|  |  |
| --- | --- |
| **팀명 : 도개교** | **일자 : 25/09/03** |

|  |  |  |
| --- | --- | --- |
| 팀원 이름(역할) | (09/03) 활동 내역 | 진행도 (전체 기준) |
| 김시온  (팀원) | 생성용 데이터 정리 및 포맷 정리, on-premise 환경 설정 및 databricks 연동 확인 | 20% |
| 향후 진행사항 : 데이터 생성 및 DB저장 |
| 김지민  (팀원) | 아니 이걸 크롤링 최적화(23일 -> 1일)를 성공하네, 하루종일 이것만 했으... 비동기 너무 어려운 것.  테이블 명세서 예제 데이터랑 컬럼만 적어놓음 | 24% |
| 향후 진행사항 : 테이블 명세서 타입이랑 길이 생각해서 정하기, flask 그리기  최적화 더 깎을 수 있으면 깎아보기, 수집해서 PostgreSQL에 적재하 |
| 박형진  (팀원) | 크롤링 코드 리뷰 및 병합, 크롤링 테스트 및 코드 수정, DB 리소스 생성 및 적재 시도 | 17% |
| 향후 진행사항 : 데이터브릭스-DB 연결, 데이터 DB 적재 |
| 송누림  (팀원) | Databricks에서 크롤링 코드 수정 및 1개직무로 테스트, DB 이관 시나리오 기획, VM 리소스 생성 및 DB 서버 설치, jdbc 방식으로 적재 시도 | 10% |
| 향후 진행사항 : 점핏 크롤링 완료, 브론즈 데이터 적재 |
| 이민지  (팀원) | 데이터 스케일링(이상치만 -> 전체, 필요 없어짐), IPF 기반 연령/성별/학력 확률 분포 계산 | 15% |
| 향후 진행사항 : 조건 설정해서 비율 맞춰서 나머지 feature 할당, DB 적재, 코드 정리 및 Databricks에 업로드 |

|  |  |
| --- | --- |
| **담당자 피드백** | 데이터브릭스 컴퓨트, 크롤링을 위한 클라우드 VM 은 많은 비용이 발생될 수 있는 리소스 입니다.  적절한 작업에 적절한 사이즈로 사용될 수 있도록 비용관리도 잘 부탁드립니다.  초반 작업 속도가 무척 빠르시네요. 역할별로 잘 하고 계신 것으로 판단됩니다.  점점 궁금해지네요. |
| **담당자 확인** | 임대건 |