

대구 빅데이터 활용센터 이용 정책

2023. 02.

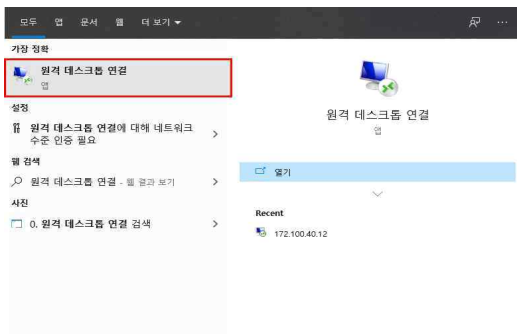
□ 분석 환경제공

- 대구지역내 공공·민간데이터를 활용 가능하도록 데이터를 수집하고, 전처리하여 이용자가 원활히 분석할 수 있는 데이터 및 인프라를 제공
- 대구시 산하 공사·기관, 연구소와 연계하여 사일로 형태로 쌓여있는 데이터를 발굴하고, 비식별 처리하여 데이터 거버넌스를 구축
- 활용 소프트웨어는 데이터 분야에서 주로 활용되는 파이썬, R 언어를 사용하고, 이용자의 수요가 있을 경우 별도의 사전 승인과정을 통해 프로그램 설치하여 지원
- 분석용 컴퓨터는 기본적으로 최대 15인까지 분석할 수 있는 환경을 제공하며, 딥러닝 학습 및 자연어 처리의 경우 별도의 워크스테이션 분석 환경을 제공
- 분석환경은 서버내 가상환경에서 분석이 수행되며, 분석 후 결과물에 한정되어 담당자 검토 후 자료 반출 가능

□ 데이터 이용

- 분석용 데이터는 원시, 원천데이터가 외부로 유출되지 않도록 폐쇄망 환경에서 관리가 되고 있으며, 이용자 데이터와 융합하여 분석이 가능
- 단, 분석 결과물은 원천 데이터가 포함되지 않도록 담당자의 보안절차를 거친 후 반출됨
- 분석 프로세스별 이용자가 분석에 어려움이 있을 경우, 분석요원이 이를 지원하여 분석과제를 수행할 수 있음

