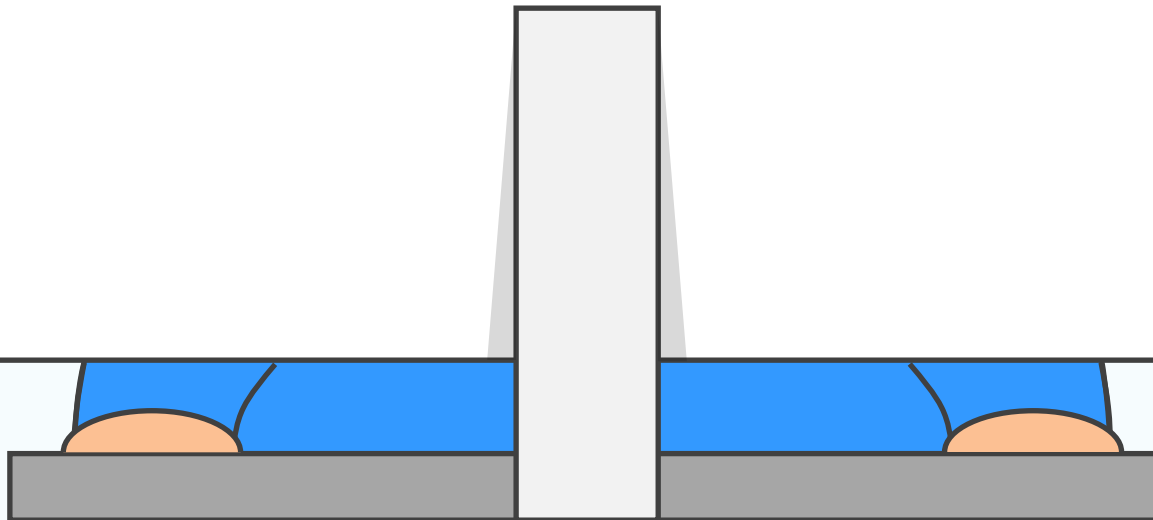




IT 취준생을 위한 직무백서

1조 - 최은정, 김민지, 정민지





목차

1 주제

- 주제 선정 배경

2 데이터 수집

- 데이터 수집 방법
- 크롤링 작업 중 핵심 소스 코드 설명

3 분석

- 분석 방법/모델
- 결론

4 참고

- 참고자료
- 과제 수행에서 어려웠던 점



IT 취준생을 위한 직무백서

1조 - 최은정, 김민지, 정민지

주제

주제 선정 배경

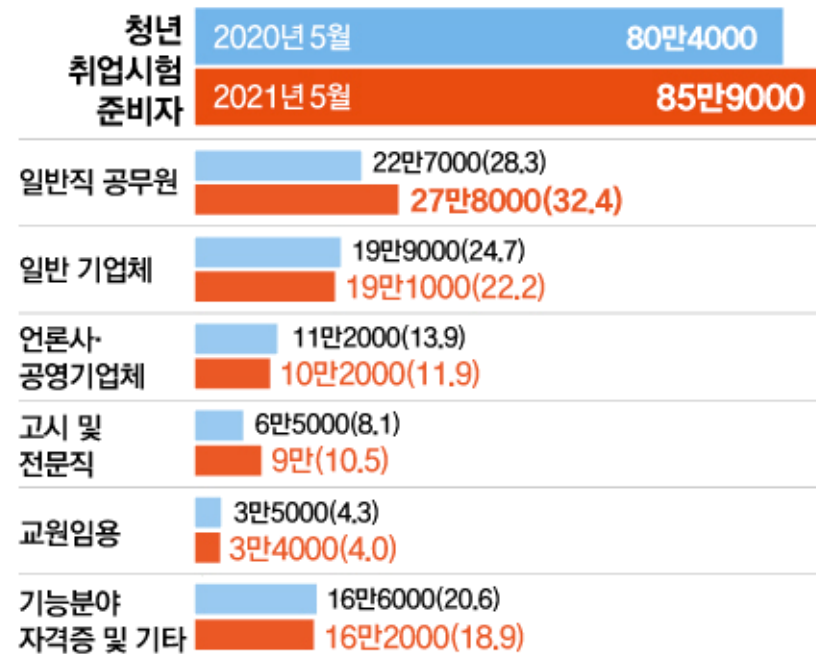
859,000

2021년
취업 준비생의 인구

주제 선정 배경

청년 비경제활동인구의 취업시험 준비 분야

단위: 명, ()안은 %



자료: 통계청

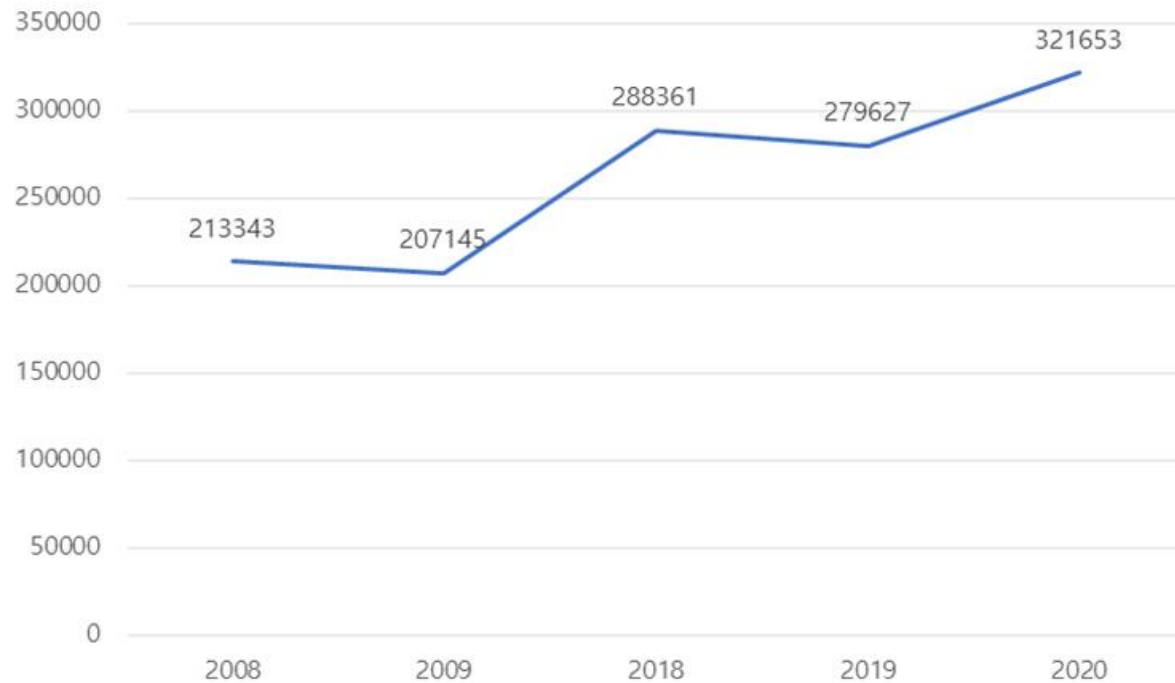
The JoongAng

코로나 19 이전에도
한파였던 청년 구직

코로나 19 이후
청년 구직 흑한기가
지속되고 있습니다.

주제 선정 배경

■ 한번도 취업하지 않았던 인구



Cold Surge

The highest Number of
