

파이썬 미니 프로젝트 기획서

1. 프로젝트 개요

프로젝트명 : 채용 분석 정보 플랫폼

1-1 팀원 구성

- 안태경 : 데이터 수집 및 전처리 담당(담당 역할 구체화) **확인 필요!!**

데이터 정제 파이프라인 구축 (중복/결측치 처리)

채용 공고 크롤링 및 API 연동

데이터베이스 설계 및 관리

- 홍영준 : 데이터 분석 및 시각화 담당(담당 역할 구체화) **확인 필요!!**

EDA(탐색적 데이터 분석) 및 통계 분석

기술 키워드 NLP 처리 (KoNLPy, Soynlp)

- 한명규 : **Streamlit** 웹 개발 담당 (담당 역할 구체화) **확인 필요!!**

Streamlit 기반 대시보드 UI/UX 기획 및 구현

인터랙티브 시각화 구현 (Plotly, Altair)(가정)

2.프로젝트 목표

“채용 시장 분석 과 커리어 의사결정 지원 플랫폼”구축

왜? : 현재 기업과 회사의 직종은 많지만, 사용자가 직접 분석할 수 있는 부분이라던가,

1.구직자 지원

- 기술 트렌드 분석 : 연도별 기술 점유율 변화 시각화 및 신규 기술 강조하여 제공
- 직무 추천 : 사용자 프로필 기반 유사 채용 공고 추천

2.교육 기관 지원

- 지역별 수요 대시보드 : 지역 채용공고 수 제공

3.기업 채용 최적화

- 경쟁사 분석 : 기술 스택 비교 분석

4.플랫폼 핵심 기능

- 실시간 데이터 갱신 : 실시간 채용 공고 업데이트
- 개인 관심 맞춤: 최근 관심 직무 기반 탭 생성 (가정)

3. 데이터 수집 계획

3-1 수집 대상 데이터

1. 수집 대상 사이트

- 사람인, 잡플래닛

2. 기본 정보 :

- 직무명, 회사명, 근무 지역, 고용 형태, 회사링크

3. 분석 정보 :

- 경력 요구 수준, 학력 및 자격요건 기준 분석

4. 공고 일정 & 분석용 데이터 :

- 공고 게시일 및 마감일

3-2 데이터 출처

1. 사람인

- URL : https://www.saramin.co.kr/zf_user/
- 수집 방법 : 웹 크롤링 (BeautifulSoup)
- 대상 데이터 : 포지션명, 회사명, 지역, 연봉, 경력 조건, 기술 키워드(본문 기반 추출), 공고 등록일/마감일

2. 잡플래닛

- URL : <https://www.jobplanet.co.kr/job>
- 수집 방법 : 웹 스크래핑
- 대상 데이터 : 기업 평점

4. 데이터 분석 방법론

4-1 전처리 목적

목표 : 중복 제거, 노이즈 정제, 형식 통일

- 결측치 처리 : 직무(null), 회사명(null), 근무지(null), 경력(null) 빈 값 삭제.
- 이상치 처리 : “해외”, “경력” 튀는 숫자 값 제외.
- 데이터 정규화 : 카테고리 정규화(라벨정리 및 통일), 직무명(같은 의미 다른 표현 통일)

4-2 분석 기법(가정)

목표 : 데이터 분석으로 가치 있는 인사이트 제공

4-2-1 탐색적 데이터 분석 (EDA)

- 지역별 채용 건수 분석

지역 컬럼을 이용해서, 어떤 지역에서 채용공고가 많은지 분석 가능.

- 경력 요건별 분포 분석

요구경력 컬럼을 이용해서, 신입 채용 VS 경력 채용 비율 분석 가능.

- 학력 요건별 분포 분석

학력 컬럼을 이용해서, 고졸, 대졸, 석사 이상 요구 비율 분석 가능.

- 기업 배지(인증) 여부 분석

배지 컬럼을 활용해서 우수기업, 채용대행, 인증기업 비율을 분석 가능.

배지가 'NULL'이면 일반 기업, 그렇지 않으면 특수 인증기업으로 분배 가능.

- 공고 마감 기간 분석

기간 컬럼을 분석해서, 마감 임박 공고, 오래 열려 있는 공고 분포 확인.

공고 기간을 날짜 데이터로 변환하여 히스토그램으로 표출 가능.

- 회사별 공고 수 상위 랭킹

회사 컬럼을 이용해, 공고를 가장 많이 올린 회사 도출 가능.

