

통계의 창

WINDOW OF STATISTICS

2022.
SUMMER
VOL.29

디지털 전환 시대, 마이데이터를 주목하자

의료·교통, 마이데이터는 우리 삶을 변화시킨다 | 금융 마이데이터, 3년의 준비를 통해 서비스 본격화
국가데이터! 인공지능 시대를 연다 | 디즈니랜드 간다고? 난 「KOSIS 통계놀이터」 간다

CONTENTS

통계의 창
2022. Summer
Vol.29

발행일 2022년 5월 31일

발행인 최연옥

발행처 통계교육원

기획 김정란, 최희정, 김병준

주소 대전광역시 서구 한밭대로 713(월평동) 통계센터 통계교육원

전화 042-366-6151, 6152

팩스 042-366-6498

이메일 ccn420@korea.kr, joonk@korea.kr

디자인 및 인쇄 (주)피그마리온(02-516-3923)

ISSN 2005-1379

©2022. 통계교육원

※'통계의창'에 실린 내용은 필자 개인의 의견이므로 필자의
소속기관이나 본지의 공식적인 견해를 대변하는 것은 아닙니다.



1

통계 ISSUE

이슈 1

디지털 전환 시대, 마이데이터를 주목하자 004
배일권 | 4차산업혁명위원회 데이터기획관

이슈 2

의료·교통, 마이데이터는 우리 삶을 변화시킨다 012
유재경 | 컴패노이드 랩스 수석 파트너

이슈 3

금융 마이데이터, 3년의 준비를 통해 서비스 본격화 018
심현섭 | 한국신용정보원 빅데이터센터장

2

통계 FOCUS

통계청 탐방 1

국가데이터! 인공지능 시대를 연다 026
오승철 | 통계청 통계데이터기획과 사무관

통계청 탐방 2

디즈니랜드 간다고? 난 「KOSIS 통계놀이터」 간다 032
정승훈 | 통계청 통계서비스기획과 주무관

PEOPLE 1

통계와 사회가 만나다, 융합 교과서 「통계와 사회」 038
정문성 | 경인교육대학교 사회과교육과 교수

PEOPLE 2

각본 없는 드라마 속에 숨겨진 비밀을 캐낸다 044
조선미, 강지연, 이한울, 최윤철
| 명지대학교 스포츠기록분석연구센터

테마 속 통계 1

쏟아지는 여론조사, 왜 들쭉날쭉해 보일까 048
신정인 | 넥스트리서치 차장

테마 속 통계 2

스포츠 경기! 승리에 앞서 무엇이 중요한가 054
김정윤 | 웨슬리퀘스트 상무

테마 속 통계 3

통계를 활용한 지능화된 보험사기의 대응 064
김주환 | SAS Korea 이사

뉴스

통계로 바라보는 세상이야기 072
신동현 | 도서출판 지일박스 대표

3

통계 EDU

실습

통계를 만들고 분석할 수 있는 분석도구, 「통그라미」 따라하기 (시각화) 076
정승호 | 영남고등학교 교사

이론

분석은 합의(合意)이다 082
강양석 | Deep Skill 대표

01 디지털 전환 시대, 마이데이터를 주목하자

배일권 | 4차산업혁명위원회 데이터기획관



왜 마이데이터가 필요한가

최근 국내·외에선 데이터를 기반으로 모든 분야에서 디지털 전환(Digital Transformation)이 활발히 진행되고 있다. 이러한 변화의 과정에서 데이터를 얼마나 잘 활용하는지가 공공·민간은 물론 개인의 혁신 역량과 성과를 결정하는 주요 경쟁력이 되었다. 데이터는 ‘21세기의 원유’라는 말처럼 그 가치가 새롭게 인식되면서, 다양한 산업 분야에서 데이터의 확보와 활용에 많은 노력을 기울이고 있다. 하지만 ‘유용한 데이터’로 불리는 데이터들은 개인정보를 포함하고 있어 여러 규제를 적용받는 경우가 많다. 따라서 개인정보보호 관련 다양한 규제를 준수하면서도 데이터를 가치 있게 활용하는 방안이 사회적 논점으로 대두되고 있다.



2020년 8월, 데이터3법 개정으로 개인정보가 포함된 빅데이터를 활용할 수 있도록 하는 ‘가명처리 및 결합’ 제도가 도입되었다. 가명처리란 개인정보 일부를 삭제·대체하여 특정 개인을 알아볼 수 없도록 처리하는 것으로, 통계·연구 등의 정해진 목적 하에 선 사전 동의 없이 사용 가능하다. 개인정보가 포함된 빅데이터를 활용할 수 있는 할로가 열렸지만 아직 본격적으로 활용되고 있지는 못하다. 이는 가명처리의 본질적인 한계에 기인한다. 가명처리의 수준을 높이면 데이터 활용 가치가 낮아지고, 반대로 가명처리 수준을 낮추면 개인 식별 가능성이 높아지는 딜레마가 있기 때문이다. 또한 데이터를 활용하려는 정보보유 기관과 프라이버시를 중요시하는 정보주체·시민사회 간 이해가 상충하는 문제 또한 가명정보 활성화의 걸림돌이 되고 있다.

