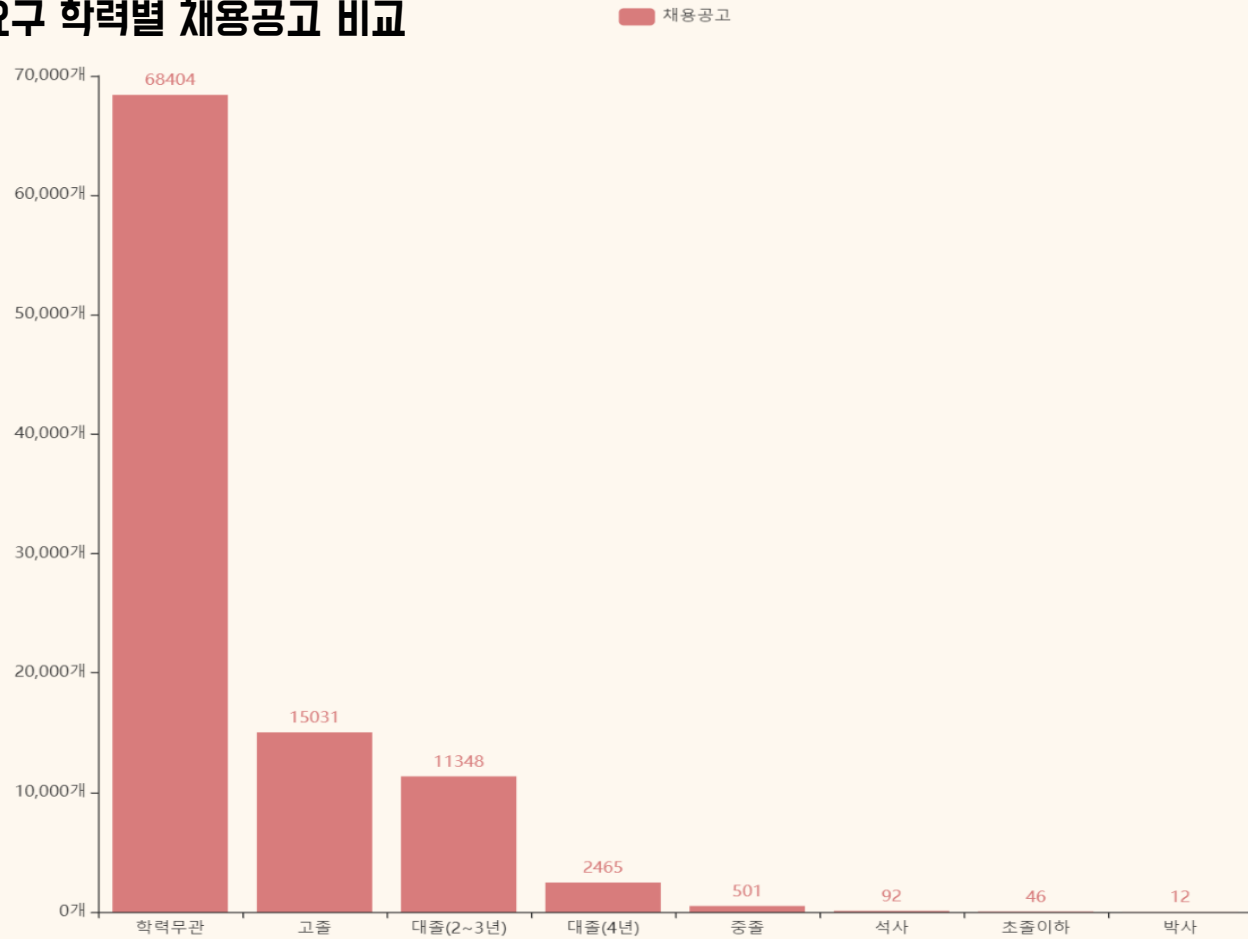


- 워크넷의 경우 중소기업 비중이 높고 외국계 기업의 비중이 낮다.
- 경기도는 많은 공업단지, 중소기업이 분포되어 있는 지역

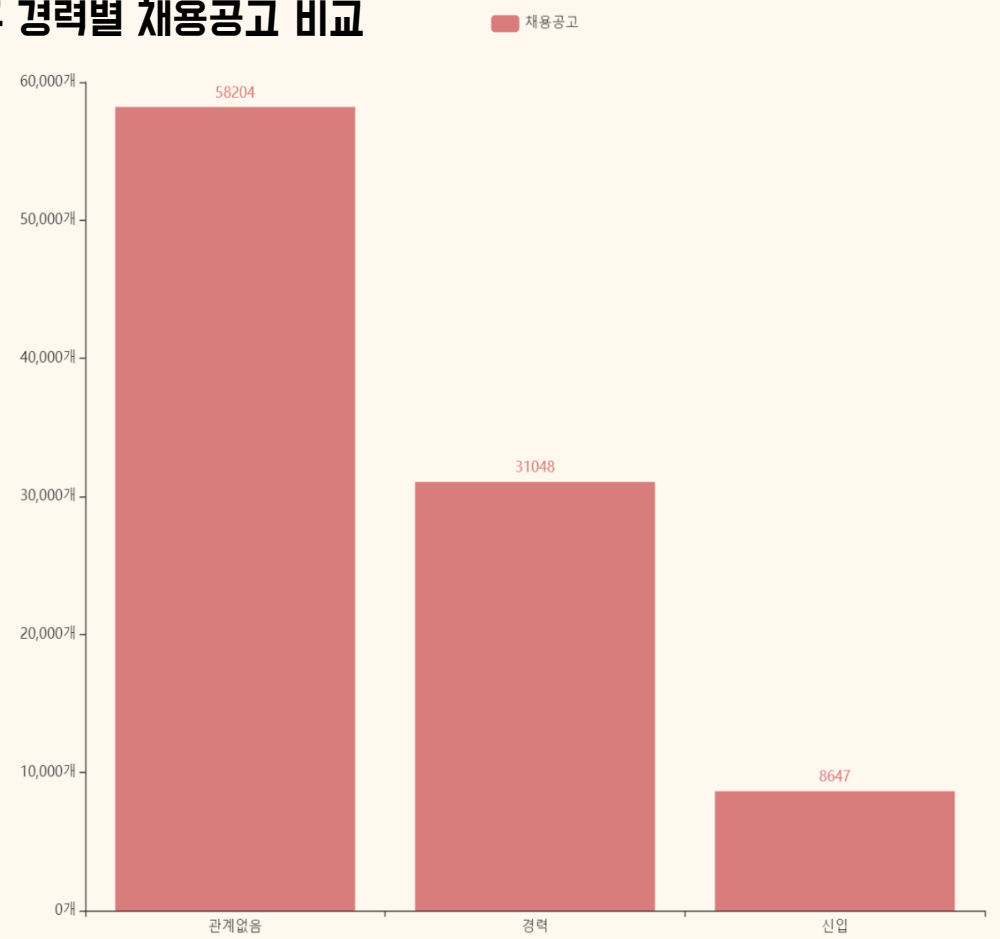
데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