Youth unemployment

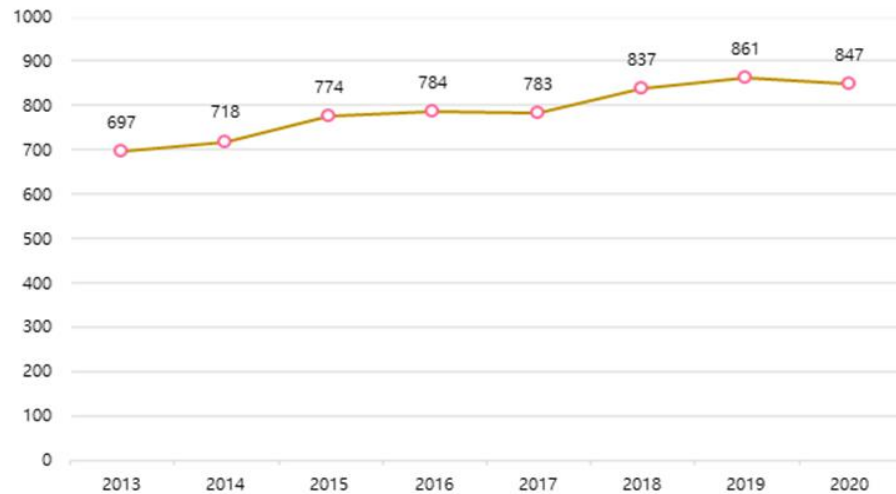
대학(전문대 포함)을 재학, 졸업한
학사·석사·박사인 25~39세 인구 중
취업 경력이 전혀 없는

‘취업 무경험자’는

2020년 32만1654명으로
집계됐다.

주제 선정 배경

꾸준히 취업률이 증가세인 IT 산업



산업별	2021. 10	2021. 09	2021. 08	2021. 07	2021. 06	2021. 05	2020
계	27,741	27,683	27,603	27,648	27,637	27,550	26,904
D 전기, 가스, 증기 및 공기 조절 공급업(35)	77	75	71	67	67	67	74
*전기·통신·통신·공용 (D,H,J,K)	3,428	3,414	3,360	3,362	3,371	3,321	3,181
J 정보통신업(58~63)	929	930	917	897	916	893	847
M 전문, 과학 및 기술 서비 스업(70~73)	1,237	1,228	1,216	1,218	1,233	1,208	1,164

IT 산업은 **성장가도!**

다른 산업들과 달리
꾸준히 IT 산업의 공고는
증가 중!

