



통계자료 더이용제도 및 데이터 결합·활용 사례 안내

2022. 2. 9.

통계청 통계데이터기획과 이지영

목 차

1. 통계자료 더이용제도
2. 데이터 결합 사례(폐암환자 분석)
3. 통계데이터 분석 · 활용대회

통계자료 더이용제도 개요

참조 : 통계데이터센터 홈페이지(data.kostat.go.kr) > 공지사항 > 통계자료 더이용제도 가이드라인

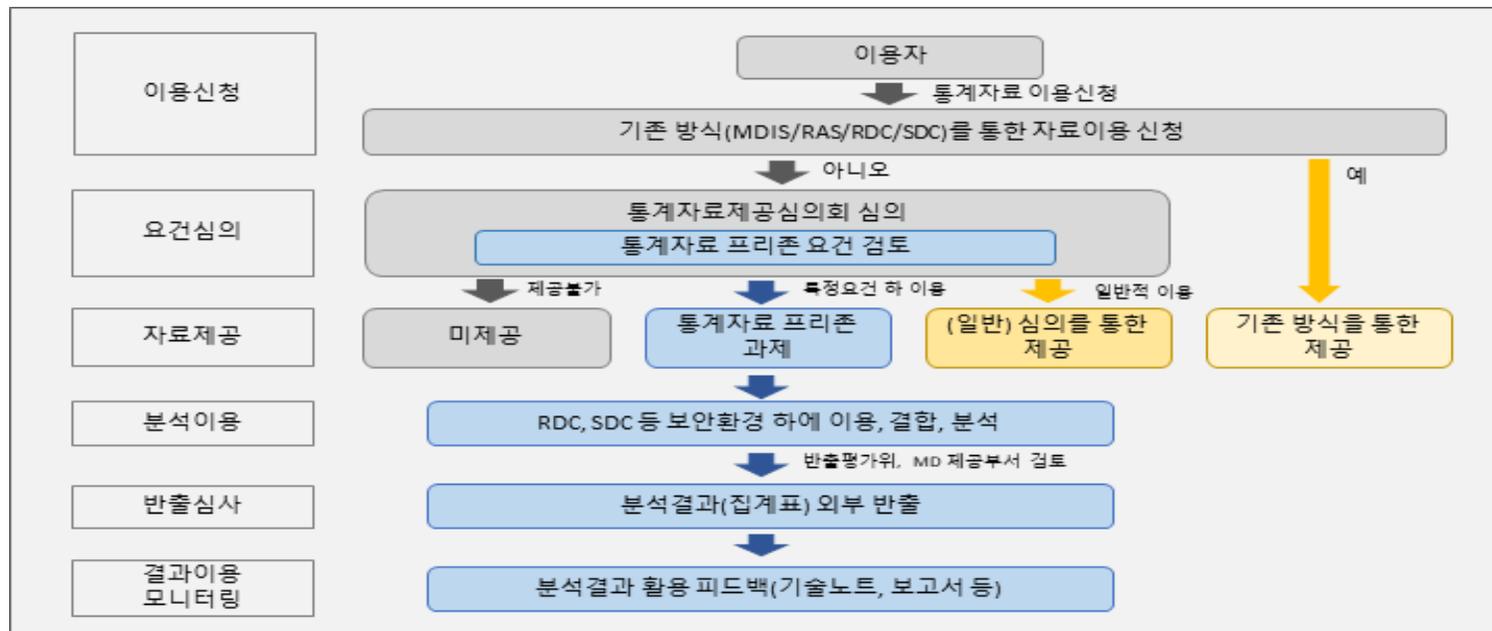
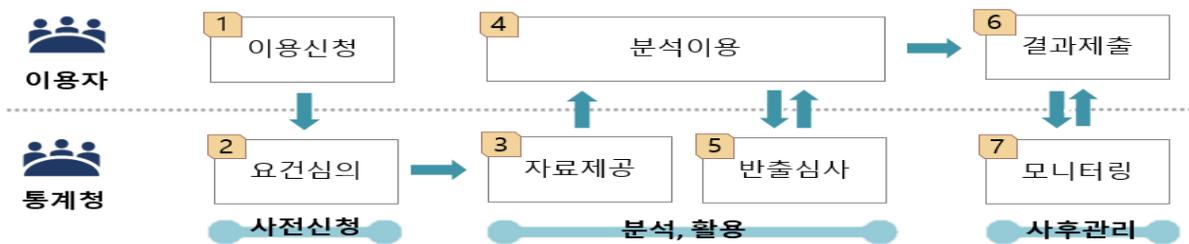
❖ 통계자료 이용의 장을 넓혀주는 규제개혁의 의미를 가진 제도로 이용자 우선 서비스(Customer Service) 일환으로 도입