요구 학력별 채용공고 비교



요구 경력별 채용공고 비교



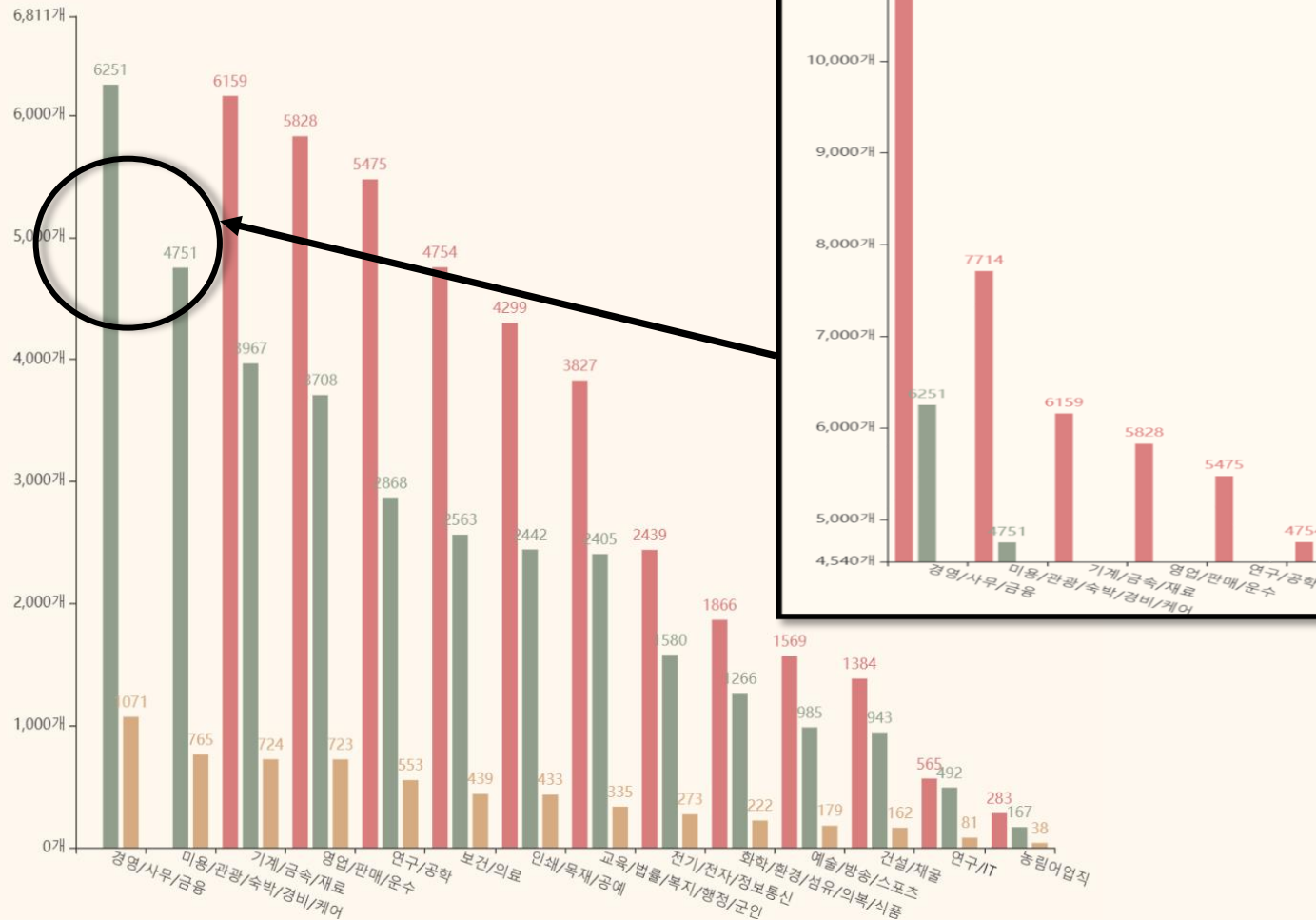
- 경력&학력 모두 **학력무관/관계없음** 가장 높음
- **Default value** 설정으로 인한 결과