특히 코로나 19 이후
비대면 산업이 증가하여
IT 직무 신입, 경력직
에 대한 수요가 증가

자료 : KOSIS 국가 통계 포털 산업별 취업자 통계

주제 선정 배경

코로나 불황 불구 이통사 등 'IT업계 채용' 활발

이통3사, 포털, 게임 등 각사별 채용 진행

세 자릿수 채용 눈길, 우수 인재 영입 목적

디지털 대전환 대비 AI, 블록체인 전문 분야 각광

IT 계열 취업 희망자를 위해
수요가 많은 IT 직무를 파악하여
취직의 성공률을 올릴 수 있도록
주제 선정



IT 취준생을 위한 직무백서

1조 - 최은정, 김민지, 정민지

테이터 수집

데이터 수집 방법

데이터
분석 결과도출



데이터 수집 방법

사람인

The screenshot shows the '사람인' (Saramin) website interface. At the top, there are navigation links: 기업리뷰, 면접후기, 연봉정보, 기업클레이션, 이직로드맵, and a search bar. Below this, there are filters for 기업형태별 평균연봉, 산업(업종)별 평균연봉, 대졸초임, and 연봉정보 FAQ. The main section is titled '전체 연봉정보' (Overall Salary Information). Underneath, it says '기업별 평균 연봉 TOP 5' (Top 5 Average Salaries by Company). There are five tabs: 대기업, 중견기업, 공사·공기업, 외국계·법인/투자, and 상장기업. The '강소·인증·중소' tab is selected. The table below shows the top 5 companies in each category with their average salaries and year-over-year changes. At the bottom, there is a search bar for company names and a salary range slider from 1,000 to 100,000,000 KRW.

기업명	평균 연봉	전년 대비
1 (주)휴맥스	20,268	(108.8% ▲)
2 에스브이인베스...	18,888	(9.4% ▲)
3 컴퍼나케이파트...	13,026	(4.5% ▼)
4 SK텔레콤(주)	12,474	(9.9% ▲)
5 (주)티에스인베...	12,369	(23% ▼)

스택 오버플로우

The screenshot shows the Stack Overflow website interface. At the top, there are navigation links: About, Products, For Teams, a search bar, and Log in / Sign up buttons. The main section is titled 'Developers first. You'll never receive recruiter spam or see fake job listings on our site.' Below this, there are filters for Jobs, Companies, Why Jobs?, Applications, and Developer Story. The 'Jobs' filter is selected. The search bar contains 'Search all jobs' and 'Located anywhere' with a dropdown menu. The results show 503 jobs. The first job is 'Senior Software Engineer - 3D Graphics' by Esri. The second job is 'Software Developers for our offices in Berlin, Hamburg, Göttingen or Abingdon' by Evotec SE. The third job is 'Senior Software Engineer (Frontend)' by Deepfield Networks. The fourth job is 'Senior Full-Stack Software Engineer' by NOW Payment Services Provider LLC. The right sidebar shows details for the first job, including the company name, location, and a list of technologies used.

데이터 수집 방법

```
1 from flask import Flask, render_template, request, redirect, send_file
2 from scraper import get_jobs
3 import csv # CSV : Comma Separated Values
4
5 db = {} # 검색어 저장 임시 데이터 베이스
6
7 app = Flask("SuperScraper")
8
9 @app.route("/") # @ decorator : 밑에 위치한 함수 실행
10 def home():
11     return render_template("home.html")
12
13 # @app.route("/username") # dynamic URL
14 # def hitHelp(username):
15 #     return f"Hello, {username}! How are you doing?"
16
17 @app.route("/report")
18 def report(): # 검색 결과를 HTML에 렌더링
19     word = request.args.get('word')
20     if word:
21         word = word.lower()
22         existingJobs = db.get(word)
23         if existingJobs: # 임시 데이터베이스에 있는 결과 즉 이미 검색했던 단어일 경우
24             jobs = existingJobs
25         else: # 임시 데이터베이스에 결과가 없다면, 즉 새롭게 검색하는 단어일 경우
26             jobs = get_jobs(word) # 스레드풀로 무에서 검색어로 관련 직업 크롤링
27             do[word] = jobs # 결과를 임시 데이터베이스에 저장
28         else: # URL 파라미터에 word 값이 없을 경우 홈으로 리다이렉트
29             return redirect("/")
30     return render_template(
31         "report.html",
32         searchingBy=word,
33         resultNumber=len(jobs),
34         jobs=jobs
35     )
36
37 @app.route("/export")
38 def export(): # 검색 결과를 CSV 파일로 저장
39     try:
40         word = request.args.get('word')
41         if not word:
42             raise Exception()
43         word = word.lower()
44         jobs = db.get(word) # 임시 데이터베이스에서 해당 검색어에 해당하는 결과는? jobs에 반환
```

1. Replit 을 이용한
협업 & 형상 관리

2. CSV 파일로 저장

Back_Jobs_Erased.CSV	2021-12-09 오후 7:04	Microsoft Excel ...	4KB
Consultant_Jobs_Erased.CSV	2021-12-09 오후 7:12	Microsoft Excel ...	4KB
Database_Engineer_Jobs_Erased.CSV	2021-12-09 오후 7:13	Microsoft Excel ...	3KB
Embedded_Engineer_Jobs_Erased.CSV	2021-12-09 오후 7:14	Microsoft Excel ...	4KB
Front_Jobs_Erased.CSV	2021-12-09 오후 7:14	Microsoft Excel ...	5KB
Network_Engineer_Jobs_Erased.CSV	2021-12-09 오후 7:14	Microsoft Excel ...	3KB
Project_Manager_Jobs_Erased.CSV	2021-12-09 오후 7:15	Microsoft Excel ...	6KB
Security_Engineer_Jobs_Erased.CSV	2021-12-09 오후 7:15	Microsoft Excel ...	4KB
Software_Architect_Jobs_Erased.CSV	2021-12-09 오후 7:15	Microsoft Excel ...	3KB
System_Engineer_Jobs_Erased.CSV	2021-12-09 오후 7:15	Microsoft Excel ...	8KB



데이터 수집 방법

Start Bootstrap

Search for...