개인정보를 활용할 수 있는 또 다른 방식으로 최근

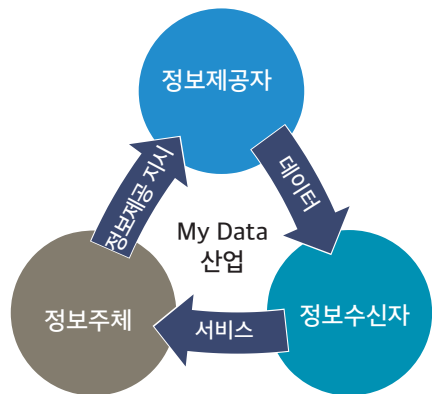
가장 주목받고 있는 방식이 ‘마이데이터’ 방식이다. 마이데이터를 통한 데이터 활용은 가명처리와 달리 정보주체의 능동적인 요청(적극적 동의)에 따라 이루어진다. 때문에 규제 이슈에서 상대적으로 자유롭고, 원본 데이터를 그대로 사용할 수 있어 활용 가치도 높은 것이 장점이다.

디지털 전환시대에서 데이터는 권력이며 힘이다. 마이데이터 제도는 데이터의 이동권을 의미하여 이는 권력의 이동을 의미한다고 할 수 있다. 이러한 권력의 이동은 기업, 정부로부터 정보주체인 개인으로의 이동을 의미한다고 할 것이다.

마이데이터가 구현되기 위한 핵심 원리는 정보주체의 「자료전송 요구권」에 기반하는데 이는 ① 정보주체가 ②정보제공자로 하여금 본인이 원하는 서비스를 제공받기 위해 ③데이터의 전송을 요구하는 권리로 정의할 수 있다.

[표1] 가명처리 방식과 마이데이터 방식 비교

구분	가명처리 방식	마이데이터 방식
①데이터 보유	기관	기관
②데이터 활용 결정	기관	정보주체
③개인정보 동의	동의 불필요	적극적 동의 기반
④활용 데이터	통계적 데이터 활용	원본 데이터 활용



자료전송 요구권 개념도

예를 들면, 기존엔 소상공인이 자금지원 신청을 위해 관공서에서 필요한 첨부서류를 발급받아 소상공인 진흥공단에 직접 제출해야 했으나, 마이데이터를 활용하면 소상공인이 관할부처로 하여금 필요한 행정 문서를 바로 소상공인 진흥공단으로 전달하도록 지시하여 첨부서류를 제출할 수 있다. 금융 분야에서는 은행, 카드, 보험 등 나의 금융거래 정보를 마이데이터 사업자로 전송토록 지시하여 이전된 데이터를 기반으로 내가 원하는 재무관리, 상품 추천 등 다양한 서비스를 받을 수 있다. 향후에 마이데이터가 전 산업으로 확산되어 나가게 되면 의료, 교육, 통신 분야 등에서도 이러한 내 정보 전송 지시를 통해 다양한 서비스를 받을 수 있게 된다.

마이데이터를 통한 데이터 전송요구권 제도의 도입은 정보 주체 측면에서 본인의 개인정보를 통제할 수 있게 한다. 그리고 산업 측면에서는 데이터의

자유로운 이동에 따라 소비자의 서비스 선택 폭이 넓어지고 서비스 질 개선과 가격 합리화 촉진 등 서비스 경쟁을 활성화한다. 또한, 다양한 신규 서비스 공급자의 시장참여의 기회가 제고되고 고객 수요에 부합하는 혁신적인 맞춤형 서비스 등장 등 '신규시장 창출'을 할 것으로 기대된다.

이렇듯 마이데이터라는 것은 데이터의 이동, 권력의 이동이기 때문에 정부가 먼저 나서서 제도를 만들어 주는 것이 핵심일 것이며, 마이데이터의 정부 정책 방향을 정확히 이해하는 것이 중요하다.

마이데이터 발전 상황은 5단계로 구분

마이데이터는 서비스 제공자가 관리하고 있는 내에 관한 데이터를 단순히 열람하거나 다운로드하는 것에서부터 다른 분야의 서비스 제공자에게 내 데이터를 전송하도록 요구하여 서비스를 제공받는 것까지 다양한 수준으로 구현될 수 있다. 정보주체의 권리 보장 및 서비스 구현체계를 기준으로 4차산업혁명위원회는 마이데이터 발전 단계를 0단계에서부터 4단계까지 총 5단계로 구분하였다.

0단계(조회)는 모바일 앱이나 홈페이지에서 나의 정보를 단순 열람할 수 있는 단계로, 마이데이터 도입 이전 단계이다. 개인정보 열람권은 일반법인 「개인정보 보호법」에 이미 반영되어 보장되고 있



으며, 열람 기능은 개인정보를 처리하는 대부분의 서비스에서 이미 구현되어 있다.

인터넷·스마트 banking에서 내 계좌 목록, 잔고 및 이체내역 등을 열람하거나, 쇼핑물의 마이페이지에서 내가 구매한 이력을 열람하는 것 등이 이 단계의 조 회 행위에 해당된다.

1단계(다운로드)는 모바일 앱이나 홈페이지에 접속 해서 내에 관한 데이터를 내려 받아 저장하는 단계로, 마이데이터 초기 단계에 해당한다. 이러한 데이터 저장 방식은 마이데이터 제도 이전부터 많은 사이트에서 구현되어 있기도 하였다. 인터넷·쇼핑물 등에서 Excel 파일 형태로 거래내역을 다운로드 하는 것이 이에 해당할 수 있다. 하지만 단순 열람·보관용이 아닌 데이터 이동을 위한 다운로드는 CSV, XML과 같은 비독점 표준 포맷으로 제공되어야 한다. 최근엔 이러한 비독점 표준 포맷으로 데이터 다운로드를 지원하는 경우가 점점 더 많아지고 있다.

2단계(전송요구)는 정보주체가 특정 서비스에 저장된 본인에 관한 정보를 다른 서비스로 전송하도록 요구할 수 있는 단계이다. 1단계처럼 정보주체가 직접 자신의 개인정보를 다운로드하여 다른 서비스에 업로드할 필요 없이, 사업자 간에 직접 전송을 지시하는 것이다. 2단계인 전송요구 단계부터 마이데이터의 핵심 원리인 '데이터 전송요구권'이 현실에서 본격적으로 구현되는 단계라고 할 수 있다.