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

월별 직종군 채용공고 비교

2022-01 2021-12 2021-11



월별 직종군 TOP5

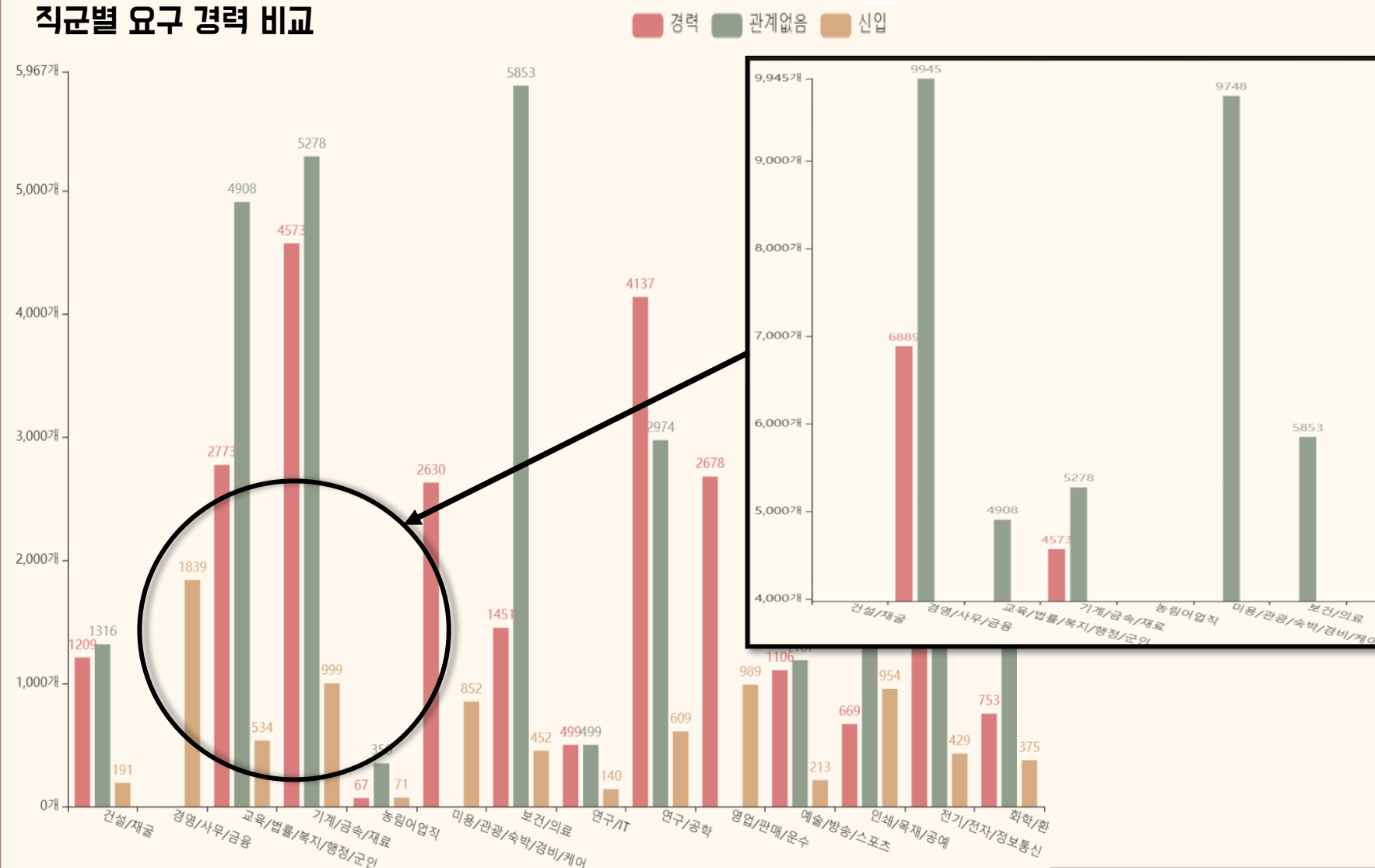
1. 경영/사무/금융
2. 미용/관광/숙박/경비/케어
3. 기계/금속/재료
4. 영업/판매/운수
5. 연구/공학