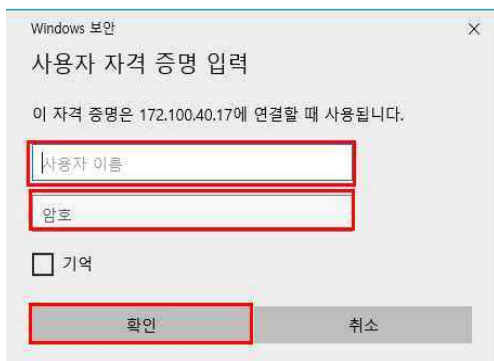
□ 분석실 사용자매뉴얼



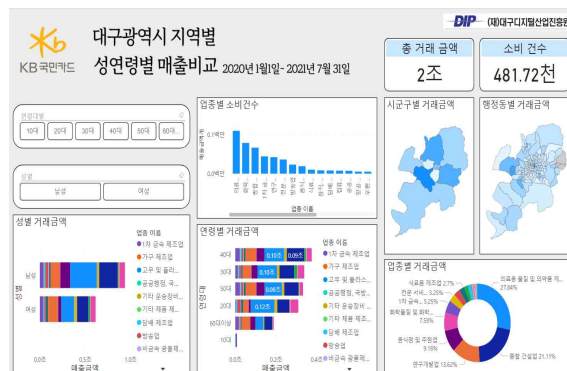
① 원격 데스크톱 연결



② 제공 받은 IP 입력



③ 사용자 계정 입력



④ 로그인 및 데이터 활용

※ 원격으로 접속하여 가상화 환경에서 분석시에는 외부 인터넷 사용이 불가함

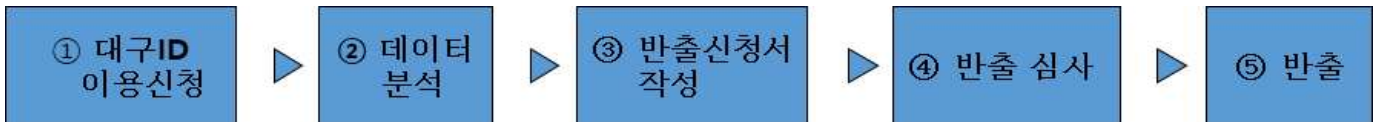
※ 가상 PC에 저장된 파일 반출 시

- ① 담당자에게 요청
- ② 담당자 검토 후 메일로 송부

□ 반출 기본 사항

- 센터내에서 분석을 완료하기를 권장함
- 반출신청서에 반출 데이터에 대한 정보(목적, 자료, 유형 등)를 상세하게 명시해야함
- 출처 표기 : 반출한 데이터를 활용할 경우 데이터의 출처를 표기
ex) 출처 : 대구시 빅데이터 활용센터, 데이터셋(이름, 기간)
- 민간데이터(통신사 생활인구, 카드 등)의 경우 공급자와의 체결한 협약과 계약 내용에 근거하여 반출기준이 설정됨

□ 이용 및 반출 절차



- ① 대구ID APP활용 이용 신청(신규회원은 ID 생성)
- ② 반출하고자 하는 데이터 확인 및 분석
- ③ 반출신청서 작성시 분석결과 요약서와 반출요청서의 서식을 다운로드하여 양식에 맞게 작성
 - 출처와 산출과정이 명확하지 않거나 내용 작성에 충실하지 않으면 재요청 및 반려
- ④ 반출 가능 여부 심사
 - 반출 및 자료제공 심사는 반출심사서에 따름
 - 내부 심사자가 분석 산출물과 자료제공을 분류하고 반출 및 제공가능 여부를 심사
 - 경우에 따라 1차 내부 심사 후 2차 제공처 심사 진행
- ⑤ 반출심사에 따라 반출

□ 반출 세부지침

- 데이터 반출의 종류는 크게 5가지로 구분되며 원본, 발췌, 통계처리, 시공간 결합, 시각화가 있음
- 이를 세부적으로 살펴보면 다음과 같음
- 첫째, 원본 데이터는 원시, 원천 데이터를 의미하며 통상적으로 공공데이터의 경우 개인정보 내용이 없을 경우 반출이 가능함
- 둘째, 발췌의 경우 원본 데이터와 유사한 데이터 보안수준을 가지며, 원본 데이터의 일부를 추출하여 반출하는 행위를 말함
- 셋째, 통계처리하는 데이터를 전처리하여 평균, 중앙값, 모형 구축 후 결과값에 대한 내용 등이 이에 포함되며, 중간단계의 데이터 보안을 요하는 수준임
- 넷째, 시공간 결합은 다양한 데이터를 융합하여 분석할 때 공간적으로 shp 또는 파이썬으로 데이터를 결합하여 새로운 데이터셋을 구축하는 단계로서 공간적으로 최소 동단위 이상일 경우 일부 데이터에서

데이터 반출이 가능함

- 여섯번째, 시각화의 경우 분석된 결과물을 이미지 형태로 반출하는 형태로 가장 높은 수준의 데이터 보안수준을 의미함
단, 이미지 파일은 통상적으로 널리 사용되는 jpg, gif, png 파일로 가능하며, 벡터형식의 데이터가 포함된 shp, tif 등의 파일형식은 불가함
- 그 외 가상화 환경에서 수행된 분석결과 보고서, 분석 코드 등은 반출이 가능함
- 보편적인 확장자로 반출신청 권장하며, 반출 심사시 데이터 확인이 어려운 경우는 반려함