3단계(대리활용)는 다양한 데이터 통합 조회·분석 서비스가 생겨날 수 있는 '대리활용' 단계라고 할

수 있다. 이 단계에서는 마이데이터 사업자들이 자신들의 어플리케이션이나 웹사이트 등을 통해 고객을 대리하여 다양한 데이터 보유자에게서 데이터를 전달받는다. 그리고 이렇게 모인 데이터를 대리자가 분석하여 여러 가지 고객 맞춤형서비스를 제공하게 된다. 대표적인 예로 최근 본격 시행된 금융 마이데이터제도를 통해 여러 마이데이터 사업자들이 제공하고 있는 금융비서, 자산관리 서비스가 있다.

4단계(전분야 확산)는 4차산업혁명위원회가 제시하는 마이데이터 발전 단계의 종착점으로, 전 산업 분야에서 분산된 나의 데이터를 내가 원하는 대로 자유롭게 모으고 이동시킬 수 있는 단계를 말한다.

마이데이터 확산을 위한 제도화, 기술·데이터 표준화 등은 각 산업별로 취급 정보의 민감도나 데이터 생태계 등이 반영된다. 따라서 초기에는 산업 분야별로 마이데이터 제도화가 진행될 것으로 전망된다.

우리나라 마이데이터 추진 현황

법제도 분야

우선, 법제도 분야에선 마이데이터를 구현하기 위한 데이터전송 요구권의 제도화는 일반법적 근거 마련과 산업 특성을 반영한 개별법 개정의 두 가지 방향(two track)으로 추진되고 있다.

개별법인 「신용정보법」, 「민원처리법」, 「전

0단계 : 조회	1단계 : 저장	2단계 : 전송요구	3단계 : 대리활용	4단계 : 전 분야 확산

[표2] 마이데이터 근거법 개정 주요 내용

법률	추진단계	주요내용
신용정보법	‘21.2월 시행	• 개인신용정보 전송요구권(제33조의2) • 금융분야 마이데이터 산업(본인신용정보관리업) 신설(제2조9의2) • 마이데이터 사업자 허가제 운영 및 의무 규정 (제4조, 제11조, 제22조의9 등)
민원처리법	‘21.10월 시행	• 민원인 요구에 따른 본인정보의 민원 처리기관 간 공동이용(제10조의2)
전자정부법	‘21.12월 시행	• 정보주체 본인에 관한 행정정보의 제공요구권(제43조의2)
개인정보보호법	개정중 (국회제출, ‘21.9월)	• 개인정보 전송요구권(제35조의2) • 개인정보관리 전문기관 지정(제35조의3)

자정부법」은 이미 개정되어 시행 중이다. 금융 분야에서 「신용정보법」이 개별법 중 가장 먼저 개정을 완료하고 2021년 2월부터 시행되어 금융 분야 마이데이터 제도 기반을 마련하였다. 공공 분야에선 정보주체가 공공기관이 가진 본인에 관한 행정정보를 제3자 기관으로 전송 요구할 수 있는 자료전송 요구권을 규정한 「민원처리법」과 「전자정부법」이 개정되고 각각 2021년 10월, 12월부터 본격 시행되었다.

이미 데이터전송 요구권이 법률 개정을 통해 반영된 금융, 공공분야 외에 마이데이터 전 분야 확산을 위한 법적 근거 마련은 「개인정보 보호법」 개정을 통해 추진 중이다. 이 법은 모든 분야에 적용될 수 있는 일반법으로, 개인정보 전송요구권 등을 반영한 개정안이 입법예고(‘21.1월~2월), 차관회의·국무회의를 거쳐 2021년 9월에 국회에 제출되었으며 현재 국회에서 논의 중에 있다. 전 산업분야에서 마이데이터 사업의 기반·동력이 마련되기 위해선 일반법인 「개인정보 보호법」의 개정이 조속히 완료되어야 한다.

비즈니스·산업·공공 분야

핀테크 업체 중심으로 추진되어온 마이데이터 비즈

니스는 금융에서 공공·의료분야로의 확대 및 제도화 과정 등을 거쳐 확산이 한창 진행 중이다. 앞서 설명한 마이데이터에 대한 발전 단계(0에서 4단계)를 기준으로 각 분야에 대한 마이데이터 진척 정도를 평가해보면 우리나라 각 산업별로 마이데이터의 발전 정도에 차이가 나는 점을 알 수 있다.

우선 마이데이터가 가장 먼저 시작된 금융 분야는 3단계로, 법률개정, 마이데이터 사업자 허가 등을 거쳐 2022년 1월부터 본격적으로 서비스가 실시되었다. 공공 분야는 「전자정부법」시행에 따라 본인에 관한 행정정보를 정보주체가 원하는 기관으로 이전시킬 수 있게 됨에 따라 2단계에 들어섰다고 할 수 있다.

통신 분야의 경우 정보주체의 요구가 있는 경우 요금납부정보를 금융 마이데이터 사업자에 일부 전송이 가능한 점에서 1.5단계로 평가된다. 의료 분야는 2021년 3월부터 서비스중인 「나의건강기록 앱」을 통해 정보주체가 진료이력, 건강검진이력, 투약·예방접종이력 등의 자료를 열람·다운받을 수 있다는 점에서 1단계 수준이나, 「마이헬스웨이 플랫폼」추진 계획을 발표하고 2~3단계로 나가기 위해 박차를 가하고 있다.

