1. 미니 프로젝트 개요

1-1. 목표:

- 배운 파이썬 기술을 통합적으로 활용해 하나의 데이터 프로젝트를 완성
- 수집 → 저장 → 분석 → 시각화 → 웹앱 과정을 실습
- 본인의 이해도, 기술 습득 정도, 문제 해결 능력 등을 종합적으로 보여주는 결과물을 제작

1-2. 팀 구성:

- 팀 인원: 3명
- 팀별 자유 주제 선정 (샘플 주제 제공 가능)
- 프로젝트 결과 발표는 5일차에 진행

1-3. 기술 스택:

- 웹 스크래핑: `requests`, `BeautifulSoup`, `Selenium`
- 데이터 분석: `pandas`
- 시각화: `matplotlib`, `seaborn`
- 데이터 저장: `pymysql`, `sqlalchemy`
- 웹 앱 제작: `streamlit`
- 그 이외 : `scikit-learn` 등 머신러닝 관련 (선택)

1-4. 타임 테이블**:**

일차	핵심 목표	세부 활동 내용
Day 1 기획 및 데이터 수집 설계	주제 선정 및 역할 분담	- 주제 선정 - 프로젝트 목표와 범위 설정 - 필요한 데이터 항목 정의 - 데이터 수집 방법 설계 (웹스크래핑 전략) - 데이터 수집 코드 초안 작성 - 일일 성과 공유 및 1일차 피드백
Day 2 데이터 수집 및 전처리	수집 및 데이터 저장	- 데이터 수집 코드 완성 및 실행 - 정적/동적 크롤링 구현 - 초기 데이터 수집 및 저장 - 수집된 데이터 전처리(결측치,이상치처리) - pandas를 활용한 데이터 구조화 - 일일 성과 공유 및 2일차 피드백
Day 3 데이터 분석 및 시각화	분석 → 시각화로 연결	- pandas를 활용한 데이터 분석 수행 - matplotlib/seaborn을 활용한 시각화 구현 - 분석 결과 해석 및 인사이트 도출 - 일일 성과 공유 및 3일차 피드백
Day 4 Streamlit 웹앱 개발	웹 UI 구현 및 통합	- Streamlit 애플리케이션 구조 설계 - 기본 UI 구현 및 데이터 연동 - Streamlit으로 대시보드 구현 - 인터랙티브 시각화 요소 추가 - 사용자 입력 기능 구현 - 일일 성과 공유 및 4일차 피드백
Day 5 발표 및 피드백	발표 및 공유	- 애플리케이션 디버깅 및 개선 - 최종 테스트 및 사용성 점검 - 발표 자료 준비 - 최종 프로젝트 발표

1-5. 프로젝트 진행시 주의사항

1. 데이터 수집 시 웹사이트의 이용약관을 확인하고 과도한 요청을 보내지 않도록 주의
2. 데이터 수집이 웹크롤링이 필요 없는 경우에는 웹크롤링 대신 분석,시각화에 좀 더 기능 추가로 구현해야 합니다.
3. 실제 서비스에 적용 가능한 인사이트를 도출하는 데 중점
4. 코드의 재사용성과 가독성을 고려하여 개발 (코드의 품질)
5. 팀원 간 역할 분담을 명확히 하고 GitHub 등을 통한 코드 공유 활성화
    - **Source**의 **GitHub** 링크를 공유해 주세요**.**

6. 매일 진행상황을 기록하고 문제점과 해결방안을 문서화
    - 매일의 작업 내용을 **Notion** 또는 **Google Docs**에 기록하시고
    - 문서가 기록 되어 있는 링크를 공유해 주세요.