4-2-2 상관관계 분석

- 연봉과 직무의 상관 관계

목적 : 직무 유형(개발/기획/마케팅)별 연봉 격차 분석

- 기술 스택과 직무의 상관 관계

목적 : 특정 직무에서 자주 요구되는 기술 패턴 발견(ex: 우대사항 or 경험)

- 경력 수준과 직무의 상관 관계

목적 : 직무별 경력 선호도(신입 VS 경력) 분석

- 연봉과 경력 수준의 상관 관계

목적 : 경력 수준(신입/경력)이 연봉에 미치는 영향 파악

- 경력 수준과 직무의 상관 관계

목적 : 직무별 경력 선호도(신입 vs 경력) 분석

4-2-3 머신러닝 모델 개발

- 연봉 예측 모델

목표 : 주어진 채용 공고의 속성(경력, 직무, 지역, 학력 등)을 기반으로 연봉을 예측 함.

- 직무 분류 모델

목표 : 주어진 채용 공고를 자동으로 직무별로 분류 함.

- 구직자-직무 매칭 시스템

목표 : 구직자의 정보(경력, 기술, 학력 등)와 채용 공고의 정보(직무, 기술 스택 등)를 비교하여 매칭 함.

- 채용 시장 분석 및 트렌드 예측

목표 : 채용 공고 데이터를 분석하여 채용 시장의 트렌드(인기 직무, 요구 기술)를 예측하고 분석.

- 이상치 탐지 모델

목표 : 채용 공고 데이터에서 이상치(잘못된 연봉, 비정상적인 경력 수준 등) 탐지.

5. 웹 애플리케이션 구현

5-1 Streamlit 구성 계획

1. 메인 대시보드

- 기업별 채용 공고 수 시각화 : 기업별 공고 수를 시각화하여 활동을 비교
- 지역별 채용 공고 인터랙티브 차트 : 지역별 공고 수를 인터랙티브하게 분석.
- 최근 기업 채용 공고 순위 : 최신 공고를 기준으로 기업 순위 시각화.
- 연봉 범위 별 채용 공고 수 분석 : 연봉 범위에 따른 채용 공고 수 분석.
- 인기 직무/기술 스택 추이 : 연도별 인기 직무 및 기술 스택 변화 시각화.

2. 기업의 상세 분석 페이지

- 상세 정보 및 매출액 추이 제공 : 기업의 정보, 채용 현황, 매출 추이 제공
- 현봉 및 복지 정보 : 연봉과 복지 정보 제공
- 유사 업계 차이 비교 : 유사 기업과의 비교 분석.
- 과거 채용 트렌드 분석 : 과거 채용 트렌드 분석(키워드/경력 변화).
- 채용 경쟁력 점수화 : 경쟁력 점수 제공(유사 기업 대비).
- 구직자 평판/리뷰 분석 : 구직자 리뷰 분석 및 평판 제공.

3. 데이터 분석 결과 시뮬레이터

- 기업정보 클릭 시 상세 정보 제공 : 클릭한 기업의 상세 정보 제공.
- 머신러닝 모델 기반 매출액 추이 예측 : 매출액 추이 예측 모델 제공.
- 기업 특징 및 아쉬운점 제안 : 기업의 강점 및 개선점 제시.
- 채용 포지션 별 경쟁력 점수 제공 : 기술 스택과 시장 수요를 비교한 점수 제공.
- 구직자 지원 이력 분석 및 맞춤형 조언 : 구직자 이력을 분석해 맞춤형 조언 제공.

4-2 UI/UX 계획

목표 : 사용자가 플랫폼을 직관적이고 효율적으로 사용할 수 있도록 인터페이스 구조와 사용자 흐름을 설계하는 전략.

● 사용자 중심 디자인

초보자도 쉽게 이해할 수 있도록 간단하고 명확한 인터페이스 구성

주요 기능(검색, 필터링)까지 최소화

- 명확한 정보 구조

대시보드-> 채용 공고 리스트 -> 세부 공고 상세 페이지 흐름 설계
정보화 카테고리화(공고 정보/ 기업 정보)

- 시각적 구성 및 디자인

데이터 시각화(bar Chart, Pie Chart, Line Chart)를 이용해 트렌드 분석 제공

- 대시보드 중심 구성

메인 화면에 주요 지표(기업별 채용 수, 지역별 채용 분포 등) 요약
사용자가 원하는 통계 지표를 필터링하여 대시보드 커스터마이징 가능

- 반응형 웹 설계

메인 화면에 주요 지표(기업별 채용 수, 지역별 채용 분포) 요약
폰트, 버튼 크기 최적화 적용

- 사용자 맞춤형 UX요소

최근 조회한 채용 공고 최근 본 목록 기능 제공
선호 직무/지역 기반으로 추천 공고 자동 필터링

- 인터랙티브 UX 강화

공고를 클릭하면 **POP-UP**으로 상세정보 제공 -> 페이지 이동 최소화
드래그/드롭방식으로 차트나 테이블을 사용자 취향에 맞게 재배치 (가정)
채용 공고 비교 가능(여러 공고 선택 후 비교) (