마이데이터의 분야별 추진 정도가 다른 것처럼 데이터를 주고받는 정보제공자·수신자 선정 방식, 주고받는 데이터 내용·표준, 데이터 보안 등 관련 기준과 절차도 분야별로 상이하다. 이에 따라, 단기적으로는 분야별 특성을 고려하여 마이데이터 활성화를 도모하되, 궁극적으로 분야의 장벽 없이 데이터가 자유롭게 흐르는 마이데이터 4단계를 지향하며 관련 법·제도를 마련해 나갈 필요가 있다.

마이데이터 발전을 위한 추진 목표와 전략

2021년 1월, 4차산업혁명위원회 내에 민·관 합동 데이터특별위원회가 발족되었다. 이와 동시에 4차위가 부처별로 분산 추진되어오던 데이터 정책에 대한 종합적인 ‘범국가 데이터 컨트롤타워’로서의 역할을 맡게 되었다. 데이터특별위원회 산하에 마이데이터분과도 신설되었고 산·학·연에서 모인 각계 전문가들이 마이데이터 분과에 참여함으로써 마이데이터가 나아가야 할 방향에 대해 심도 있는 논의를 시작하였다.

그 결과 2021년 6월, 4차위는 관계부처 합동으로 「마이데이터 발전 종합정책」을 발표하고 대한민국 마이데이터 발전의 청사진을 제시하였다. 4차위는 마이데이터 제도를 통해 정보주체에게는 자기결정권 확립과 데이터 기반의 다양한 서비스 제공을, 산업적 측면에서는 데이터 개방·활용을 통한 디지털경제를 촉진한다는 비전을 비전과 함께 아래와 같은 단계적 추진 목표를 수립하였다.

- (1) ‘21년~‘22년: 법제도 기반 마련, 사업 개시
- 「개인정보 보호법」 및 시행령 개정 추진
- 정보주체 인증, 전송, 저장·관리 등 전반에서의 정보보호·보안 기준을 확립
- 분야별로 준비가 마무리되면 사업을 시작하고, 초

기 단계에서 다양한 국민체감 서비스 제공에 주력

(2) ‘22~‘23년: 모든 분야 마이데이터 준비완료 및 분야별 활성화

- 구체적인 정보대상 범위, 진입수준 등에 대한 제도를 완성하고, 분야별 마이데이터 제도를 본격 시행
- 민간의 마이데이터 수요를 지속적으로 파악하고 발굴하여 마이 데이터 산업 활성화를 지원

(3) ‘24~‘25년: 데이터의 공유·분석과 광범위한 활용

- 데이터의 안전한 활용을 지원하는 기술개발(예: 동형암호) 및 적용 등을 통해 산업간 데이터 공유를 촉진
- 데이터 상호 연계·공유 시 안전한 정보 이전, 사고 시 책임주체 명확화 등 보안관리 강화

더불어, 4차위는 이러한 단계를 밟아 나가기 위해 몇 가지 추진 전략과 원칙을 제시하였다. 요약하자면 아래와 같다.

(1) 「개인정보 보호법」 개정을 통해 마이데이터 구현 상술했듯이 이미 법 개정이 완료되어 진행 중인 금융분야(신용정보법), 공공·행정분야(민원처리법, 전자정부법)를 제외한 나머지 분야는 법률 개정 없이 일반법인 「개인정보 보호법」의 개인정보 이동권을 적용받아 마이데이터의 법적 기반을 마련하게 될 예정이다.

(2) 마이데이터 거버넌스 구축

「개인정보 보호법」을 통해 제도 기반이 마련될 예정임에 따라 마이데이터 제도 시행을 위한 세부적인 사항은 「개인정보 보호법」 하위 규정에서 정하되, 관계부처 협의를 통해 구체화해 나갈 계획이다.

개인정보보호위원회는 법·제도를 추진해나감과 동시에 정보전송 체계수립을 위한 「마이데이터 표준화 협의회」를 우선적으로 출범(‘21.11.)시켰다. 개

[표3] 개인정보 보호법제하의 제도 마련 계획

개인정보 보호법 개정 (개인정보위, '22.上)	시행령 개정 (개인정보위, '22.下)	분야별 고시 제정 (각 부처, '22.下)
• 전송요구권 도입 • 개인정보관리전문기관(정보수신자) 역할 정립	• 정보제공자 지정 기준 • 정보수신자 지정 기준·절차, 관리·감독 • 정보제공 수수료	• 정보제공자 지정 : 매출액, 개인정보 규모 등을 고려한 분야별 대상 기관 • 정보수신자 지정 : 개인정보 관리 수준, 보안체계 구비 수준 등 평가 • 제공 정보 구체화 : 종류, 범위, 표준, 보안 절차 등 업계 의견수렴 진행

인정보보호위원회, 4차산업혁명위원회, 금융위원회, 교육부, 과학기술정보통신부, 보건복지부 등 마이데이터 관계부처 국장들이 참여하는 협의체를 통해 마이데이터 표준화 및 각 부처별 마이데이터 추진 내용에 대한 연계·협조가 이루어지고 있다.

(3) 개인정보 자기결정권 제고 및 데이터주권 확립
마이데이터 시스템 상에서는 정보주체의 전송요구권 행사에 따라 개인에 관한 상당히 많은 정보가 제3자에게 전달된다. 이러한 정보 전송이 마이데이터 사업자의 영리 추구만을 위해 이루어지거나, 개인이 원하거나 필요하지 않는 이상으로 전송요구권이 행사되지 않도록 하기 위한 정책들이 필요하다.

우선 정보주체가 전송되는 정보, 전송되는 대상 등을 명확히 인지하고 전송요구 할 수 있도록 '알고하는 동의'를 추진한다. 또한 정보주체의 전송요구뿐만 아니라 전송요구권 행사에 대한 철회, 그리고 이미 전송된 데이터에 대한 삭제 요구권도 보장토록 할 예정이다. 그리고 고객이 권리를 행사할 수 있는 정보의 범위가 지속 확대될 수 있도록 분야별 협의체 마련 또한 필요하다.