※ 스타트업과 대기업이 주로 분포된 연구/IT 직군의 비중이 매우 낮음

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

직군별 요구 경력 비교



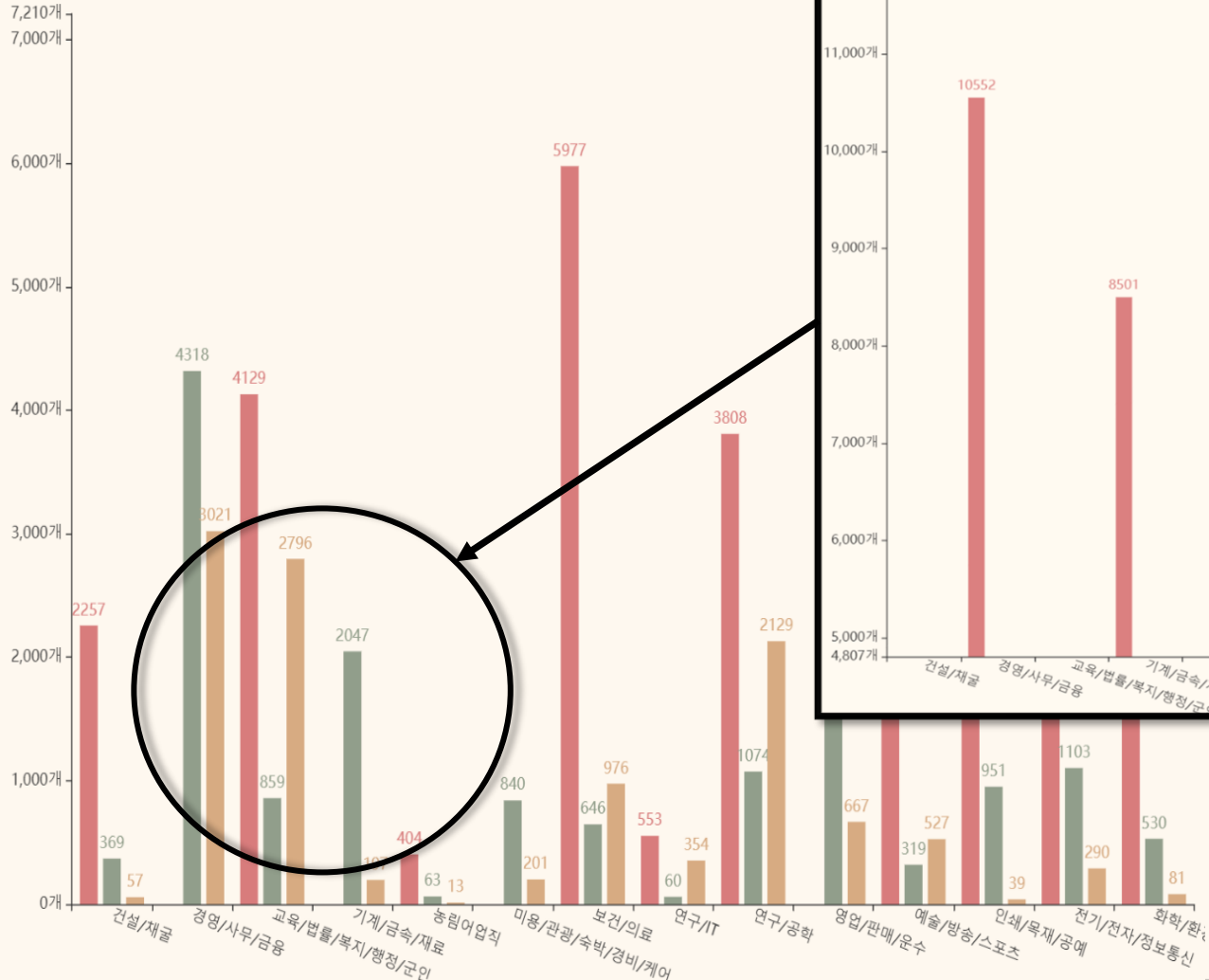
- 대부분 항목에서 경력무관 우세
- 경영/사무/금융, 미용/관광/숙박/경비/케어 직종 가장 우세
- 연구/공학, 연구/IT 유일하게 경력 우세

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

직군별 요구 학력 비교

학력무관 고졸 대졸(2~3년)