7. 프로젝트를 진행할때 최대한 **AI**를 많이 활용해서 진행 해주기 바랍니다**.**

- 시나리오 작성**,** 문서작성**,** 코딩**,** 화면설계 등

2. 미니 프로젝트 발표 템플릿 **(필수) - AI** 최대한 활용

2-1. 프로젝트 소개

- 프로젝트 제목
- 팀명 및 팀원 소개
- 주제 선정 이유 (문제의식, 동기 등)

2-2. 데이터 수집

- 데이터 출처 및 수집 방식 (웹사이트, API 등)
- 사용한 기술: requests, BeautifulSoup, Selenium
- 수집 과정 스크린샷 (가능하다면 포함, 선택)

2-3. 데이터 전처리 및 분석

- 주요 컬럼 설명
- 데이터 정제 및 가공 방법
- 분석 목적 및 접근 방식

2-4. 시각화 결과

- 주요 분석 결과 시각화 (matplotlib/seaborn 사용 그래프)
- 시각화 결과 해석
- 인사이트 도출

2-5. Streamlit 대시보드 구성

- 대시보드 전체 구조 소개 (UI/UX 설명)
- 주요 기능 (필터, 슬라이더 등)
- 구현 화면 캡처 (선택)

2-6. 프로젝트 수행 후기

- 어려웠던 점 & 해결 방법
- 배운 점 및 느낀 점
- 개선하고 싶은 점 (Next Step)

2-7. Q&A 및 데모 시연 (PPT 없이 시연만 하는 것도 가능)

- 실시간 시연 (웹페이지 링크)
- 질문 응답 시간

3. 미니 프로젝트 평가 및 배점표 **(100**점 만점**)**

3-1. 기본 항목 **(90**점 **)**

항목	세부평가요소	배점
1. 프로젝트 기획 및 주제 적절성	- 주제 선정의 타당성 - 주제 설명 및 문제 정의	10점
2. 데이터 수집 및 정제 - 웹크롤링은 필요에 따라 선택하실 수 있음	- 웹스크래핑 코드의 적절성 - 데이터 구조화 및 정제 노력 - requests,beautifulsoup,selenium을 활용한 데이터 수집	15점
3. 데이터 분석 및 시각화	- pandas를 활용한 분석 - seaborn/matplotlib 시각화 활용 - 인사이트 도출 여부	30점
4. 대시보드(웹앱) 구현	- 구성의 일관성 - 사용자 인터페이스(UI)의 직관성 - 결과 시각화 연동 - Streamlit을 활용한 웹앱 구현	20점
5. 발표 및 문서 정리	- PPT 및 문서화 자료의 완성도 - 발표 전달력 및 흐름	15점
기본 총점		90점

3-2. 가산점 항목 **(**최대 **+10**점**)**

항목	세부평가요소	배점
1. 추가 기능 구현	- 주제 선정의 타당성 - 주제 설명 및 문제 정의	+5점
2. 창의성 및 완성도	- 참신한 주제나 시각화 방식 - 깔끔한 UI/UX 디자인	+5점
가산 총점		10점

평가 기준

- 데이터 수집 및 데이터 전처리의 완성도
- 분석 결과의 정확성 및 도출된 인사이트의 품질
- Streamlit 앱의 사용자 경험 및 기능성
- 팀워크 및 협업 능력 / 발표의 전달력 및 질의응답 대응력

데이터 수집 가능한 사이트 추천 (참고용)

1. 한국어 기반 사이트

사이트명	설명	특징
공공데이터포털 https://www.data.go.kr/	정부에서 제공하는 각종 공공 데이터 주제 : 인구, 지역상권, 교통, 기후	CSV 다운로드 + 오픈API
서울열린데이터광장 https://data.seoul.go.kr/	서울시 공공 데이터 주제 : 따릉이, 미세먼지, 공공시설, 인구분포	CSV + JSON + API 제공
잡코리아 https://www.jobkorea.co.kr/ 인크루트 https://www.incruit.com/	채용정보 크롤링 + 텍스트 분석 (자격요건, 기술스택) 주제: 인기 직무 기술 분석, 요구 역량 워드클라우드	Selenium 활용 추천
네이버 블로그,쇼핑/지식 https://developers.naver.com/products/service-api/search/search.md	블로그글, 상품명,가격,리뷰,Q&A 등	텍스트 분석시 유용 다양한 검색서비스API