Project Manager Jobs

[Dashboard](#) / [Tables](#)

DataTables is a third party plugin that is used to generate the demo table below. For more information about DataTables, please visit the [official DataTables documentation](#).

DataTable Example

10 entries per page

Search...

title	company	location	apply_link
(Senior) Product Manager (all genders)	Elinvar GmbH	No office location	https://stackoverflow.com/jobs/560839
(Senior) Product Manager, Production (f/m/x)	AUTO1	No office location	https://stackoverflow.com/jobs/554540
Big Data Platform Product Owner	JPMorgan Chase Bank, N.A.	Jersey City, NJ	https://stackoverflow.com/jobs/560174
Ceph and software-defined storage product manager	Canonical	No office location	https://stackoverflow.com/jobs/522975
Data Solutions - Product Manager	Canonical	No office location	https://stackoverflow.com/jobs/467497
IAM & Cloud Product Management Lead	JPMorgan Chase Bank, N.A.	Columbus, OH	https://stackoverflow.com/jobs/561191
Intern - Product Management Shop Applications (m/f/d)	ABOUT YOU SE & Co. KG	Hamburg, Deutschland	https://stackoverflow.com/jobs/332770
Intern Product Management - CRM (m/f/d)	ABOUT YOU SE & Co. KG	Hamburg, Deutschland	https://stackoverflow.com/jobs/465961
Juju - Product Manager	Canonical	No office location	https://stackoverflow.com/jobs/451982
Kubernetes operators - Product Manager	Canonical	No office location	https://stackoverflow.com/jobs/451981

Showing 1 to 10 of 47 entries

1 2 3 4 5 >

크롤링 작업 결과물을
사이트 페이지로 구현

시각적으로 차이를 확인
효율적인 데이터 분석

#자동_오름_내림차순
#검색_기능_제공



핵심 코드 설명

```
5 # 사람인 평균 연봉 정보 크롤링
6 def saramin_salary(result):
7     for page in range(1, 501): # 1페이지당 20개의 항목들이 있다.
8         # 페이지를 제외한 사람인 사이트 url
9         saramin_url = 'https://www.saramin.co.kr/zf_user/salaries/total-salary/list?
10         page=%d&order=rank&industry_cd=&company_cd=&rec_status=&search_company_nm_org=&
11         search_company_nm=&min_salary=&max_salary=&request_modify_company_nm=' %page
12         html = urllib.request.urlopen(saramin_url)
13         saramin_soup = BeautifulSoup(html, 'html.parser')
14
15         tag_ul = saramin_soup.find("ul", {"class": "list_salary"}) # list_salary 클래스
16         ul
17         tag_list = list(tag_ul.find_all('li')) # ul 내의 li 태그들을 다 찾아 리스트화
18
19         for in_li in tag_list: # li 태그 내의 정보
20             company_name = in_li.find("a", {"class": "link_tit"}).text # 기업 이름
21             company_dd = in_li.find_all('dd') # 기업 정보
22             company_type = company_dd[0].text # 기업 형태
23             company_work = company_dd[1].text # 산업(업종)
24             company_salary = in_li.find('span', {"class": "txt_g txt_avg"}).text # 평균
25             연봉 (단위: 만)
26             result.append([company_name, company_type, company_work, company_salary]) #
27             result 에 추가
28
29     return result
```

#사람인_사이트

#1000개_회사

#산업별_연봉을_한번에

#회사_규모별_연봉차이



핵심 코드 설명

```
12 # (3) 한 페이지 당 내부 태그 순회
13 def extract_job(html):
14     title = html.find("a", "s-link")["title"]
15     # company_row 리스트 안에 어떤 요소가 몇 개 올지 알고 있을 때 -> unpacking 기능
16     company, location = html.find("h3", "mb4").find_all("span", recursive=False) #
17     # recursive : 더 깊게 안 들어가게 함, 즉 다 가져오지 않고 같은 레벨의 요청 값만 가져옴. 예를
18     # 들어 span 안에 또 span이 있으면 그 span은 가져오지 않음
19     company = company.get_text(strip=True)
20     location = location.get_text(strip=True)
21
22     job_id = html['data-jobid']
23
24     return {'title': title, 'company': company, 'location': location, "apply_link":
25            f"https://stackoverflow.com/jobs/{job_id}"}
26
27 # (2) 페이지 각각을 돌면서 extract_job을 호출하여 각 페이지의 정보 크롤링
28 def extract_jobs(last_page, url):
29     jobs = []
30     # 첫 페이지부터 끝 페이지까지 순회
31     for page in range(last_page):
32         result = requests.get(f"{url}&pg={page+1}")
33         soup = BeautifulSoup(result.text, "html.parser")
34         results = soup.find_all("div", class_="-job")
35         # print(f"Scraping StackOverflow: Page {page+1} //", len(results))
36         for result in results:
37             job = extract_job(result)
38             jobs.append(job)
39     return jobs
40
41 # (1) stackoverflow 에서 가지고 온다.
42 def get_jobs(word):
43     url = f"https://stackoverflow.com/jobs?q={word}"
44     last_page = get_last_page(url) # 해당 url이 가진 마지막 페이지 값 구하는 함수
45     jobs = extract_jobs(last_page, url)
46     return jobs
```