- ✓ 소지역, 인구/가구, 사업체 특성 등에 따라 세분화된 자료, 서로 다른 데이터 셋을 연계한 자료 등 이용범위 확대
- ✓ 과학적 연구, 정부정책 근거 마련을 목적
- ✓ 통계자료제공심의회 심의
- ✓ RDC, SDC를 통해 안전한 데이터 이용(보안환경, 결과반출 심의)



통계자료 더이용제도 개요

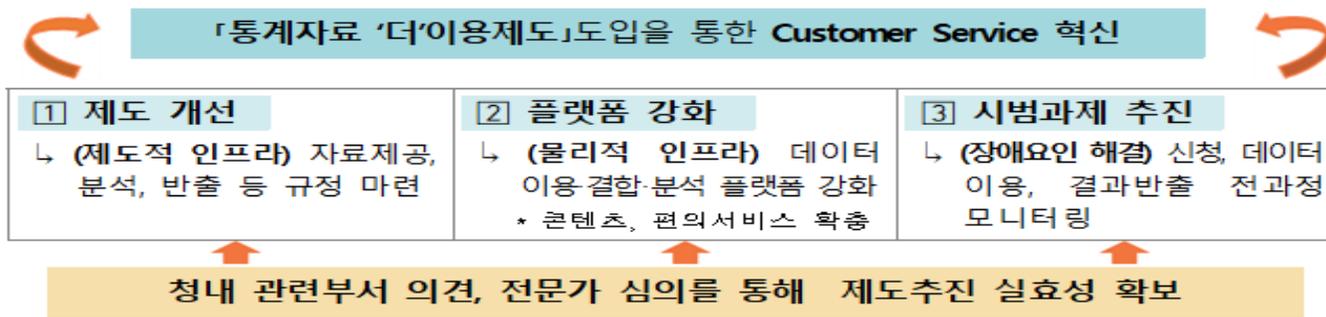
❖ 이용 절차



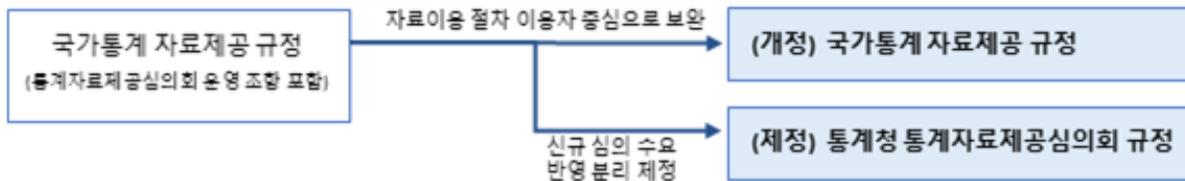
통계자료 더이용제도 개요

참조 : 통계데이터센터 홈페이지(data.kostat.go.kr) > 공지사항 > 통계자료 더이용제도 가이드라인

❖ 제도 도입을 위하여 ①제도 개선, ②플랫폼 강화, ③시범과제 추진



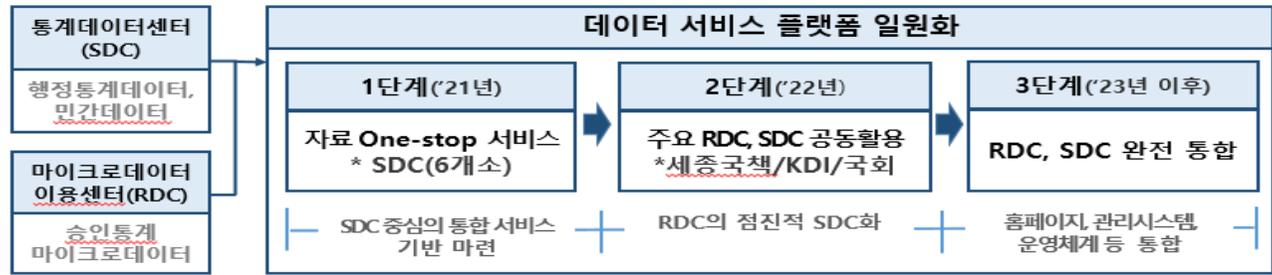
✓ ① 제도 개선 : 이용자 중심 통계자료 이용 절차별 준수사항 보완, 결합전문기관 운영, 통계자료 '더'운영제도 심의 등 신규 심의 수요 반영



통계자료 더이용제도 개요

참조 : 통계데이터센터 홈페이지(data.kostat.go.kr) > 공지사항 > 통계자료 더이용제도 가이드라인

- ✓ ② 플랫폼 강화 : RDC와 SDC 자료를 모두 이용할 경우 각각의 웹사이트에서 개별 신청 → SDC 웹사이트를 통해 ONE STOP 신청·분석하도록 서비스 개선



- ✓ ③ 시범과제 : 데이터 간 연계, 세분화 자료 이용 등 3개 시범과제 운영

시범과제	비고
1. 신혼부부-종사자DB 통합자료	행정통계, 등록부 간 연계(2.25.심의)