(4) 개인에 관한 모든 데이터 제공 원칙
정부는 정보주체의 요청에 따라 데이터가 자유롭게 이전되도록 제공 데이터 범위를 폭넓게 인정해나갈 방침이다. 다만, 정보제공자의 부담 또한 고려해야 하므로, 보유 개인정보의 규모, 산업별 특성

등을 종합적으로 감안하여 단계적인 개방을 해나가는 것이 바람직하다. 중·장기적으로 이종산업간 데이터가 전송·활용될 수 있도록 기술표준을 마련하고 데이터 규격화 등을 지속 추진해갈 예정이다. 또한 정보제공 대상 범위가 커질수록 정보제공자의 부담도 커질 수밖에 없으므로, 정부는 정보제공자(기업·기관)들이 공동으로 이용할 수 있는 중계시스템 등을 지원하는 한편 정보제공에 따른 적절한 보상·인센티브 방안도 설계해나갈 계획이다.

(5) 마이데이터 사업자 진입규제 최소화와 면밀한 사후관리
마이데이터 사업자가 되기 위한 자격, 인적·물적 요건들이나 심사제도 등에 따라 어떤 사업자들이 시장에 진입할 수 있는지에 대한 차이가 발생한다. 정부는 민간 기업 등 시장의 다양한 참여자가 정보주체가 필요로 하는 서비스를 다양하게 제공할 수 있도록 합리적으로 진입제도를 설계할 방침이다. 진입 규제가 낮은 만큼 보안이 소홀해지지 않도록 더 큰 관심과 노력을 기울여야 한다. 따라서 마이데이터 사업자에 대한 엄격한 보안 관리기준을 적용하는 등 촘촘한 사후관리 의무를 부과할 방침이다.

(6) 공공은 생태계 활성화에 도움이 되는 기반 조성에 주력
마이데이터는 데이터 분석을 통한 초개인화 서비스 등 혁신 서비스가 많이 발현될 수 있는 모델이다. 따라서 공공기관은 공익적 필요성이 특히 큰 분야, 인

[표4] 마이데이터 생태계에서 공공의 업무 영역(예시)

구분	역할
심사 및 사후관리 지원	• 보안 대책 컨설팅 및 사전 심사 지원 • 사후 점검·실사 지원 등
공익성 인프라 운영	• 인증서 발급·중계 등 통합 인증 인프라 운영 • 마이데이터 지원·관리 서비스 포털 구축·운영
의견수렴 및 표준화 추진	• 산업분야의 의견 조율을 위한 워킹 그룹 운영 • 데이터 및 API 표준화 방안 수립

프라 제공 등에 집중하고, 서비스 개발·제공 등의 비즈니스는 전문성이 있는 민간 기업이 수행토록 함으로써 데이터 경제 활성화에 기여할 수 있다.

(7) 주요기업의 마이데이터 공유와 협력
전 산업분야에 걸친 마이데이터 제도화는 정착되면 상당한 시일이 소요될 수밖에 없다. 따라서 정부는 법령 개정 이전이라도 주요 민간 기업들이 마이데이터 형식으로 데이터를 공유할 수 있도록 협력체계를 구축하겠다는 방침이다. 이를 통해 기업들의 마이데이터에 대한 인식을 제고할 수 있으며, 기업들은 사전에 데이터 표준화, 보안 확보 방안 등을 점검하고 구현하는 기회로 활용할 수 있다.

또한 공공성을 가지면서 기술·보안 전문성이 있는 한국인터넷진흥원, 금융보안원 등의 기관들을 통해 예비 사업자들에게 전문가 컨설팅을 제공하고 예산을 보조하는 등의 지원을 함으로써 대표 기업들의 마이데이터 참여를 유도할 계획이다.

**우리의 마이데이터 제도가
성공하려면**

마이데이터 생태계가 활성화되기 위해서는 무엇보다도 정보주체인 개인이 데이터 주권에 대한 인식을 확실히 가져야 한다. 내 데이터를 내 뜻대로 활용한다는 마음을 가지고 적극적으로 전송 요구권을

행사하여 스스로에게 맞는 서비스를 받고 맞지 않는 서비스엔 철회권이나 삭제 요구권을 행사하여야 한다. 데이터를 가진 기업들 또한 개인의 데이터를 기업이 소유한다는 관점에서 벗어나 개인이 주권을 가진 개인정보를 기업이 잘 이용하여 고객에게 좋은 서비스를 제공한다는 관점으로 접근해야 새로운 패러다임에 빠르게 적응하고 앞서나갈 수 있다. 마지막으로 정부는 관련 부처들이 서로 협동하여 신뢰할 수 있는 마이데이터 생태계를 구현하는데 필요한 제도 및 인프라 조성 역할을 충실히 수행해야 할 것이다.

마이데이터는 이제 막 걸음마를 댄 만큼 아직 가야 할 길이 무궁무진하다. 또한 이런 범국가적인 마이데이터 제도의 도입 및 운영은 다른 어느 나라도 아직 가보지 못한 전인미답의 길이므로 어떻게 흘러갈지 쉽게 예측하기도 어렵다. 하지만 데이터 기반의 디지털 전환 시대를 맞아 우리나라가 독보적으로 앞서가고 있는 데이터 활용 분야인 만큼, 잘 성공시킨다면 우리가 전 세계의 데이터 패러다임을 바꾸는 역할을 하게 될 수도 있다. 또한 마이데이터가 활성화된다면 국민들에게 플랫폼 기반의 선제적, 맞춤형 행정서비스를 제공할 수 있을 것이며, 관련 데이터 산업의 성장에 따른 일자리 창출뿐만 아니라 다양한 국가적 부가가치 창출을 견인해줄 것으로 기대한다.



