수원시 후생복지 분석 매뉴얼

분석 보고서 작성자: 배준호 | 매뉴얼 작성자: 배준호

개요_

개발 기간

• 2021.05

활용 기술

• Spark SQL, IRIS DSL, IRIS Studio

활용 데이터

- 후생복지 참여자 명단 (2016~2020)
- 후생복지 참여 희망자 명단 (2021)

설명_

목적

• 수원시 후생복지 참여 희망자 중 이전에 복지 혜택을 적게 받은 희망자 우선 선별

상세 내용

- 한정된 복지 자원의 효율적 분배를 위해 희망자 선별 작업이 필요
- 과거 5년 참여 데이터를 기반으로 신규 참여 희망자 중 우선순위 파악
- 사용자가 보고서 기능을 통해 직관적으로 조건에 맞는 희망자 추출 가능

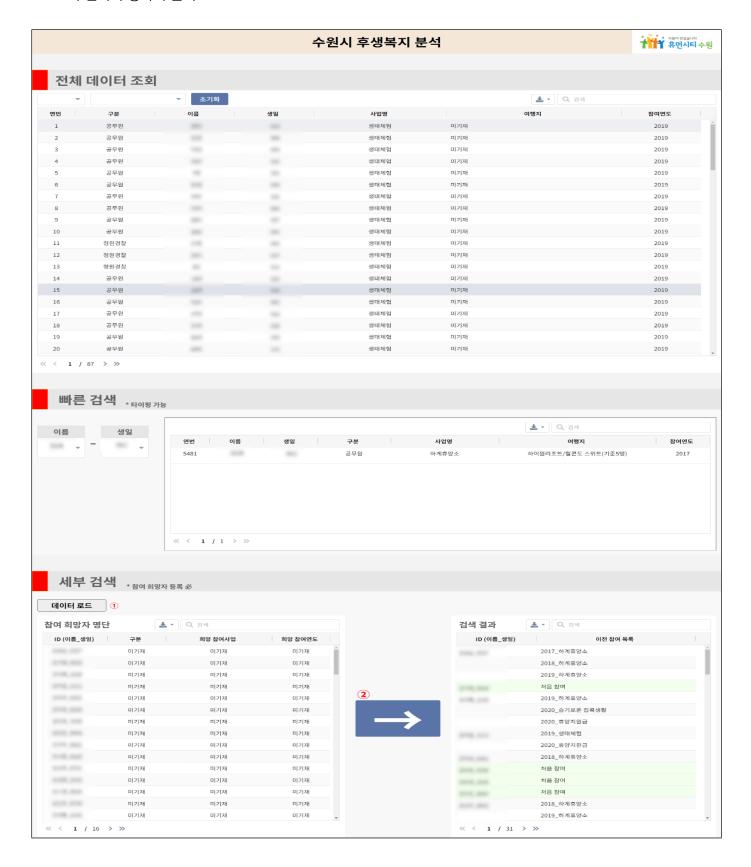
목차

- 1. 화면 구성
- Page.01
- 2. 사용자 매뉴얼
- Page.02~04
- 3. 유지관리 매뉴얼
- Page.05~07

1. 화면 구성

화면 1

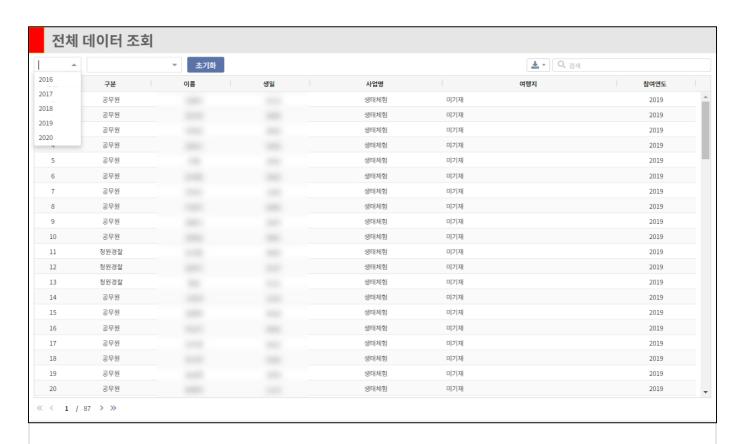
• 수원시 후생복지 분석



2. 사용자 매뉴얼

전체 데이터 조회

• 2016년 ~ 2020년 후생복지 참여 데이터



- 2016년부터 2020년까지 참여한 후생복지자원 명단을 테이블로 시각화
- 데이터에 포함된 순번 인덱스 기준으로 전체 년도를 표출하는 기본 테이블 구성
- 효율적인 데이터 파악을 위해 콤보 박스를 이용해 연도, 여행지 별로 필터링 가능하도록 구성
- 초기화 버튼 클릭을 통해 검색 후 기본 테이블로 전환 가능
- 검색 기능을 통해 이름, 직업 등 콤보 박스에 없더라도 각 칼럼에 존재하는 내용 검색 가능
- 우측 상단의 버튼을 통해 필터링된 데이터를 CSV 형식으로 해당 다운로드 가능