#스택_오버플로우

#NCS_기준_직무검색

#Front_Back_구분

#직무별_고용수_비교

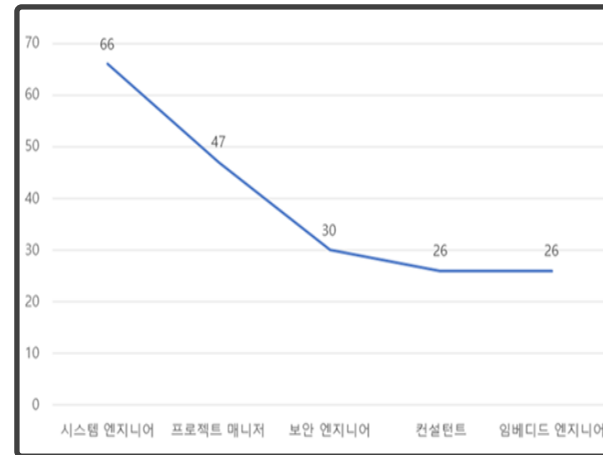
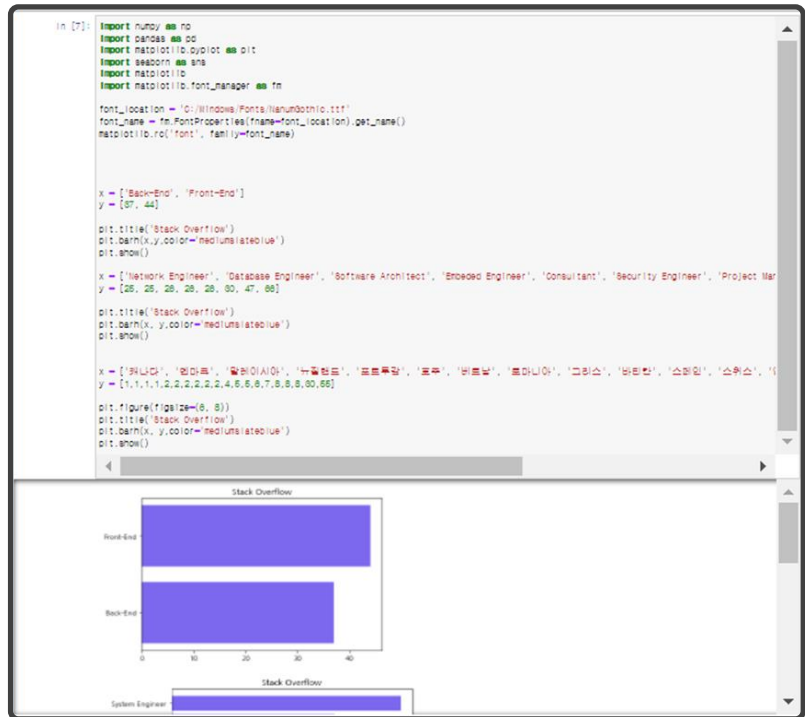


IT 취준생을 위한 직무백서

1조 - 최은정, 김민지, 정민지

분석

분석 방법 및 모델



순위	기업명	평균 연봉
1	카카오	10,308
2	현대오트모터	8,108
3	에릭슨엘지	7,450
4	넷마블	7,354
5	씨제이이엔엠	6,976

크롤링을 통해 얻은 데이터를 Matplot 을 통해 시각화 하여 분석하고,
이를 라인그래프와 바차트, 표로 표현

1000 기업 평균 연봉 TOP 5

순위	기업명	기업 형태	산업	평균 연봉
1	부국증권	대기업 코스피	금융	20,641
2	휴맥스	대기업 코스닥	제조,화학	20,268
3	에스브이인베스트먼트	중소기업 코스닥	금융	18,888
4	지누스	코스피	제조,화학	18,557
5	한양증권	대기업 코스피	금융	18,149

1000 기업 평균 연봉 LEAST 5

순위	기업명	기업 형태	산업	평균 연봉
1	노랑풍선	중소기업 코스닥	서비스업	1,758
2	한국코퍼레이션	코스닥	서비스업	1,971
3	인터파크	중견기업 코스닥	IT	2,037
4	씨앤에스자산관리	중견기업	서비스업	2,107
5	참좋은여행	중견기업 코스닥	서비스업	2,278

1000 기업 평균 연봉 차이 비교



평균 연봉
20,641



평균 연봉
1,758



 1개 : 2,000만원
500만원 : 반올림 진행

IT 기업 평균 연봉 비교

Top

순위	기업명	기업 형태	평균 연봉
1	카카오	대기업 코스닥	10,308
2	현대오토에버	대기업 코스피	7,545
3	풀어비스	중소기업 코스닥	7,545
4	에릭슨엘지	대기업	7,450
5	다쏘시스템코리아	중견기업	7,421

Least

순위	기업명	기업 형태	평균 연봉
1	인터파크	중견기업 코스닥	2,037
2	위드이노베이션	중소기업	2,793
3	로지시스	코스닥	3,020
4	케이티아이에스	대기업 코스피	3,041
5	인피닉	중소기업	3,150