2. 소상공인실태 세분화MD 제공	설계시 공표단위 이외자료 제공(4.12.심의)
3. 가계동향조사 MD(가구연계키) 제공	연간소득 추정을 위한 연계키 제공(4.29.심의)

통계자료 더이용제도 개요

❖ RDC, SDC 설치현황(음영은 '21년 신규 개소)

▶ 국립암센터(RDC), 서울상암센터(SDC), 대구센터(RDC·SDC통합) 3개 추가 개소							
	수도권	충청권	동남권	대구·경북	광주·전남	전북·제주	강원
RDC (14개소)	국회도서관, 한국통계진흥원, 서강대, 서울대, 국립암센터	한국개발연구원, 세종국책연구단지, 통계센터	부산 통합 청사	대구 센터	광주 통합 청사	전북대, 제주대	건강 보험 공단
SDC (6개소)	서울중구센터, 서울상암센터	통계센터				* 음영은 '21년 개소	

• 마이크로데이터 이용센터(Research Data Center): 통계청 승인통계 마이크로데이터 제공
 • 통계데이터센터(Statistics Data Center): 행정통계DB 및 민간데이터 제공, 자료 간 연계 지원 목적

데이터 결합전문기관 개요

❖ 데이터3법 개정('20.1월) 이후 개인정보보호위 개인정보 유출 감독 사무 수행, 가명정보 결합을 위한 결합전문기관 지정

- ✓ 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의없이 가명정보 처리 가능
- ✓ 서로 다른 개인정보처리자 간의 가명정보의 결합은 보호위원회 또는 관계 중앙행정기관의 장이 지정하는 전문기관이 수행
- ✓ 결합전문기관 지정현황('21.11.1.기준)