빠른 검색

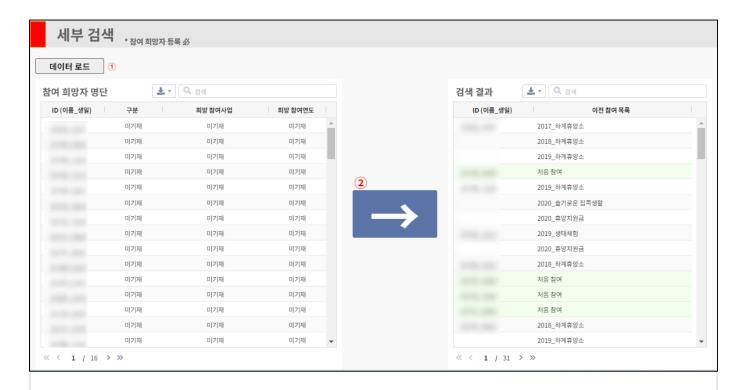
• 2016년 ~ 2020년 후생복지 참여 데이터



- 2016년부터 2020년까지 참여한 후생복지자원 명단을 테이블로 시각화
- 콤보 박스에서 이름을 선택하거나 타이핑하면 우측 테이블에 필터링된 결과가 표출됨
- 검색 기능을 통해 참여연도, 직업 등 콤보 박스에 없더라도 각 칼럼에 존재하는 내용 검색 가능
- 동명이인이 있을 경우 생년 선택을 통해 세부적으로 검색하여 필터링 가능
- 우측 상단의 버튼을 통해 CSV 형식으로 필터링된 데이터 다운로드 가능

세부 검색

• 2021년 후생복지 참여 희망자 데이터

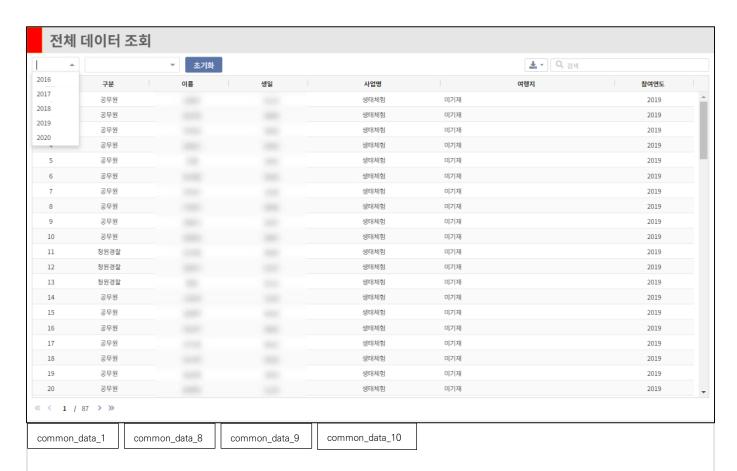


- 2021년 후생복지자원 신규 참여 희망자를 참여 희망자 명단 테이블로 시각화
- 반드시 참여 희망자 데이터를 미리 등록해야 하며, 기존 참여자 명단과 칼럼 개수가 같아야 등록이 가능
- 사용자가 참여 희망자 데이터 지표 등록 후 '데이터 로드' 버튼을 클릭해야만 최신 데이터가 참여 희망자 명단 테이블에 표출
- 동명이인을 구분하기 위해 참여 희망자 이름과 생일을 합쳐 ID 칼럼이 자동으로 생성
- 데이터 로드 후 화살표 버튼을 클릭하면 기존 참여 데이터를 참조하여 검색 결과 테이블이 표출
- 각 ID가 이전에 참여했던 목록이 나타나며, ID 하나의 결과값이 여러 개일 경우 모든 이력이 행으로 이어서 표시됨
- 이전에 참여 경험이 없는 ID의 경우 '처음 참여'로 표시되며 초록색으로 강조됨
- 우측 상단의 버튼을 통해 CSV 형식으로 필터링된 데이터 다운로드 가능