IT 기업 평균 연봉 차이 비교



평균 연봉
10,308



평균 연봉
2,037



 1개 : 1,000만원

500만원 : 반올림 진행



IT 기업 형태 별 평균 연봉 TOP 5

순위	기업명	평균 연봉
1	카카오	10,308
2	현대오토에버	8,108
3	에릭슨엘지	7,450
4	넷마블	7,354
5	씨제이이엔엠	6,976

대기업

순위	기업명	평균 연봉
1	풀어비스	7,545
2	두나무	6,724
3	우아한형제들	5,918
4	아우토크립트	5,878
5	이지케어텍	5,631

중견기업

순위	기업명	평균 연봉
1	다쏘시스템코리아㈜	7,421
2	넷마블	7,354
3	더블유게임즈	7,350
4	다우기술	6,218
5	신한디에스	6,159

중소기업

IT 기업 형태 별 평균 연봉 LEAST 5

순위	기업명	평균 연봉
1	케이티아이에스	3,041
2	에스씨지솔루션즈	3,822
3	케이티엠앤에스	4,023
4	피에스앤마케팅	4,189
5	카카오브이엑스	4,458

대기업

순위	기업명	평균 연봉
1	인터파크	2,037
2	커리어넷	3,316
3	아이지엑스	3,799
4	케이지모빌리언스	3,895
5	솔로몬테크노서플라이	4,202

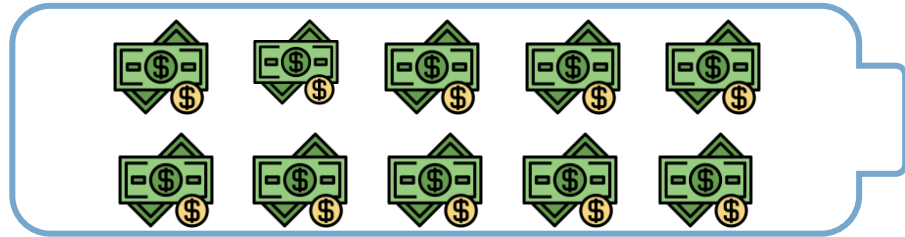
중견기업

순위	기업명	평균 연봉
1	위드이노베이션	2,793
2	인피닉	3,150
3	유플러스아이티	3,231
4	서울데이터통신	3,407
5	비플라이소프트	3,504