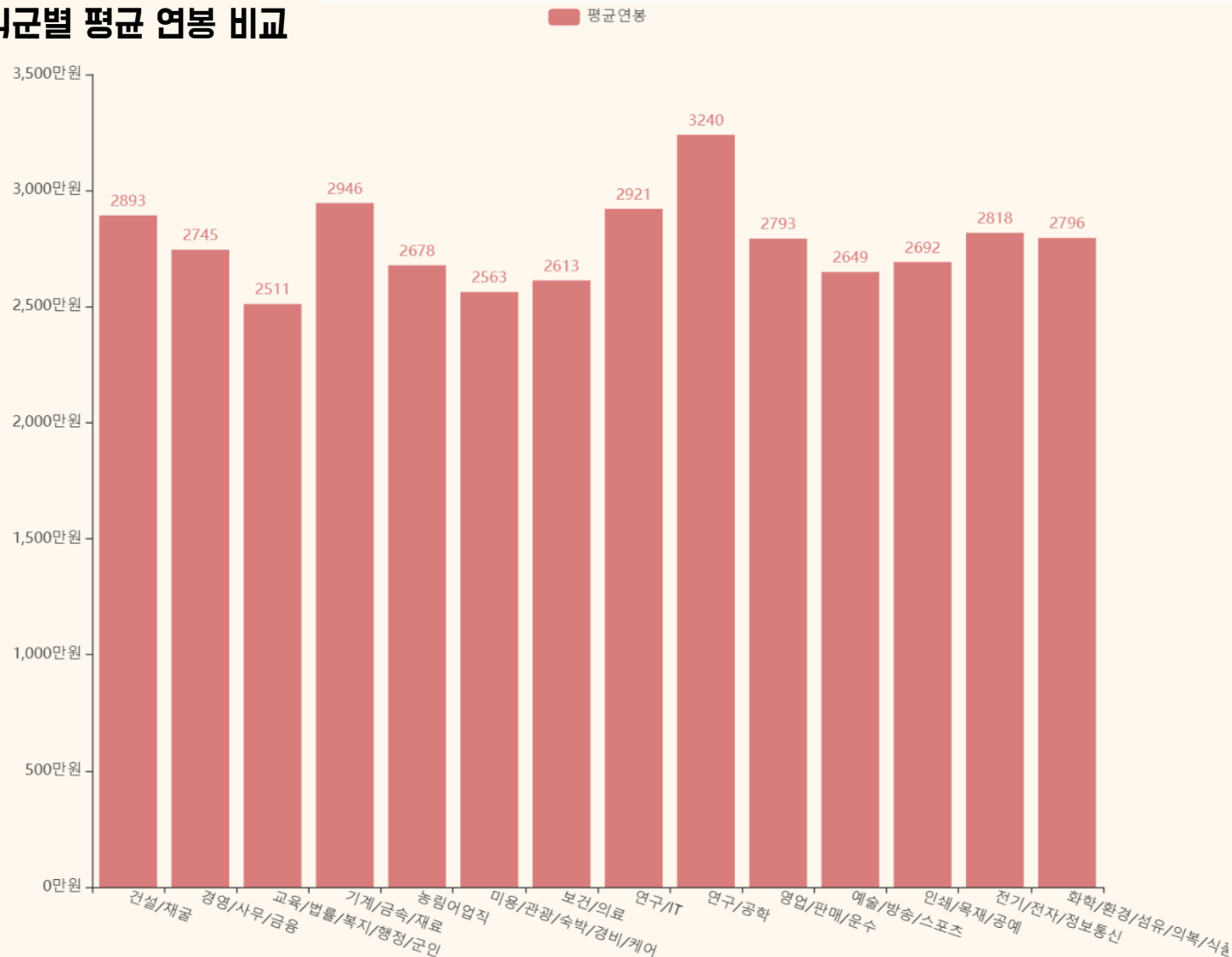


- 대부분 항목에서 **학력무관** 우세
- 건설/채굴, 미용/관광/숙박/경비/케어** 직종 가장 우세
- 교육/법률/복지/행정/군인** 유일하게 **고졸** 항목 우세

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

직군별 평균 연봉 비교



- **연구/공학**
가장 높은 수준 급여 조건 제시
- **교육/법률/복지/행정/군인**
가장 낮은 수준 급여 조건 제시
- 평균적으로 **2700** 수준의 연봉을 제시

※ 급여 데이터 중 범위로 지정된 급여값의 경우 가장 낮은 값을 기준으로 분석

#전체 데이터 분석 #IT직종 세부 분석

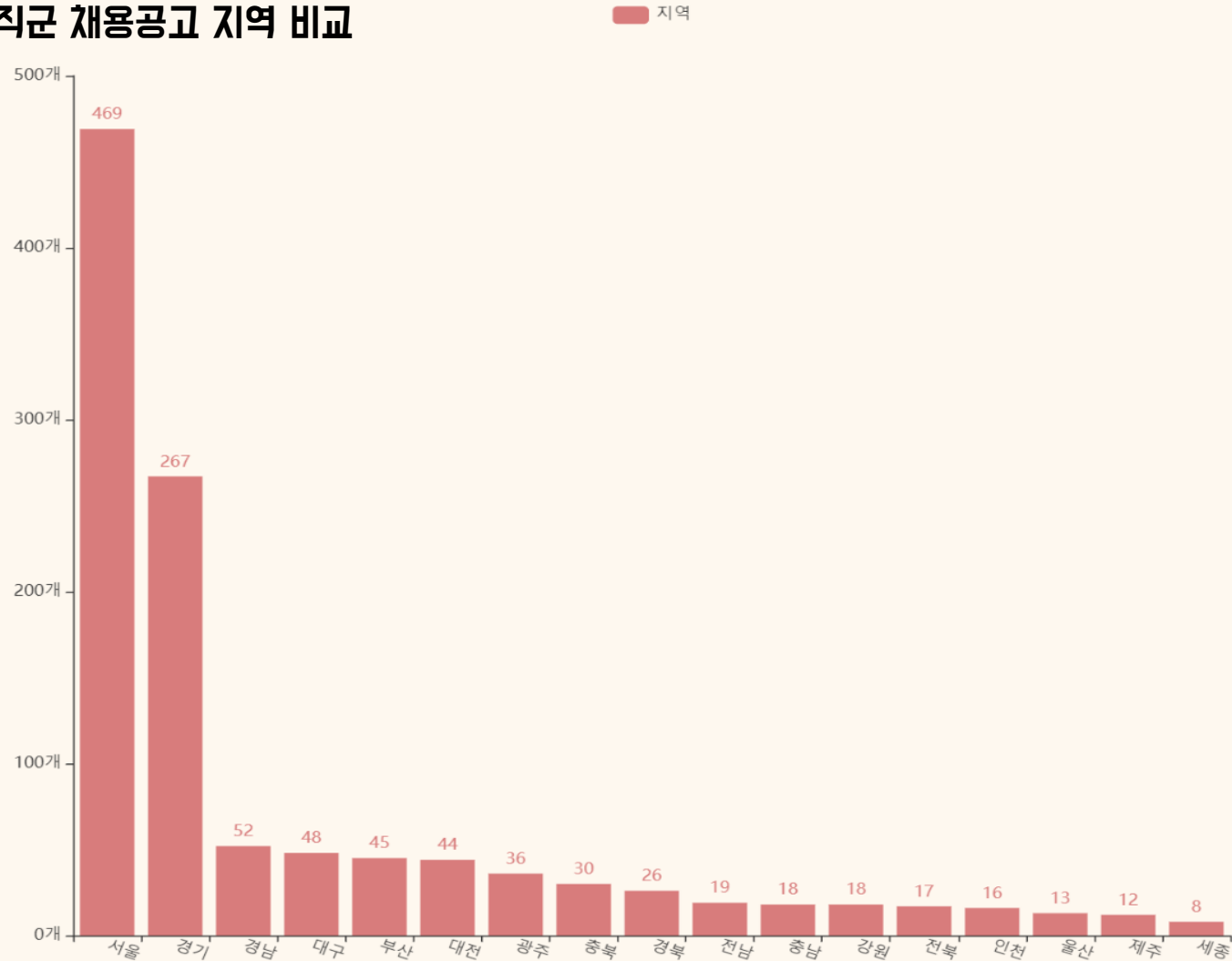
[illegible]