의료·교통, 마이데이터는 우리 삶을 변화시킨다

유재경 | 컴패노이드 랩스 수석 파트너

마이데이터란, 문자 그대로 나의(My) 데이터(Data)를 의미한다. 마이데이터 사업은 데이터를 생산한 정보 주체가 '자신의 데이터(My data)'를 직접 관리하고 활용할 수 있도록 하는 것을 주요 골자로 하고 있다. 마이데이터 사업의 시행은 정보화 사회에서 4차 산업혁명으로 넘어오며 데이터 경제(Data Economy)가 급부상한 것과 관련이 깊다.



마이 헬스웨이 사업 도식화 (출처=보건복지부 보도자료)



모든 사람들은 아침부터 잠들기 전까지 수십 가지의 디지털 기기(Digital Device)와 디지털 서비스(Digital Service)를 사용한다. 모바일 기기로 서비스를 이용하며 발생하는 데이터에서부터 웨어러블 기기를 통해 축적되는 데이터, IoT 기기의 센서를 통해 축적되는 데이터까지, 수천만가지의 데이터가 실시간으로 쌓이고 있다. 지금까지 고객(사용자와 소비자를 통칭) 데이터는 정보 보호 혹은 프라이버시 침해 방지 측면에서만 논의되어왔다. 그러나 고객 데이터의 활용은 데이터 경제의 핵심 요소로, 데이터의 활용 주체의 권리에 대한 논의가 상대적으로 부족하다는 지적이 동시에 제기되어왔다.¹⁾ 이에 따라 EU에서 GDPR(The General Data Protection Regulation)을 제정하며 개인 데이터 활용에 대한 새로운 패러다임을 제시하였고, 이는 국내를 포함한 전세계 마이데이터 사업의 도화선이 되었다.

정리하자면 마이데이터 사업은 데이터 주권(主權, Data

Sovereignty)을 기업이 아닌 개인에게 돌려줌으로써 시장에서의 활발한 데이터 활용을 촉진하고자, 정부 주도로 진행하고 있는 새로운 변화의 시작으로 볼 수 있다. 그렇다면 우리는 어떠한 데이터를 활용해서 더 나은 시장과 공동체를 만들어낼 수 있을까? 의료와 교통 분야의 마이데이터 사업은 어디까지 와 있으며, 어디까지 변화할 수 있을까? 새로운 혁신을 이끌어내기 위한 선결(先決) 과제는 무엇일까?

의료분야 마이데이터를 통해 건강 상태를 진단하고 예방

의료 분야는 금융 분야에 이어, 마이데이터 사업이 가장 활발하게 논의되는 영역 중 하나이다. 20년도에 금융 분야와 함께 8대 주요 마이데이터 활용 분야에 선정되어 사업이 추진되고 있으며, 이후 실시한 수요조사에서 66개 기관이 실제 사업 참여 의사를 밝혔다.²⁾

2021년 2월부터 본격적으로 ‘마이 헬스웨이(My Healthway)’ 구축 사업과 ‘나의건강기록(PHR)’ 서비스 개발이 추진되고 있다. 마이 헬스웨이는 흩어진 개인의 의료 데이터를 한 곳에 모아 다양한 서비스에 연결하여 활용할 수 있도록 게이트웨이의 역할을 해주는 플랫폼이다. 개인은 정보 활용 동의를 통해 의료기관 진료정보(진료 기록, 약물 처방, 검사 결과, 상담 기록 등)와 개인 건강정보(맥박, 혈당, 생활 습관, 운동량 등), 공공기관 정보(건강보험, 예방접종, 날씨 등)를 한 눈에 확인할 수 있다.³⁾

이외에도 의료분야의 마이데이터 사업을 진행하고 있는 기업으로는 PHR(Personal Health Record) 상용화 플랫폼 ‘라이프 레코드’를 기반으로 사업을 확장하고 있는 **라이프 시맨틱스**와, 블록체인 기반 헬스케어 전문기업 **메디블록**, ‘페이퍼리스’ 서비스를 통해 병원과 약국, 카드사, PG사간 시스템 연동으로 진료 예약과 결제를 제공하는 **레몬 헬스케어** 등이 있다. 마지막으로 아래 표를 통해 의료분야 마이데이터 실증서비스들에 대한 정보를 확인해볼 수 있다.

[표1] 의료 마이데이터 실증서비스 요약 (출처=데이터산업진흥원)