중소기업



IT 대기업 평균 연봉 차이 비교



평균 연봉
10,308



평균 연봉
3,041

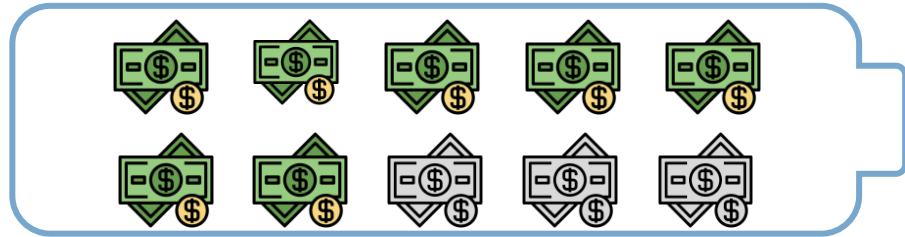


 1개 : 1,000만원

500만원 : 반올림 진행



IT 중견기업 평균 연봉 차이 비교



평균 연봉
7,421



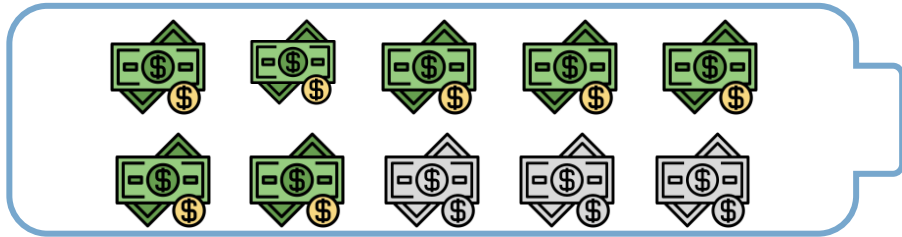
평균 연봉
2,037



 1개 : 1,000만원
500만원 : 반올림 진행



IT 중소기업 평균 연봉 차이 비교



평균 연봉
7,545



평균 연봉
2,793

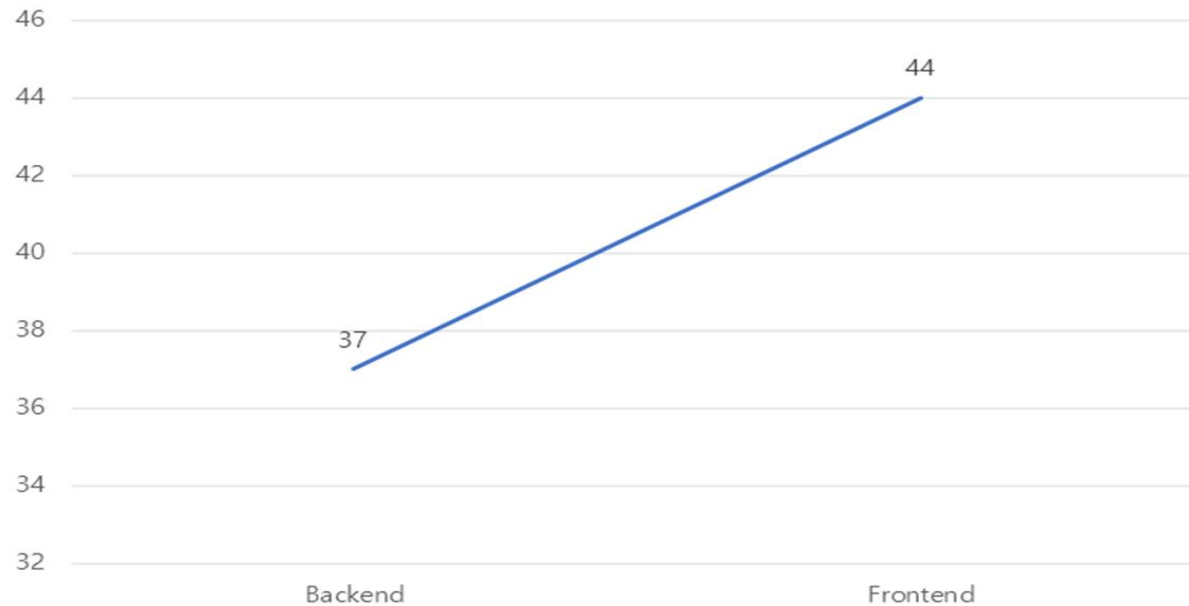


 1개 : 1,000만원
500만원 : 반올림 진행



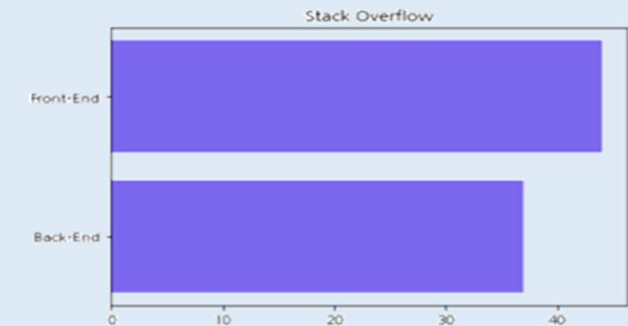
스택 오버 플로우 직무 분석 (Front/Back)

■ 직무 (FRONT / BACK) 공고 수



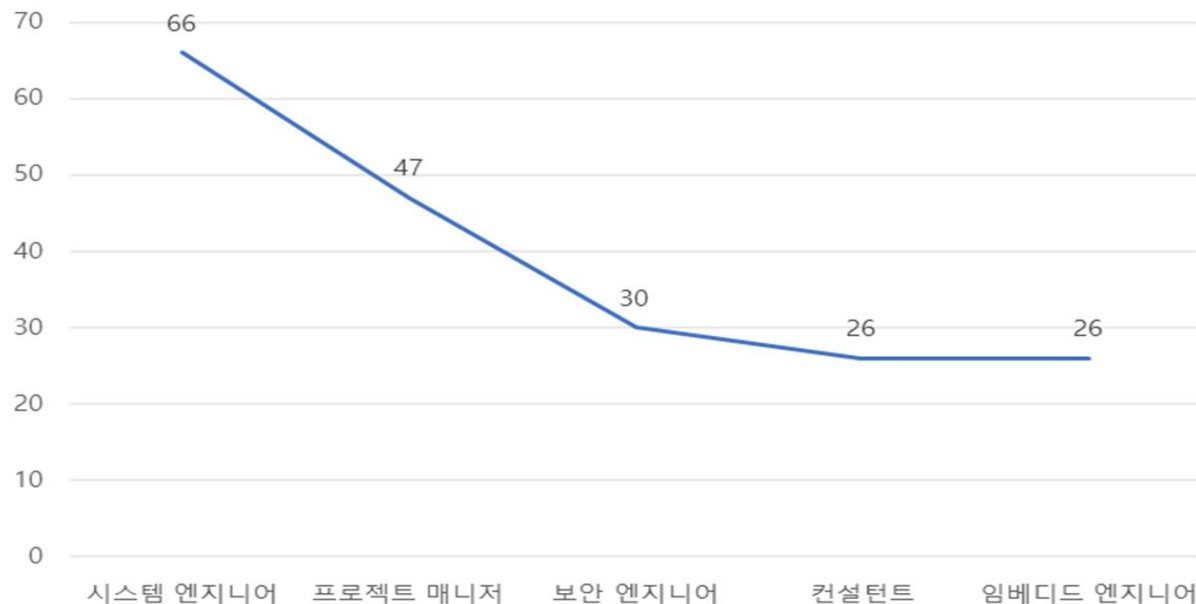
Front-end

Front 와 Back 라는 단순한 기준으로는
공고를 잘 내지 않는 것으로 분석됨



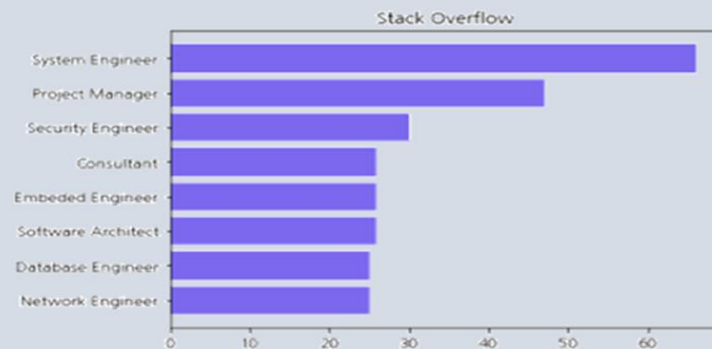
스택 오버 플로우 직무 분석 (NCS)