<p>개인정보보호위원회</p> <ul style="list-style-type: none"> ▶ 통계청(2020) ▶ 삼성SDS(2020) ▶ (주)케이씨에이(2021) ▶ 한국지역정보개발원(2021) ▶ 롯데정보통신(2021) 	<p>과학기술정보통신부</p> <ul style="list-style-type: none"> ▶ 한국지능정보사회진흥원(2021) ▶ SK주식회사(2021) ▶ 더존비즈온(2021) ▶ 한국데이터산업진흥원(2021) ▶ 비씨카드(주)(2021) 	<p>보건복지부</p> <ul style="list-style-type: none"> ▶ 국민건강보험공단(2020) ▶ 건강보험심사평가원(2020) ▶ 한국보건산업진흥원(2020) ▶ 국립암센터(2021)
<p>국토교통부</p> <ul style="list-style-type: none"> ▶ 한국도로공사(2020) 	<p>교육부</p> <ul style="list-style-type: none"> ▶ 한국교육학술정보원(2021) 	<p>산업통상자원부</p> <ul style="list-style-type: none"> ▶ 한전KDN(2021)

가명정보결합 시범사례

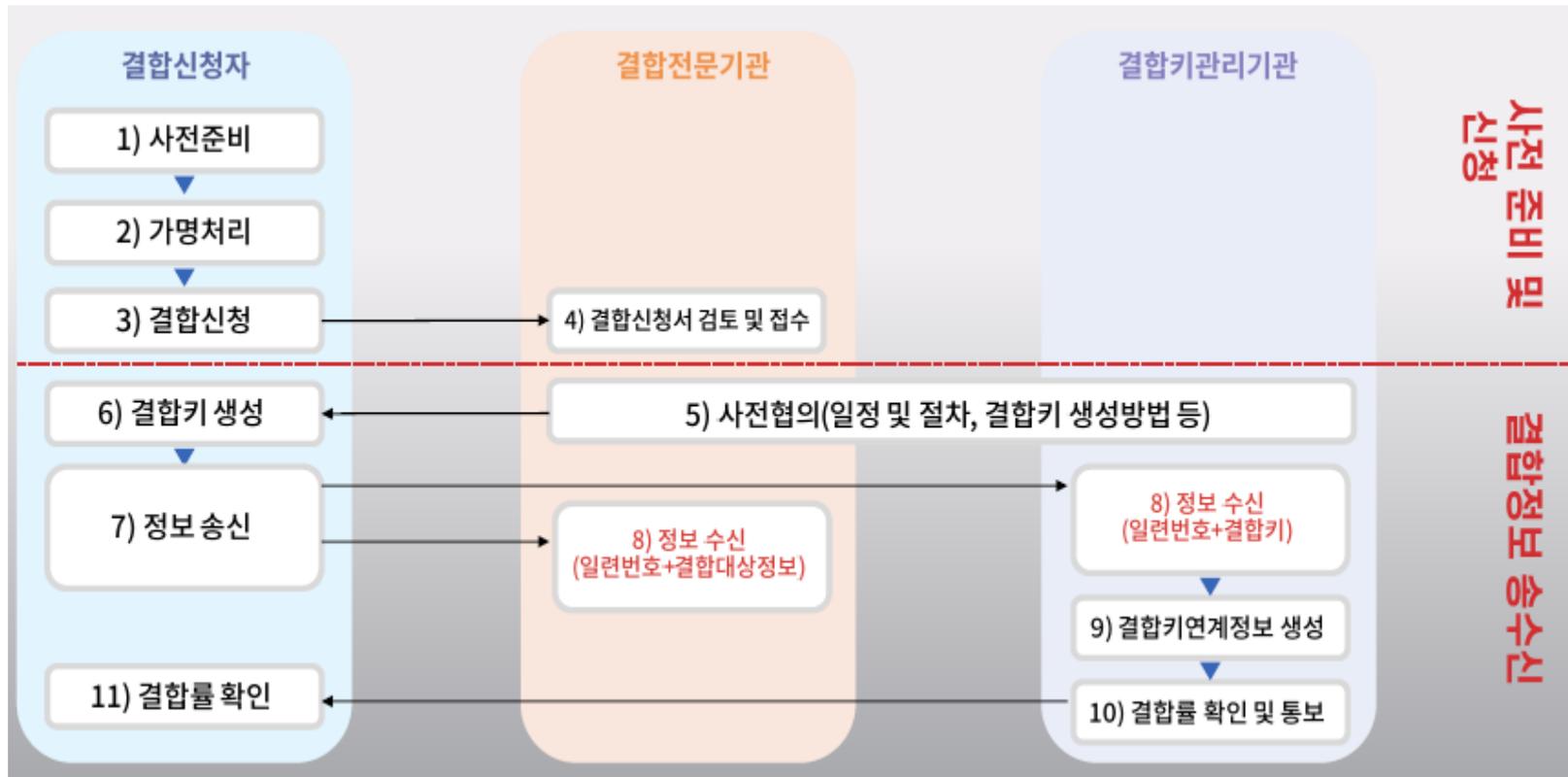
❖ 개인정보보호위에서 가명정보 결합 시범사례 선정

- ✓ 분야별 7건 중 통계청은 의료+인구분야 수행
- ✓ 결합데이터에 기반한 폐암 환자의 사망동향 및 사망예측 연구
- ✓ 통계청, 국립암센터, 건보, 한국인터넷진흥원, 개보위, 보건복지부 협업
- ✓ 가명정보결합 진행과정(11회 TF회의 실시)
- ✓ 결합대상 : 임상정보(암센터) & 진료정보(건보) & 사망정보(통계청)
- ✓ 결합키/결합률 : 성명, 성별, 생년월일 / 92.51%