- 주요 관심 분야인 **IT분야**를 세부적으로 분석 진행
- 연구/IT 직군에 대한 주요 **키워드 분석 결과**
- 주요 키워드
 - ✓ 개발자
 - ✓ 경력
 - ✓ 개발
 - ✓ 엔지니어
 - ✓ 웹

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

IT직군 채용공고 지역 비교



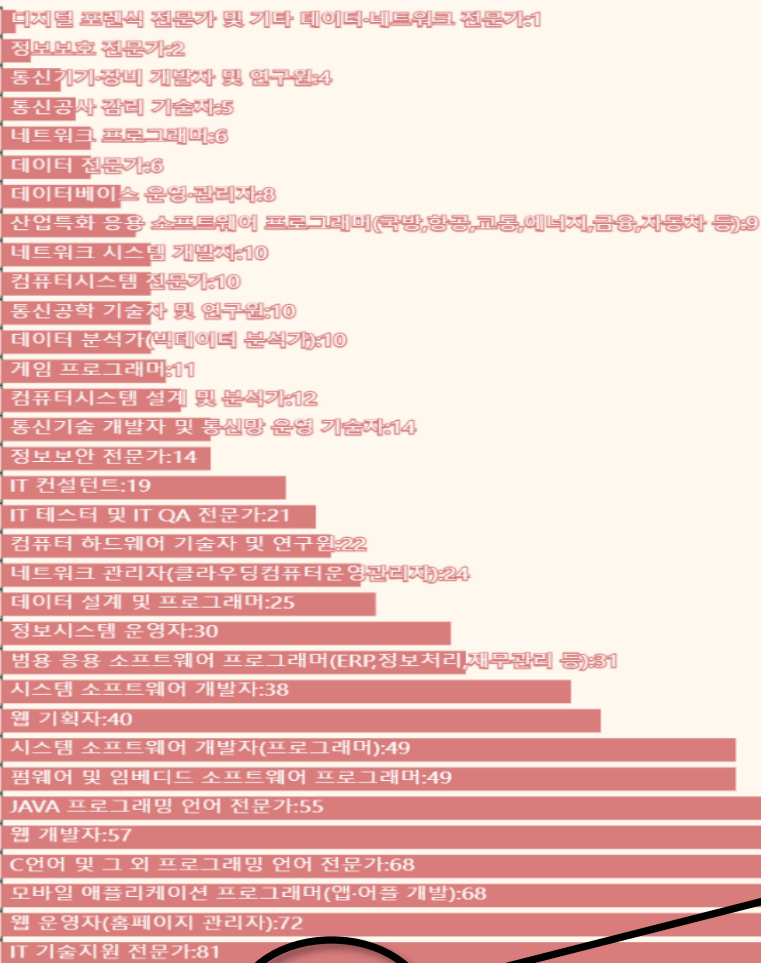
- 서울, 경기권에 채용공고 집중
- 실질적인 IT기업의 경우 대부분 서울, 성남, 판교, 고양 등 테크노밸리에 집중
- 그 외에는 경남 지방이 대부분의 비중을 차지하고 있다.

※테크노밸리
중앙정부 및 경기도 주도하여 국가 경쟁력 및 신도시의
자족기능 강화를 위하여 조성한
IT, BT, CT, NT 및 융합기술 중심의 첨단
혁신클러스터(연구개발단지)

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

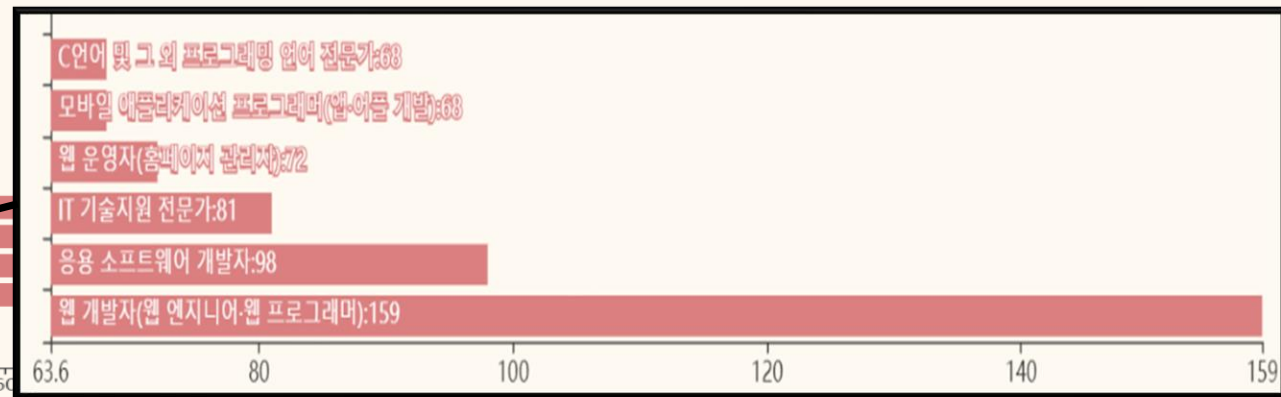
연구/IT직종 상세 분류



연구/IT 세부 직종군 TOP5

1. 웹 개발자
2. 응용 소프트웨어 개발자
3. IT 기술 전문가
4. 웹 운영자
5. 모바일 애플리케이션 프로그래머

- 빅데이터 분야 채용공고의 경우 워크넷에서 많은 부분을 차지하지 못하고 있다.