□ [별첨 1] 용어 정의

반출신청	원본데이터 또는 분석결과물을 외부로 반출하도록 신청
원본데이터	분석산출물이 아닌 원시, 원천 데이터
분석산출물	데이터를 활용하여 도출해 낸 중간결과나 분석하여 작성한 소스코드, 데이터 및 결과보고서 등 문서, 분석산출물을 작성하는 방식은 발취, 통계처리, 시공간결합, 시각화로 정의
발취	원본데이터를 형태 그대로 일부분을 추출 처리
통계처리	데이터의 평균, 중앙값, 모형 결과 등의 분석 결과물 ※ 통신사 유동인구, 생활인구, 카드 매출실적 데이터의 통계처리 최소 단위는 동단위임
시공간 결합	시공간적으로 데이터를 결합하여 산출된 중간단계의 분석결과물 ※ 통신사 유동인구, 생활인구, 카드 매출실적 데이터의 공간결합 최소 단위는 동단위임
시각화	원본데이터를 활용하여 그래프, 지도 등으로 시각화한 그림파일 형태(png, jpg)의 분석결과물 *SVG,VML,HTML 등 확장자로 구현한 벡터방식 시각화는 화면을 캡처한 그림파일 형태(png,jpg)로만 가능

□ [별첨 2] 데이터 반출가능 여부

NO	데이터명	제공기관	원본	발체	통계 처리	시공간 결합	시각화	비고
1-1	신한카드 데이터	신한카드			○	○	○	
1-2	대구은행 고객데이터	대구은행			○	○	○	
1-3	대구은행 BC카드 결제 데이터	대구은행			○	○	○	
1-4	대구은행 지점별 방문 데이터	대구은행			○	○	○	
1-5	KB 국민카드 데이터	국민카드			○	○	○	
1-6	현대 카드 데이터	SK텔레콤			○	○	○	19·20년부터는 승인전수 및 매출단가 칼럼 추가
2-1	통신사 관광지 방문인구 데이터	SK텔레콤			○	○	○	행정동 이상의 데이터 반출 가능
2-2	맛집, 호텔, 명소 리뷰데이터	소셜네트워크					○	
2-3	대구시 관광지 데이터(블로그, 카페)	소셜네트워크					○	
2-4	대구 관광 리뷰데이터	소셜네트워크					○	
3-1	EDS 학교정보공시	한국교육학술정 보원			○	○	○	
4-1	사회경제지표, OD	한국교통연구원			○	○	○	
4-2	도로정보 및 교통량	한국교통연구원			○	○	○	
4-3	택시 승하차(TIMs)	교통안전공단			○	○	○	
4-4	대구시 버스 데이터	대구광역시	○	○	○	○	○	
4-5	대구 도시철도 수송 및 수입실적	대구 도시철도공사	○	○	○	○	○	
4-6	나드리콜 승하차 데이터	대구시설공단			○	○	○	
4-7	달성 행복 택시 데이터	달성군			○	○	○	
5-1	IT/SW 산업 실태조사	(재)대구디지털산 업진흥원			○	○	○	

5-2	대구 경북 ITSW 기업데이터	한국기업데이터			○	○	○	
6-1	통신사 유동인구	SK텔레콤			○	○	○	행정동 이상의 데이터 반출 가능
6-2	통신사 생활인구 데이터	SK텔레콤			○	○	○	행정동 이상의 데이터 반출 가능
7-1	매일신문 기사데이터	매일신문사			○	○	○	
8-1	행정구역 지리정보	국토교통부	○	○	○	○	○	
9-1	유동인구 소블럭 기준	SK텔레콤			○	○	○	
10-1	도서관 자료 데이터	대구시			○	○	○	
10-2	대구광역시 페이스북 페이지	소셜네트워크					○	
10-3	북구 민원 상담 자료	북구청					○	
10-4	한국소비자원 소비자 상담 자료	한국소비자원			○	○	○	
11-1	고용조사 데이터	한국고용정보원	○	○	○	○	○	
11-2	AI 데이터셋 - 보험료	한국교육학술정 보원			○	○	○	
11-3	AI 데이터셋 - 수학성적	한국교육학술정 보원			○	○	○	
11-4	AI 데이터셋 - 와인품질	한국교육학술정 보원			○	○	○	
11-5	AI 데이터셋 - 자전거대여	한국교육학술정 보원			○	○	○	
11-6	AI 데이터셋 - 타이타닉 생존자 예측	한국교육학술정 보원			○	○	○	
12	대구 빅데이터 마트 - 환경							
13	대구 빅데이터 마트 - 문화							
14	대구 빅데이터 마트 - 복지							
15	대구 빅데이터 마트 - 상권							