- 개인화된 대시보드 제공

사용자가 관심 있는 산업군, 직무군을 선택하면 맞춤형 데이터 대시보드 자동 생성 .
알림 기능(선호 조건에 맞는 새로운 공고 발생 시 알림)

- 검색/필터 최적화

지역 필터, // 키워드 검색, 연봉 범위 필터를 **Streamlit** 등으로 구현
최근 검색어 자동 저장 및 추천

- 로딩 개선 / 사용자 친화적 피드백

데이터 로딩 시 “Loading...”표시
공고가 없을 때, 친절한 문구 출력

6. 일정 계획

일차	주요 작업 내용	예상 시나리오
1일차(16)	<ul style="list-style-type: none"> - 미니 프로젝트 구조 논의 및 예상 시나리오 - 수집 대상 사이트 구조 분석 - 웹 스크롤러 초기 버전 개발 - 수집 데이터 항목 리스트 확정 	팀 회의 후 역할 분배 완료 ↓ 사이트별 HTML 구조 파악 ↓ 크롤러 기본 뼈대 개발 시작
2일차(17)	<ul style="list-style-type: none"> - Daily meeting - 주요 사이트 데이터 수집 완료 - 크롤링 데이터 품질 검토 - 데이터 정제 파이프라인 구축 - mysql 데이터베이스 구축 및 데이터 저장 	데이터 수집 70% 이상 완료 ↓ 결측치, 중복 데이터 발생 ↓ 정제 로직 추가 적용 ↓ DB 구축 성공
3일차(18)	<ul style="list-style-type: none"> - Daily meeting - EDA(탐색적 데이터 분석) 진행 - 주요 지표 시각화(지역, 연봉, 경력별 분포) - 머신러닝 예비 모델 학습(연봉 예측, 직무 분류) - 시각화 결과 기반 Streamlit 연동 준비 	데이터 정리 완료 ↓ EDA 에서 지역별/연봉별 트렌드 도출 ↓ 간단한 예측 모델 학습 완료 ↓ Streamlit 연동 테스트 시작
4일차(19)	<ul style="list-style-type: none"> - Daily meeting - Streamlit 웹 기본 UI/UX 구축(페이지 라우팅) - 데이터 시각화 차트 삽입 - 머신러닝 모델 연동 - 북마크/스크랩 기능 초기 버전 개발 	Streamlit 대시보드 초안 완성 ↓ 모델 예측값 출력 가능 ↓ 북마크 기능 1차 버전 구현 완료
5일차(20~21)	<ul style="list-style-type: none"> - Daily meeting(최종 meeting) - 전체 기능 통합 점검 및 디버깅 - 오류 수정 및 UX개선 최종 발표 자료(PPT 제작) - 데모 시연 준비 및 리허설 	최종 점검 및 작동 구연 확인. ↓ 발표용 PPT 완성 ↓ 데모 리허설 1~2회 진행

7. 종합 기대 효과 및 발전 방향

7-1 기대 효과

구직자 측면

- 구직자의 효율적인 의사결정 지원
 - 실시간 지원 현황 분석 리포트 제공
 - 개인 맞춤형 경력 성장 경로 제시
 - 연봉 예측, 직무 추천 트렌드 분석을 기반으로 장기 경력 로드맵 제안
- 구직 활동 스트레스 감소
 - 정확한 정보 제공으로 “잘못된 선택”에 대한 스트레스 완화
- 채용시장 투명성 향상
 - 신뢰도 높은 데이터 제공으로 구직자 만족도 향상

기업 측면

- 브랜드 인지도 강화
 - 채용 플랫폼 내 기업 프로필 최적화
- 지원자 행동 데이터 확보 및 채용 광고 최적화
- 경쟁사 분석
 - 산업별 연봉/복지/채용 트렌드 비교 분석
- 우수 인재 확보 전략 수립
 - 데이터 기반 역량 매칭

플랫폼 측면

- 데이터 기반 서비스 고도화
 - 사용자 행동 데이터 분석을 통한 서비스 개선
- 광고 및 유료 기능 확장 기반 마련
 - 프리미엄 서비스(기업 우상위 공고), 타겟팅 광고(교육·스킬 서비스 연계) 제공
- 시장 분석 기반 비즈니스 모델 다변화
 - B2B 솔루션(채용 분석 대시보드)
 - 글로벌 확장(현지화 서비스) 추진