데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

세부 직군별 평균 연봉 비교



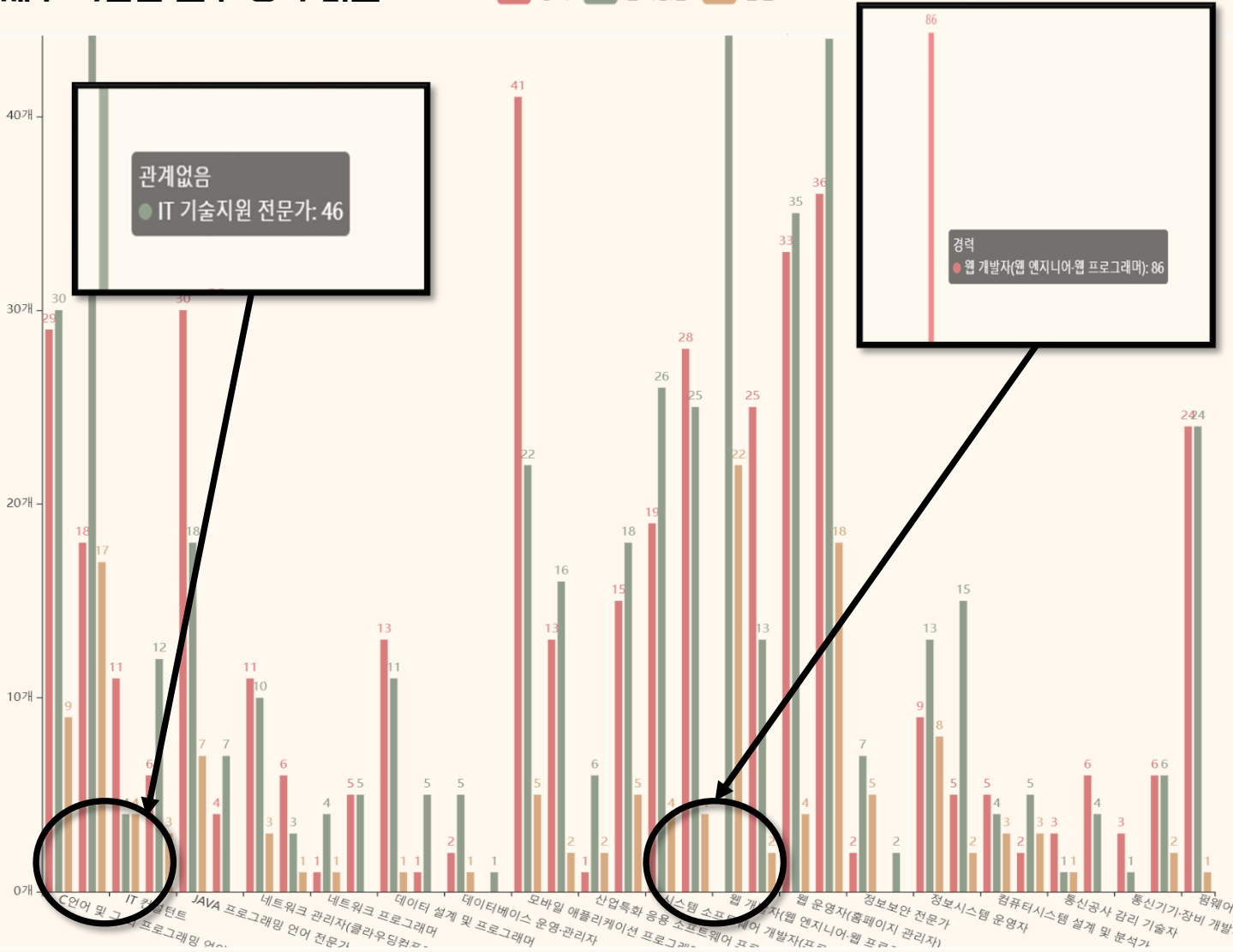
- 평균 연봉이 가장 높은 직군
- 통신기기/장비 개발자 및 연구원
- 평균 연봉이 가장 낮은 직군
- 데이터 분석가(빅데이터 분석가)
- IT직군 전체 평균 급여는
2000 후반 ~ 3000초반으로 구성
- 연구/IT 평균 급여 순위는 전체
직군 중 2위로 데이터 분석가를
제외하면 전체 직군 중 가장 높게
나타난다.

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

세부 직군별 요구 경력 비교

경력 관계없음 신입



IT 직군의 경우 타 직군에 비해 경력을 우대

- ✓ 모바일 애플리케이션 프로그래머
- ✓ JAVA프로그래밍 언어 전문가
- ✓ 네트워크 관리자
- ✓ 데이터 설계 및 프로그래머
- ✓ 웹 개발자