3. 유지관리 매뉴얼

전체 데이터 조회

common_data_1, 8, 9, 10



- 생년월일을 concat, substr로 정제하여 개인 식별을 위한 4자리의 생년월 파생변수 추출
- 여행지를 기재하지 않은 경우 모두 미기재로 표시하여 사용자의 검색을 도움 (fillna TOUR_PLACE 미기재)
- 사용자가 콤보 박스에서 참여년도를 선택하면 트리거 기능을 통해 테이블 갱신 (where PARTICIPATION_YEAR = '\${combo_4}')
- 세부 검색을 위해 사용자가 참여년도 \${combo_4} 선택 후에 추가적으로 콤보 박스에서 사업 명을 선택하면 트리거 기능을 통해 테이블 갱신 (where BUSINESS_NAME = '\${combo_5}')
- 최종적으로 사용자는 2016~2020 5년 간의 후생복지 참여자 전체 목록과 선택한 참여년도 별 목록, 해당 참여년도 내에서 선택한 사업명 별 목록을 확인 가능

빠른 검색

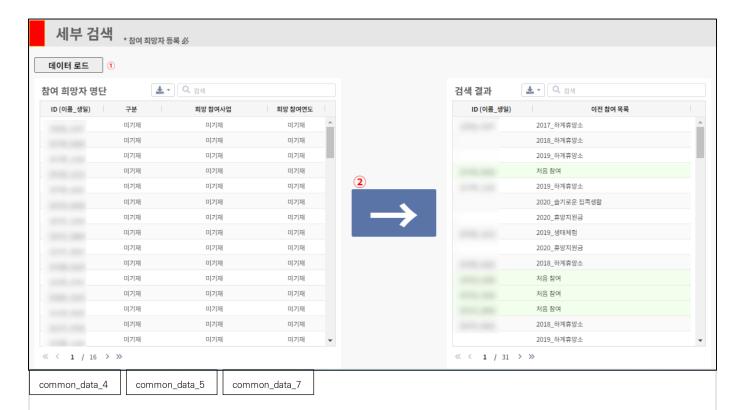
common_data_2, 3, 6



- 과거 5년 전체 참여자 이름을 오름차순으로 정렬하여 콤보 박스에 표시
- 사용자가 콤보 박스에서 참여자 이름을 선택하면 트리거 기능을 통해 테이블 갱신 (where NAME = '\${combo_2}')
- 동명이인에 대한 구별을 위해 사용자가 이름 \${combo_4} 선택 후에 추가적으로 콤보박스에서 생일을 선택하면 트리거 기능을 통해 테이블 갱신 (where BIRTHDAY ='\${combo_3}')

세부 검색

common_data_4, 5, 7



- 데이터 로드 버튼에 데이터 모델을 연동하여 사용자 클릭 시 캐시 초기화 구현 (* | command cache clear)
- 버튼을 누른다고 실제로 데이터가 업로드 되는 것은 아니며, 사용자가 참여 희망자 지표 등록을 하거나 기존 데이터를 갱신 할 시 프로세스가 바로 반영되지 않아 발생할 수 있는 데이터 로드 오류를 막기 위함
- 이름과 파생 변수 '생년'을 합쳐 개인을 식별한 수 있는 ID 칼럼 생성 (concat NAME, "_", BIRTH as ID)
- 사용자가 화살표 버튼을 클릭하면, 후생 참여 희망자 데이터와 기존 참여자 데이터에서 ID를 추출한 데이터 모델 뷰 '후생 참여자 ID'를 조인 (join outer '후생 참여자 ID' 후생 참여 희망자.ID = 후생 참여자 ID.ID)
- ID 오름차순으로 정렬한 뒤 SQL을 사용해 같은 ID를 하나의 Rank로 묶는 DIS 칼럼을 생성 (sql "select*, rank() over(order by ID) as DIS from angora")
- ID 오름차순으로 정렬된 기존 테이블에서 INDEX 순서로 숫자를 붙이는 NUM 칼럼을 생성 (numbering NUM)
- 같은 사람이 이전에 여러 번 참여 기록이 있을 경우 행에 연속해서 이름이 표출되기 때문에 이를 막기위해 NUM과 ID 별 랭크를 갖고 있는 DIS 칼럼으로 같은 이름을 판별 후 값을 공백으로 채우는 result 칼럼 생성 (case when NUM = DIS then ' ')
- 해당 result 칼럼은 NUM=DIS면 ''라는 공백 문자가 들어가 있으며, NUM과 DIS가 같지 않을 때는 Null 값을 가짐
- 위와 같은 조치는 IRIS DSL에서 concat이 칼럼에 Null을 포함하면 합쳐진 결과값이 Null로 나오기 때문에 NUM=DIS의 유효한 값 조차 사라지는 것을 막기 위함
- ID와 result를 합쳐 여러 값이 있을 경우 이름이 중복 표출되지 않고 공백으로 비워두는 RID 칼럼 생성 (concat ID, result as RID)
- 이전 참여 목록이 공백으로 존재하는 ID의 경우 이전에 참여 기록이 없는 ID이기 때문에 표출 후 테이블 시각화 기능으로 강조 표시 (fillna concated '처음 참여', RID'')