**프로젝트 기획안**

**2021년 9 월 13 일**

**과정명 : 데이터 사이언스/엔지니어링 전문가**

|  |  |
| --- | --- |
| 팀 명 | 아이멀캠유 |
| 팀 원 | 팀장: 황수빈  팀원:여정문, 왕기영, 이새봄, 이주남 |
| 프로젝트 주제 | 서울시 구별 창업하기 좋은 요식업종과 장소 추천 알고리즘 학습용 데이터 구축 |
| 프로젝트 수행 방향 | * 수집시스템 구축 * 구별 음식점 현황, 임대시세 및 추정매출(상권), 폐업률 대비 개업률, 자치구단위 서울생활인구, 지역 인프라, 주민등록인구 및 가구유형, 등록외국인 오픈API 수집 * 웹 크롤링 : 인스타그램, 네이버블로그, 카카오맵, 네이버맵 * DB 스키마 설계   첨부한 ERD 사진 참고   * 데이터 현황 시각화   연도별 통계, 업종별 통계, 구별 통계 등   * 데이터 분석 예상   요식업종 및 장소 추천 알고리즘 |
| 역할 분담 및 일정 | ● 역할분담  1. 데이터 수집, 전처리, 시각화 및 웹 크롤링(전원)  황수빈: 서울특별시 일반음식점 인허가 정보 수집 및 전처리, 시각화  여정문: 서울특별시 일반음식점 인허가 정보 수집 및 전처리, 시각화  왕기영: 매출액, 개업률•폐업률, 임대료 데이터 수집 및 전처리, 시각화  이새봄: 자치구단위 서울 생활인구, 서울특별시 주민등록인구, 서울특별시 유형별 가구 데이터 수집 및 전처리, 시각화  이주남: 서울특별시 일반음식점 인허가 정보 수집 및 전처리, 시각화  ● 일정  9/8~ 9/10: 주제 선정 및 일정 수립  9/13~9/17: 웹 크롤링, aws 서버 구축 및 하둡, 스파크 설치/ 데이터 전처리  9/18~9/20: 데이터 시각화  9/21~9/25: 데이터 DB 저장  9/28~9/29: 발표준비 및 포트폴리오 제작  9/30: 프로젝트 발표 |
| 프로젝트 수행 도구 | ● 수집 기술 패키지  requests, beautifulsoup4  ● 빅데이터 수집, 저장, 분석 시스템  Apache Hadoop, Apache Spark  ● 데이터베이스  MySQL, ORACLE  ● 시각화 기술 패키지  Matplotlib, Seaborn |