서비스 명	서비스 설명	주관기관
헬스톡포미	• 병원의 건강검진, 문진 등의 건강 정보를 통합 관리 • 데이터를 활용한 맞춤형 건강관리 및 식단 추천 서비스 제공	연세대 산학협력단 (2019)
퍼스터 (Firster)	• 응급 진료기록 및 일상생활 속 건강기록 보관하여 진료와 처방에 활용 • 응급환자를 대상으로 제공하는 개인건강지갑 서비스	브이티더블유 (2019)
마이헬스데이터	• 스마트 폰으로 언제 어디서나 의료데이터 조회 • 의료데이터의 항목별 설명과 건강 등급 제공	서울대병원 (2019)
메디패스	• 원내 의료 정보 확인 및 서류없이 간편 보험청구 가능 • 병원 진료 현황 파악을 통한 예약, 대기 서비스 제공 • 복약 지도를 위한 알림 서비스 제공	서울대병원 (2019)
라이프월릿 (Life Wallet)	• 건강검진, 처방전, 유전자 검사, 장내미생물 검사, 구글 fit 정보, 혈당, 혈압 기록 등을 통합 관리 • 보장 보험, 장내 미생물 균형, 코디네이터 추천 등 맞춤형 서비스 제공	엔디에스 (2020)
하이엠디	• 검진결과, 유전체 검사 등 의료 데이터 조회 • 맞춤형 도시락 추천, 암 유전자 검사, 생활습관 및 정신 건강관리를 포함한 홈케어 서비스 제공	평화이즈 (2020)
아바타 빈즈 (Avatar Beans)	• 만성콩팥병 환자 데이터를 기반으로 식단, 운동 코칭 • 신약 개발 임상 시험 매칭 등 서비스 제공	서울대 산학협력단 (2021)
아이케어미 (iCareMe)	• 개인건강기록, 유전체 정보, 임상기록 등 데이터 활용 • 당뇨 등 만성 질환의 예방 및 관리	아이크로진 (2021)
메디팡팡	• 개인데이터 저장소(PDS)를 제공 • 서로 다른 병원에 존재하는 환자의 과거진료이력을 공유할 수 있는 진료기록 공유 서비스	에비드넷 (2021)

새로운 건강관리 방법의 등장

의료분야 마이데이터 사업은 “누구를 대상으로 어떤 의료데이터를 수집하여 유용한 기능 혹은 유의미한 정보를 제공할 것인가?”를 설계하는 것에서부터 시작한다. 수집하는 의료데이터는 병원에 저장된 상담·진료·치료 기록부터 웨어러블 기기를 통해 수집하는 생체기록 데이터 및 활동 데이터, 유전자 검사를 통한 유전체 정보 등이 될 수 있다. 더 나아가 금융과 같은 다른 분야의 데이터를 수집하거나 디지털 기기의 사용패턴을 수집하여 디지털 표현형(Digital Phenotype)을 통해 사용자의 건강 상태를 예측하려는 시도들도 많아지고 있다.

이러한 데이터 수집은 결국 사용자를 구체적으로 정의하고, 사용자의 맥락(context)을 정확하게 분석하여 개개인에게 맞는 심리스(seamless)한 건강관리 경험을 제공하기 위함이다. 예를 들면, 단순 진료

기록을 통해서만 사용자의 과거 질병의 치료 기록만 정보화할 수 있지만, 생체 데이터 혹은 행동 데이터를 추가적으로 해석하여 사용자의 이상 징후를 빠르게 포착하고 치료시기를 앞당길 수 있다. 더 나아가 궁극적으로 사용자의 유전체 정보를 포함한 다양한 데이터가 쌓이면, 예측 모델을 적용하여 특정 질병의 발병 시점을 근접하게 예측하거나 발병 확률을 계산하는 것이 가능해질 수 있다.

의료분야 마이데이터 사업의 방해 요인

현 시점에서 의료 분야 마이데이터 사업이 저해되는 원인을 다음과 같이 정리해볼 수 있다. 첫 번째는 사용자의 낮은 신뢰감 및 거부감으로 인한 데이터 활용의 제약이다. 보건의료데이터의 특성상, 개인이 느끼는 정보의 민감도는 매우 높을 수밖에 없다. 그러나 현재 보건의료데이터 거래 시스템은 안정성이 보장되지 않으며, 보건의료데이터의 2차 활용



용에 대한 법적 규제가 미비하다. 4차위의 조사 결과에서도 의료데이터의 안정성 및 투명성에 대한 낮은 사회적 신뢰도와 데이터 활용에 대한 높은 심리적 거부감이 의료 분야 데이터 활용의 걸림돌로 제기된 바 있다.⁴⁾

두 번째는 폐쇄적이고 독점적인 데이터 활용으로, 표준화된 보건의료데이터 구축이 쉽지 않다는 점이다. 보건의료데이터는 각 병원에서 사용하는 전자 의무기록(Electronic Medical Record, EMR) 시스템에 의해 수집되고 관리된다. 즉 1차, 2차 병원 및 상급종합병원에 이르기까지 저장하는 데이터의 종류와 형태(저장 방식)가 모두 다를 수 있다는 가능성을 내포한다. 이를 해결하기 위해 개인건강정보(Personal Health Record, PHR) 서비스의 필요성과 사업화 및 적용 방식이 논의되고 있으나, 폐쇄적으로 운영되는 EMR 데이터의 유출에 대한 우려와 PHR을 도입하는 것에 대한 회의가 뒤섞여 실질적으로 정착되기까지에는 시일이 걸릴 것이다.⁵⁾ 마지막은 의료분야의 특수성을 고려한 접근이 부족하다는 점이다. 보건의료데이터는 생명과 직결된 정

보이기 때문에 신중하게 접근해야 한다. 데이터를 잘못 해석하고 병원을 계속 방문하여 의료비 급증을 유발하거나, 지역 의료기관보다 상급종합병원으로 이용이 쏠리는 등의 문제가 발생할 수도 있다. 따라서 전문가들과 함께 가능성과 위험요소를 충분히 검토하여 사업을 전개해야 한다.