반대로 관계없음이 앞도적으로 우세한 직군

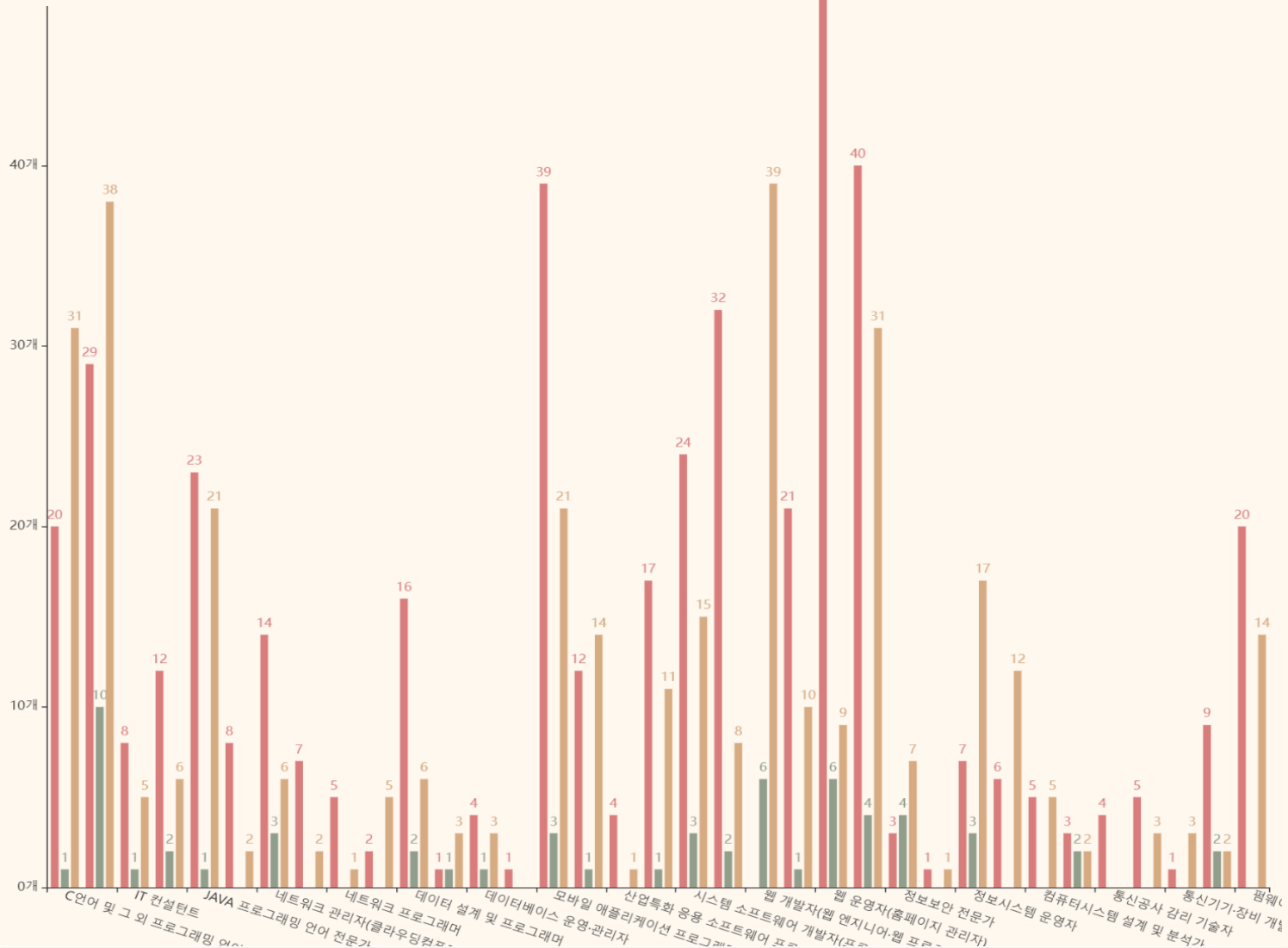
- ✓ IT 기술지원 전문가

데이터 분석

#전체 데이터 분석 #IT직종 세부 분석

세부 직군별 요구 학력 비교

학력무관 고졸 대졸(2~3년)



- 대부분의 직군에서 **학력무관**이 가장 우세
- 학력보단 **경력, 공모전 활동, 포트폴리오** 등 직무와 직접적 연관 있는 스펙의 중요도 ↑

4

마치며..

마치며..

분석의 한계



워크넷 API의 한계

- 워크넷 오픈 API에서 제공하는 채용공고 데이터의 경우 누락된 데이터(기업정보, 기업구분 등) 많다.
- 분석 도중 요청 파라미터의 값이 변경 되는 등 환경이 변화하며 데이터를 수집하는데 어려움이 있었다.



워크넷 구조적 한계

- 워크넷의 경우 고용노동부와 연계하여 채용을 매칭시켜준다. 그렇기 때문에 국가 정책에 따라 채용공고가 선정되고, 높은 채용률을 위해 대기업, 외국계 기업보단 중소기업이 대부분이다. 그에 더해 직종 또한 일반 사무,행정 생산직 등 단순 직종의 비중이 많아 IT업종과 같은 4차 산업 직종군의 비중이 매우 적어 분석에 어려움이 있었다.

분석의 의의



그럼에도 불구하고 워크넷!
























- 위와 같은 한계점이 명백할지라도 워크넷은 명실상부한 국내 최대 채용 플랫폼이다.
- 또한 오픈 API에서 제공하는 데이터의 경우 타 플랫폼과 다르게 다양한 파라미터값을 통해 이용자가 원하는 정보만을 빠르게 추출하는 것이 가능하다.



많은 정보에 따른 정확성 확보

- 제공하는 데이터의 양이 매우 방대해 일부 경우를 제외하고는 분석을 할 경우 실질적 통계 수치와 유사한 형태를 나타낸다.

마치며..

29	newage		구직자를 위한 기업 정보 한눈에!
30	초코우유__	  	구직자를 위한 취업 가이드북
31	소공소공	  	여성 구직자를 위한 근로 환경 분석 및 시각화
32	김고든		워크넷 오픈 API를 활용한 채용공고 분석 & IT직종 세부 분석
33	은행나무	   	은행 나 무조건 취업한다
34	DACONIO		팀 병합 방법
35	연희민서	   	경기도 기업 정보 시각화
36	호랑이 기운	  	구직자들이 원하는 트렌드 분석!
37	복맹이	  	NFT(Notice Follower's Trend) 팔로워의 트렌드에 주목하라 !!

이미지 클릭시 해당 링크로 이동