7-2 확장 가능성

1. 기술적 확장

- **AI 고도화**
 - 실시간 채용 트렌드 예측 (GPT-4급 대화형 인터뷰 코칭)
 - Deepfake 면접 분석: 표정·말투 평가 → 문화적 적합성 진단
- **블록체인 적용**
 - 검증된 경력/자격증을 NFT 기반 이력서로 관리 (위조 방지)

2. 서비스 영역 확장

- **Post-Hiring(채용 후)**
 - 온보딩·교육·성과 관리 통합 (기업 HRM 시스템과 연동)
 - 이직 리스크 예측: 재직자 데이터 분석 → 기업에 경고
- **글로벌 크로스보더**
 - 번역+비자+세금 통합 지원 (예: "한국↔베트남" 특화 채널)

3. 데이터 활용 확장

- **실시간 산업 리포트**
 - "5년 후 유망 직무" "지역별 임금 격차" 등 유료 데이터팩 판매
- **구직자 Lifetime Value 관리**
 - 20대 청년 → 중장년까지 커리어 전주기 플랫폼 화

4. 수익 모델 다변화

- **B2B2E(Business-to-Business-to-Employee)**
 - 기업이 직원 교육 비용 부담 → 플랫폼에서 스킬 업 크레딧 판매
- **Web3 연계**
 - 토큰 보상 시스템 (구직자 활동 → 채용 코인 적립 → 현금화)

5. 생태계 파트너십

- **교육 플랫폼 M&A**
 - 코딩 부트캠프 인수 → "교육→채용" 원스톱 서비스
- **공공기관 협력**
 - 실업률 데이터 기반 정부 지원금 연계 채용 프로그램

8. 참고 문헌 및 자료

- 채용 플랫폼 데이터 정책 : 사람인, 잡플래닛 공식 사이트
- Python 웹 크롤링 : BeautifulSoup, Selenium 공식 문서
- 텍스트 분석 : KoNLPy, soynlp 라이브러리
- 웹 개발 : Streamlit, Plotly, Altair 공식 문서
- AI 활용 : DEEP SEEK, CHAT GPT

9. 프로젝트 수행 후기

9-1. 어려웠던 점 & 해결 방법

- 안태경 : 프로젝트를 진행하면서 가장 어려웠던 점은 초기 아이디어 선정 과정이었습니다. 다양한 아이디어를 제시하긴 했지만, 관련 자료를 찾는 데 큰 어려움을 겪었습니다. 이를 해결하고자 'Gemini - Deep Research'를 활용해 자료를 찾아보기도 했고, 이를 통해 일반적인 웹서핑으로는 찾기 어려운 몇몇 사이트를 발견하기도 했습니다. 그러나 여전히 자료의 신뢰도나 적절성이 부족하다고 판단되어, 결국 해당 아이디어는 폐기하게 되었습니다.
이후 팀원들의 적극적인 아이디어 제안과 끈질긴 자료 조사 끝에 지금의 주제와 데이터셋을 구축할 수 있었습니다. 아이디어가 정해진 후에도 팀원들과 지속적인 브레인스토밍을 통해 다양한 기능과 방향성을 논의했고, 기존 수업에서 배운 여러 코드들을 응용해 구현해 나갔습니다. 그러나 구현 과정에서 코드가 지나치게 복잡해지고 스파게티처럼 꼬여 있는 점이 문제로 드러났습니다. 이를 해결하기 위해 ChatGPT를 활용하여 코드 리팩토링을 진행했고, 덕분에 더욱 깔끔하고 유지보수가 쉬운 구조로 개선할 수 있었습니다.
- 한명규 :
- 홍영준 :

9-2. 배운 점 및 느낀 점

- 안태경 : 확실히 아이디어가 중요한 시대라고 느꼈습니다.
- 한명규 :
- 홍영준 : 제가 생각했던 것 보다 훨씬 복잡했습니다. 처음 접해보는 프로젝트여서 많이 걱정되었습니다. 태경님과 명규님께서 많이 이끌어주신 결과 완벽하게는 아니어도 잘 따라갈 수 있었습니다. 프로젝트를 진행 하면서 두 분께서 많이 가르쳐주시고 알려주셔서 많이 배울점이 많아 도움이 되었습니다. 같이 프로젝트를 할 수 있어 영광 이었고, 두분께 감사를 표합니다.

9-3. 개선하고 싶은 점(Next Step)

- 현재 계획에 있는 부분을 많이 실행을 하지 못 하였지만 한반도를 시각화 하여 전국구 채용 공고수를 확인 or 최다 공고수 지역 등 비교분석할 수 있는 데이터를 출력하는것이 제일 개선하고 싶은 점 입니다.

10. Q&A 및 데모 시연