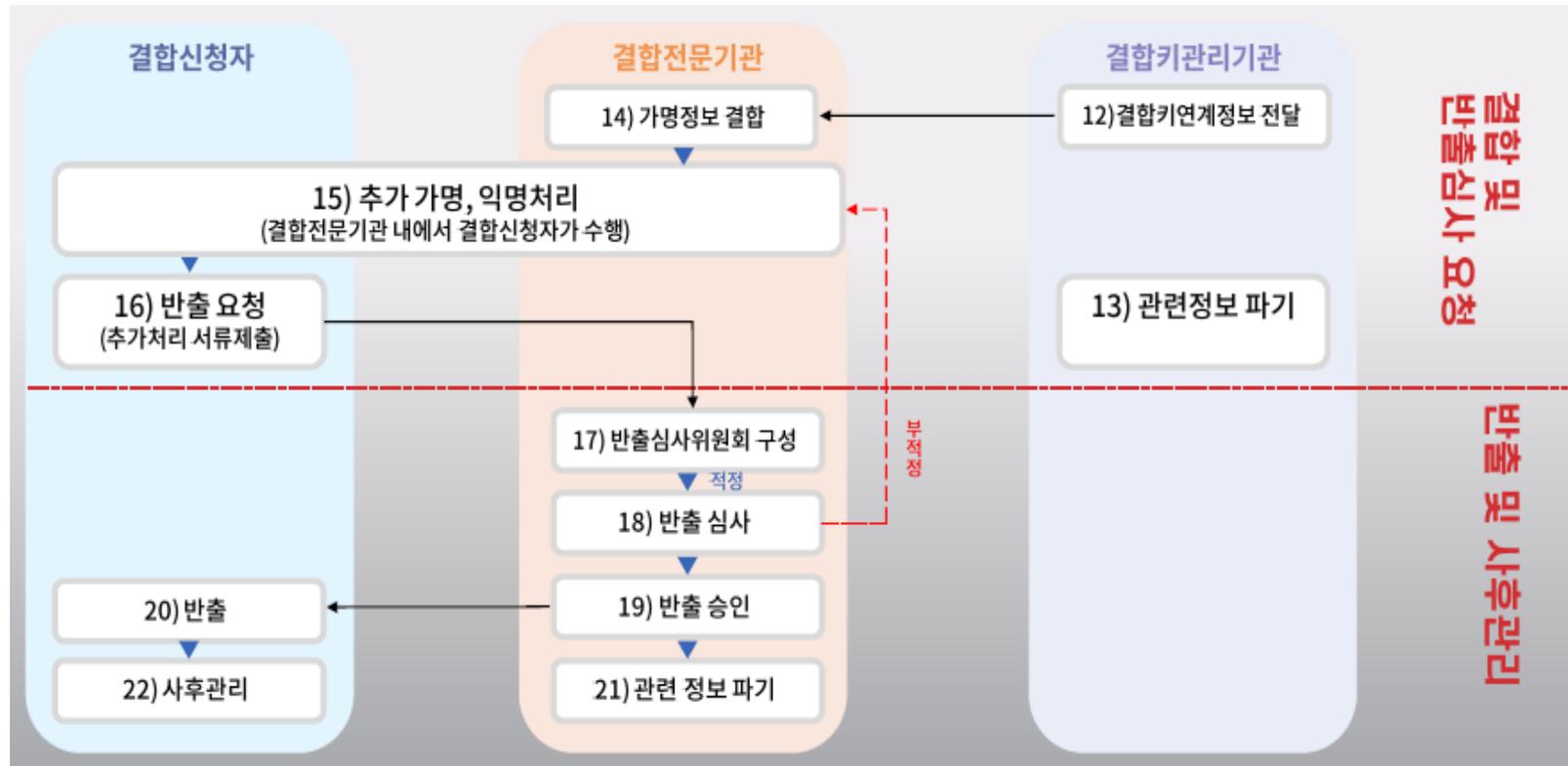
가명정보결합 시범사례 결합, 반출 절차

❖ 사전협의 및 결합정보 송수신



가명정보결합 시범사례 결합, 반출 절차

❖ 결합과 반출



가명정보결합 시범사례 성과 및 검토사항

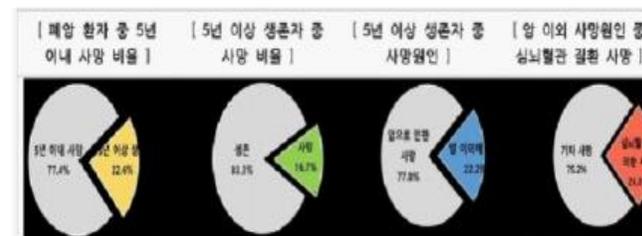
❖ 가명정보결합 첫 사례('20.5.28.)

- ✓ 최초의 결합사례
- ✓ 결합 모범사례 제시 ⇨ 데이터 활용도 제고
- ✓ 의료+인구데이터 결합 ⇨ 사망, 질환 예측모델 개발

❖ 검토사항

- ✓ 사전 협의과정 체계화 필요 * 결합신청 전 단계 상당 시일 소모(시범사례 4개월)
- ✓ 정형화된 협의 과정, 신청자에 대한 안내 필요
- ✓ 데이터 요청 시 제공 범위, 제공 가능여부에 대한 즉각적 답변
- ✓ 가명결합에 대한 데이터 담당자들의 이해
- ✓ 외부 심사위원 선정 과정 체계화

폐암 생존자, 심뇌혈관질환 관리
중요...가명정보 첫 결합분석



통계데이터 분석 활용대회 개요(2021년 제3회 기준)