2. 영문 기반 오픈 데이터

사이트명	설명	특징
Kaggle Datasets https://www.kaggle.com/datasets	고품질 데이터셋 저장소	분석 중심 프로젝트에 유리
IMDB https://www.imdb.com/	영화 데이터 장르별 영화 평점/리뷰 분석 + 대시보드	영화, 배우, 평점 크롤링 가능
Indeed https://kr.indeed.com/?from=gnav-homepage	채용 공고 IT 직군 요구 기술 트렌드 분석	Selenium 활용 추천

3. 주제별 분류 리스트 추천

주제 분류	데이터 수집 사이트	추천 사이트/데이터 출처
금융/경제 분석	한국은행 경제통계시스템 (ECOS)	https://ecos.bok.or.kr/
	금융감독원 전자공시시스템 (DART)	https://dart.fss.or.kr/
	네이버 금융	https://finance.naver.com/
영화/엔터테인먼트 분석	영화진흥위원회 통합전산망	https://www.kobis.or.kr/
	TMDB API	https://www.themoviedb.org/documentation/api
	네이버 영화	https://movie.naver.com/
	인터파크 티켓	https://tickets.interpark.com/
	KMDB	https://www.kmdb.or.kr/eng/main
부동산/주택 시장 분석	국토교통부 실거래가 공개시스템	https://rt.molit.go.kr/
	한국부동산원	https://www.reb.or.kr/
날씨/환경 데이터 분석	기상자료개방포털	https://data.kma.go.kr/
	에어코리아	https://www.airkorea.or.kr/
뉴스/미디어 분석	네이버 뉴스	https://news.naver.com/
	다음 뉴스	https://news.daum.net/
	빅카인즈(뉴스 빅데이터)	https://www.bigkinds.or.kr/

파이썬 미니 프로젝트 기획서 (필수 작성)

1. 프로젝트 개요

프로젝트명

국내 영화 흥행 분석 및 예측 대시보드

팀원 구성

- 홍길동: 데이터 수집 및 전처리 담당
- 김철수: 데이터 분석 및 시각화 담당
- 이영희: Streamlit 웹 개발 담당

프로젝트 목표

- 영화진흥위원회 API와 네이버 영화 데이터를 활용하여 국내 영화 흥행 패턴 분석
- 장르, 배우, 감독, 개봉 시기 등 다양한 요소가 흥행에 미치는 영향 파악
- 사용자가 영화 정보를 입력하면 흥행 예상 관객수를 예측하는 인터랙티브 웹 애플리케이션 개발

2. 데이터 수집 계획

수집 대상 데이터

- 기본 정보: 영화 제목, 개봉일, 장르, 등급, 러닝타임, 국가
- 제작 정보: 감독, 주연 배우, 제작사, 배급사
- 흥행 정보: 일별/주간 관객수, 누적 관객수, 매출액
- 평가 정보: 네이버 평점, 관람객 리뷰 (감성 분석용)

데이터 출처

1. 영화진흥위원회 통합전산망 (KOBIS) API
    - URL: https://www.kobis.or.kr/kobisopenapi/
    - 수집 방법: API 호출 (requests 라이브러리)
    - 대상 데이터: 2018년~2024년 개봉 영화 (약 1,500편)
2. 네이버 영화
    - URL: https://movie.naver.com/
    - 수집 방법: 웹 스크래핑 (BeautifulSoup, Selenium)
    - 대상 데이터: 영화별 상세 정보, 평점, 리뷰

