**프로젝트 진행과정**

**2023년 05월 19일**

**과정명 : 기업프로젝트기반AI빅데이터 융합전문가 양성과정**

|  |  |
| --- | --- |
| 팀 명 | Log人 |
| 팀 원 | 정지용, 김만기, 김성진, 이가연, 김성동, 신호철 |
| 프로젝트 주제 | **로그데이터 및 관련 결합데이터를 수집, 분석하여 신규숙박업소 지점 선정을 위한 비즈니스 인사이트 도출** |
| 프로젝트 수행 방향 | * 웹크롤링: 월요일에 끝마침. * DB 구축: 수집 및 가공한 데이터를 모두 데이터베이스로 관리하고 데이터간 관계도를 만들어 정리 * 데이터 전처리: 수집한 데이터의 결측치, 이상치 등을 처리하여 분석에 적합한 형태로 가공, 데이터 간의 결합 및 필요한 변수 생성. 관광, 호텔, 크롤링 데이터를 적절히 결합하여 분석에 용이하게 데이터프레임 만들기. * 데이터 분석: 버스, 지하철 정류소 1km 개수 새는 코드 수정 후 다시 작동시킨후 최종데이터로 사용. 클러스터링 활용하여 지하철, 버스 정류장 및 관광지 밀도가 높은 지역 그룹 찾기. 크롤링 데이터 분석 밎 시각화, 다른 데이터와 결합. * 시각화: 서울 관광, 교통 데이터를 folium 이용 지도 시각화 밎 클러스터링 결과 시각화. 크롤링된 데이터 단어랭킹, 워드클라우드로 시각화. * 웹 구현: Django와 html등등의 툴을 이용해 분석 및 시각화 결과를 웹 서비스로 제공. Google Colab과 비슷한 UX를 구현하기 위해 crispy\_forms같은 여러 라이브러리를 사용하여 팀원들의 코드를 직접 웹사이트에 구현. 시각화를 위해 folium등 파이썬 라이브러리를 이용함. * 협업: Github, 구글 드라이브 등을 이용하여 데이터, 코드, 발표 자료 등의 공유 및 협업. * 발표: Gamma를 사용하여 발표물을 온라인으로 실시간으로 공유하며 작성. |
| 역할 분담 및 일정 | 조장: 정지용   * 비즈니스 인사이트와 결론도출: 전원참여 * 데이터분석 및 시각화: 정지용, 이가연, 김성진, 신호철, 김만기 * DB구축: 김만기 * 웹 구현: 김성동 * 발표물 제작: 전원참여 |
| 프로젝트 수행 도구 | * 데이터분석: python, pandas, numpy, matplotlib, seaborn, plotly, scikit-learn, powerBI, folium * 웹구현: vsc, html, css, javascript, Django, django.crispy\_forms, widget\_tweaks, django.dash, folium, pandas, matplotlib등 * DB구축: PostgreSQL * 협업: Github, 구글드라이브 * 발표: 파워포인트 |