참조 : 통계데이터센터 홈페이지(data.kostat.go.kr) > 분석활용대회

- ❖ **(목적)** 통계데이터를 활용한 다양한 연계 분석사례 발굴, 대중적 데이터 분석 활용성 제고, 통계데이터센터 활성화
- ❖ **(공모분야)** ① 데이터분석 보고서 ② 데이터분석 포스터
- ❖ **(신청자격)** 개인 또는 팀(3명 이하) 형태로 누구나 참여 가능
- ❖ **(활용데이터)** 통계데이터센터 제공자료 1건 이상 활용
 - * SDC 행정통계자료, 민간자료, RDC 제공자료, 공공데이터, 지도데이터 등
- ❖ **(일정)** 참가신청(5~6월) ⇨ 자료분석기간(6~7월) ⇨ 결과물제출(7월)
 - ⇨ 심사(8월) ⇨ 시상(9.1., 통계의 날)

통계데이터 분석 활용대회 개요(2021년 제3회 기준)

❖ (참가현황) 154개팀 신청 접수(보고서 132개팀, 포스터 22개팀)

- ✓ 최종결과는 보고서 54개팀, 포스터 11개팀 제출
- ✓ 통계데이터센터 방문 79개팀, 온라인 참여 75개팀

❖ (심사) 3차례 심사위원회를 거쳐 최종 선발

- ✓ (1차) 내외부 위원 2배수 선정 ⇒ (2차) 전문가심사(70%)+국민평가(30%) 수상대상자 선정
⇒ (3차) 내외부위원이 최종순위 및 상급 결정
- ✓ 대회 취지 부합성, 주제선정 중복성, 분석방법 다양성, 보고서 완결성 등 고려

❖ (시상내역) 13팀 선정, 상금 총 9백만원

구 분	데이터분석 보고서				데이터분석 포스터		
	대상	최우수	우수	장려	최우수	우수	장려
인 원	1명/팀	1명/팀	2명/팀	3명/팀	1명/팀	2명/팀	3명/팀
상 금	300만원	150만원	50만원	25만원	100만원	50만원	25만원
표 창	통계청장(7점) / 한국통계진흥원장(6점)						
부 상	통계데이터센터 24주 무료이용권 제공(수상자 전원)						

통계데이터 분석 활용대회 수상작(2021년 제3회)

❖ 데이터분석 보고서 분야 수상작(7개팀)

수 상	출 품 작	비 고
대 상	탄소저감을 위한 도시숲 최적입지 선정	팀(3)
최우수	넥스트노멀 시대 - 전통시장 DT 활용 방안	팀(3)
우 수	각 시도별 도시 경쟁력 지표 설립 및 이를 통한 시도별 인구이동의 분석과 이해	팀(2)
우 수	연령구조를 고려한 지역별 구급 출동 건수 예측	개인
장 려	강서구 소방 안전센터 최적 입지 선정	팀(2)
장 려	일회용컵 사용 절감을 위한 그린 뉴딜정책 개선 방안 제시	팀(3)
장 려	서울특별시 '여성안심홈세트지원서비스'정책 분배적절성 분석	팀(3)

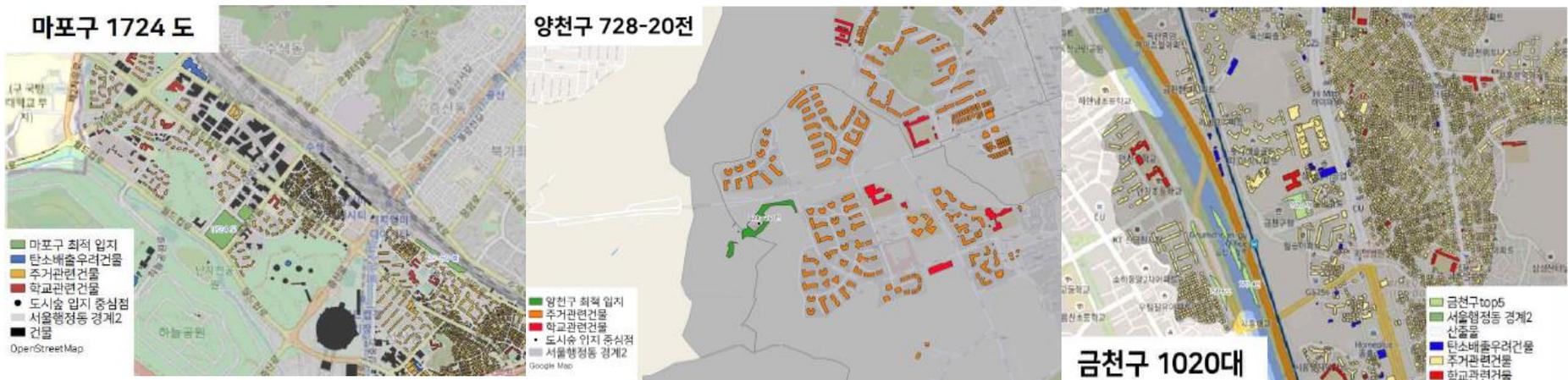
통계데이터 분석 활용대회 수상작(2021년 제3회)