3. 데이터 분석 방법론

전처리 계획

- 결측치 처리: 중요 정보 누락 영화는 분석에서 제외
- 이상치 처리: 비정상적 흥행 패턴 (코로나19 시기 등) 별도 분류
- 데이터 정규화: 개봉 스크린 수 대비 관객수 계산 등
- 텍스트 데이터 전처리: 리뷰 형태소 분석 및 정제

분석 기법

1. 탐색적 데이터 분석 (EDA)
    - 개봉 시기별/장르별/등급별 흥행 분포 분석
    - 감독/배우 파워 분석 (출연작 평균 흥행 성적)
    - 제작사/배급사별 흥행 추이 분석
2. 상관관계 분석
    - 개봉 스크린 수와 최종 관객수 간의 상관관계
    - 네이버 평점과 실제 흥행과의 상관관계
    - 개봉 첫 주 관객수와 최종 흥행 간의 상관관계
3. 머신러닝 모델 개발
    - 알고리즘: Random Forest, Gradient Boosting
    - 입력 변수: 장르, 등급, 개봉 시기, 감독/배우 파워 지수, 스크린 수 등
    - 출력 변수: 예상 총 관객수

4. 웹 애플리케이션 구현

Streamlit 구성 계획

1. 메인 대시보드
    - 연도별/월별 영화 흥행 트렌드 시각화
    - 장르별/등급별 흥행 분포 인터랙티브 차트
    - 최근 개봉작 실시간 흥행 순위
2. 영화 상세 분석 페이지
    - 개별 영화 선택 시 상세 정보 및 흥행 추이 제공
    - 유사 영화와의 흥행 패턴 비교
3. 흥행 예측 시뮬레이터
    - 사용자가 영화 정보(장르, 개봉일, 주연 배우 등) 입력
    - 머신러닝 모델 기반 예상 관객수 예측 결과 제공
    - 흥행 성공 요인 및 개선점 제안

UI/UX 계획

- 반응형 레이아웃으로 PC/모바일 모두 최적화
- 직관적인 차트와 그래프 활용
- 필터링 기능으로 사용자 맞춤형 데이터 탐색 가능

5. 일정 계획

일차	주요 작업 내용	담당자
1일차	• 프로젝트 기획 확정 • API 키 발급 및 초기 데이터 수집 테스트 • 웹 스크래핑 코드 개발	전체 팀 홍길동 홍길동
2일차	• 데이터 수집 완료 • 데이터 전처리 및 정제 • 데이터베이스 구조화	홍길동 홍길동 김철수
3일차	• 탐색적 데이터 분석 • 시각화 작업 • 머신러닝 모델 학습 (선택)	김철수 김철수 이영희
4일차	• Streamlit 기본 UI 개발 • 데이터 시각화 연동 • 예측 모델 웹 연동	이영희 김철수 이영
5일차	• 웹 애플리케이션 완성 및 디버깅 • 최종 테스트 • 발표 자료 준비 및 발표	이영희 전체팀 전체팀

6. 기대 효과 및 활용 방안

기대 효과

- 국내 영화 산업의 흥행 패턴에 대한 데이터 기반 이해 증진
- 영화 제작/투자 의사결정에 활용 가능한 예측 모델 개발
- 파이썬 데이터 분석 및 웹 개발 실무 역량 강화

확장 가능성

- 해외 영화 데이터 추가를 통한 글로벌 비교 분석
- SNS 데이터 연동을 통한 영화 화제성 분석 추가
- 영화 마케팅 전략 추천 기능 개발

7. 참고 문헌 및 자료

1. 영화진흥위원회 API 문서: https://www.kobis.or.kr/kobisopenapi/homepg/apiservice/searchServiceInfo.do
2. Streamlit 공식 문서: https://docs.streamlit.io/
3. 파이썬 웹 스크래핑 가이드: https://www.scrapingbee.com/blog/web-scraping-with-python/
4. 영화 흥행 예측 관련 연구 논문: "머신러닝을 이용한 영화 흥행 예측" (홍길동 외, 2023)