직무 (NCS 기준) 공고 수



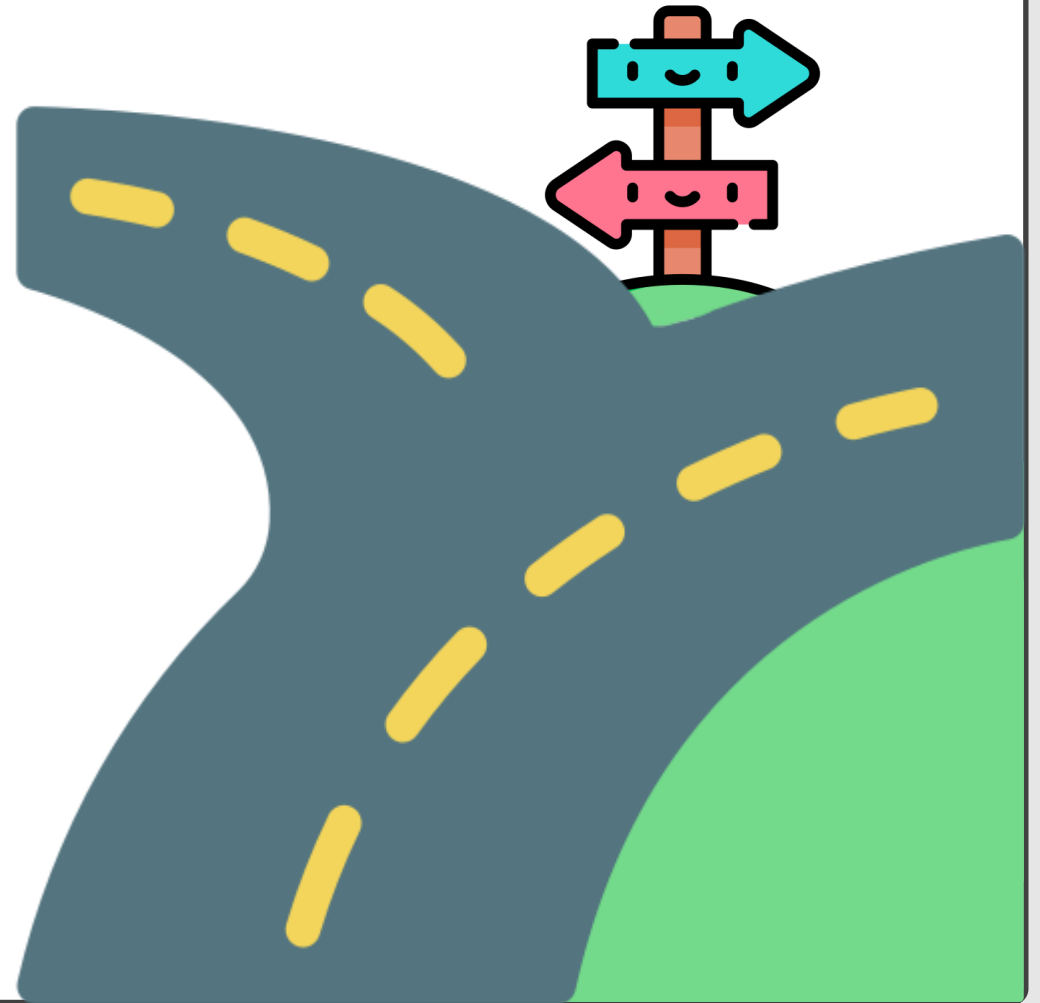
System Engineer

시스템 엔지니어가 8개의 직종 중 66개의 공고 수로 가장 많은 수요가 있다는 것을 알 수 있었습니다.



결론

자신의 가치관에 따라
기대 연봉 / 직무를 결정
하는 것이 가장 중요!



IT 취준생을 위한 직무백서

1조 - 최은정, 김민지, 정민지

참고

애로사항

애로사항	원인	해결
User-Agent 에서 크롤링을 허락해주는 사이트 중 URL 에 페이지의 숫자가 나타나는 사이트를 찾기 어려웠다.	각 사이트에서 보안을 위해 URL 을 연속해서 크롤링하는 것을 막기 때문이다.	구직 사이트 (사람인, 잡플래닛, 자소서 닷컴 등) 에서 크롤링에 최적화된 사이트를 선정
크롤링 결과에서 중복 데이터가 발생	크롤링 사이트 내부 구조로 인해 마지막 페이지를 반복했다.	CSV 에서 중복 데이터 삭제



SITE

이름	참고 사항	URL
사람인	1000개의 기업 정보 크롤링	https://www.saramin.co.kr/zf_user/salaries/total-salary/list
스택 오버 플로우	직무에 따른 공고 크롤링	https://stackoverflow.com/jobs
중앙일보-[Data & Now] 취준생 86만명 역대 최고, 셋 중 한 명은 공시족	청년 비경제활동 인구의 취업시험 준비 분야	https://www.joongang.co.kr/article/24110097
KOSIS	산업별 취업자	https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1DA7E06S&conn_path=I3