❖ 데이터분석 포스터 분야 수상작(6개팀)

수 상	출 품 작	비 고
최우수	코로나19로 달라진 소비 패턴 - 연령별 분석	팀(2)
우 수	사회 주거 환경이 노인에게 미치는 영향	팀(2)
우 수	코로나 피해 취약계층 분석 및 지원 방안 제시	팀(3)
장 려	초미세먼지 멈춰	팀(3)
장 려	친환경자동차 수와 일반 자동차와 각 지역별 대기오염도 현황	팀(3)
장 려	저출산 및 고령화에 따른 다빈도질환 분석	팀(3)

통계데이터 분석 활용대회 보고서 분야 수상작(2021년 제3회)

- ❖ **출품명** : 탄소저감을 위한 도시숲 최적입지 선정 – 서울시를 대상으로
 - * 서울시 계획(2024년 까지 매년 1곳씩 3개 도시숲 조성)에 맞추어 도시숲 입지 선정
- ❖ **활용데이터** : SDC(자치구 평균연소득), 기타(토지소유, 매립지, 산업단지, 교통량 등), 18개월
- ❖ **데이터 분석** : 도시숲에 대한 후보지역, 공급지수, 필요지수(탄소배출, 탄소배출 외) 척도 이용
 - * 도시숲 후보지 클러스터링 → 입지선정 알고리즘 적용, 3개 자치구 선정 → 면적 및 인구를 고려하여 최종입지(마포, 양천, 금천) 선정
- ❖ **분석결과**



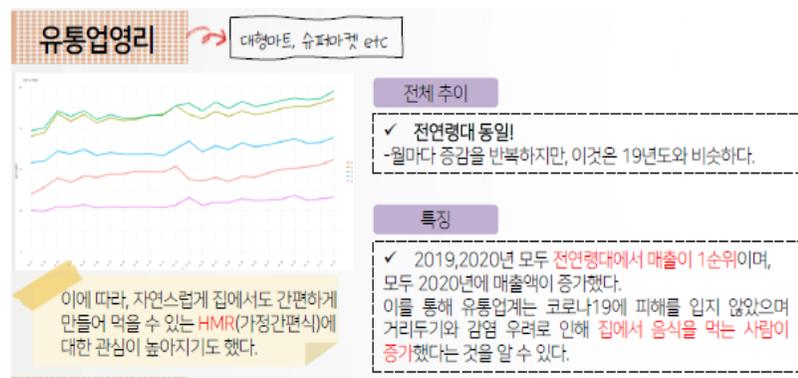
통계데이터 분석 활용대회 포스터 분야 수상작(2021년 제3회)

❖ 출품명 : 코로나19로 달라진 소비 패턴? 연령별 분석

* 코로나19가 한국인들 소비 습관에 끼친 영향력을 연령별로 분석

❖ 활용데이터 : SDC(성연령별 카드 매출), 기타(한국관광통계, 한국발 입국자에 대한 조치 현황), 2개년

❖ 분석결과





감사합니다