교통분야 마이데이터를 통해 이동수단과 방법을 설계한다

교통 분야의 마이데이터 사업은 대부분 지자체에서 개인의 MaaS(Mobility as a Service) 데이터를 활용하여 대중교통의 사각지대를 발견하고 이를 적절하게 정책에 반영하기 위한 것을 목적으로 진행된다. MaaS는 자차, 대중교통과 같은 보편적 교통수단뿐 아니라 공유차(Sharing car), 공유 자전거, 자율주행차 등 새롭게 등장한 모든 것을 교통수단으로 포함하고 이를 바탕으로 다양한 개인의 요구를 충족시켜주는 서비스로 정의할 수 있으며⁶⁾, MaaS 데이터는 이 과정에서 수집 가능한 데이터를 의미한다. 교통 분야의 마이데이터 실증서비스는 아래의 표를

[표2] 의료 마이데이터 실증서비스 요약 (출처=데이터산업진흥원)

서비스 명	서비스 설명	주관기관
스마트시티 패스 (CityPass)	• 실증서비스 참여 시 1만 마일리지 보상 제공 • 교통 이용데이터를 부천시의 교통정책 수립에 활용 • 대중교통의 사각지대 해소 및 최적의 교통수단 재배치 방안 분석 목표	부천시 (2020)
마이데이터 기반 장애인 이동지원 실증 서비스	• 개인 데이터(개인정보, 생활정보, 활동정보, 디지털 패턴 분석)를 도시데이터, 공공데이터와 연결 • 마이데이터 기반의 교통약자 이동지원 서비스	대전시, KISTI (2020)
마이티(My-T)	• 개인위치정보(GPS), 대중교통 운행정보, 교통카드 승하차 정보, 교통수단별 Wi-fi 정보를 수집 • 대중교통 혼잡도 기반 길찾기 • 코로나19 확진자 이동 경로 겹침 알람	서울시, 한국교통연구원 (2020)
안심대리	• 대리기사의 운행데이터, 근로데이터를 활용한 이력 및 실시간 위치 알림 • 개인 소득 통합관리 서비스 • 승객의 안전이용을 위한 안심 대리운전 서비스	인플랩 (2021)

통해 확인 가능하다.

모빌리티 서비스는 ‘라스트마일(Last Mile)’을 매우 중요하게 생각하며, 라스트마일 경험을 적절하게 설계하기 위해 많은 노력을 기울인다. 예를 들면 현대자동차에서는 이러한 라스트마일 모빌리티를 고객의 여정을 마무리하는 마지막 이동으로 ‘도보’, ‘마이크로모빌리티 차량’, ‘공유 자전거’, ‘전동 스쿠터’ 등을 활용하여⁷⁾, 고객의 라스트 마일을 만족시켜주기 위해 노력하고 있다.

이동 수단 별 노선과 운영 방침을 세워야 한다면, 먼저 사람들의 동선을 고려하여 그들의 '라스트 마일'을 최소화해주는 방향으로 설계해야 한다. 약속장소로 가기 위해 지하철 혹은 버스를 이용했는데, 내려서 한참 걸어간 적이 있지 않은가? 혹은 내가 자주 타던 대중교통의 배차간격이 갑자기 길어지거나, 해당 버스 노선이 없어지는 경우가 있지 않은가? 지자체에서는 효율적이면서도 적절한 대중교통의 운영을 위해 노선과 운행 시간, 운행 빈도 등을 조절한다. 이 과정에서 사용자의 MaaS 데이터가 주요한 지표로 활용될 수 있다. 이에 더하여, 주행에 따른 안전운전 점수를 획득하여 자동차 보험료를 할인받는 사례처럼 MaaS 데이터가 다른 분야에 적용되어 새로운 경험을 제공하는 경우도 찾아볼 수 있다.

혁신은 늘 반복되는 일상 속에서 갑자기 다가온다

금융 분야에 대한 마이데이터 사업이 본격적으로 시행된 지 약 4개월 정도의 시간이 흘렀다. 하지만 마



이데이터 사업의 격전지가 될 것으로 예상했던 금융 분야조차 정책 수혜자들에게 인지 가능한 수준의 서비스를 제공하지 못하고 있다는 평가가 우세해지면서, 사업의 안정화를 꽤나 먼 시점으로 내다보는 경우가 많아졌다. 뚜껑을 열어본 사람들의 평가는 생각보다 회의적인 듯하다.

그러나 동시에 사업을 전개하며 실패 사례들을 많이 목격하고 새로운 니즈와 인사이트를 마주함에 따라, 머지않아 마이데이터 사업의 주도권을 잡은 키플레이어(key player)가 나올 것이라는 예측이 지배적이다. 그리고 이 불명한 예측 이면에는 자신의 비즈니스가 바로 그 '키플레이어'가 되어야 한다는, 될 것이라는 열망이 서려있다. 모름지기 혁신은 꾸준한 변화 속에서 갑자기 일어나는 법이다. 앞으로 이어질 다양한 시도 속에서, 분야를 넘어선 데이터의 융합으로 유의미한 정보 제공에 성공할 키플레이어의 탄생을 지켜보자.

1) 윤수영. (2018). 4차 산업 혁명 시대의 소비자 데이터 주권에 대한 고찰: EU GDPR을 중심으로. 소비자학연구, 29(5), 93-115.
2) 보건복지부 보도자료(2021.12.), “마이 헬스웨이 활용기관 수요조사 결과 발표”
3) 보건복지부 보도자료(2021.02.), “마이 헬스웨이 구축 시작”
4) 보건복지부(2021.6.), “보건의료 데이터 인공지능 혁신전략”
5) 김홍진(2021.9.), “PHR 취지는 좋다, 그러나 ‘어떻게 득 될지’ 고민 먼저” URL: <http://www.hitnews.co.kr/news/articleView.html?idxno=35687>
6) 서울연구원 정책리포트(2019.9.), “서울형 통합교통서비스 도입방안”
7) 현대자동차 홈페이지, <https://www.hyundai.com/worldwide/ko/brand/last-mile-mobility>

03

금융 마이데이터,
3년의 준비를 통해
서비스 본격화

심현섭 | 한국신용정보원 빅데이터센터장

70여 기관이 참여하여 금융분야
마이데이터 추진

금융분야 마이데이터는 2018년 7월, 금융위원회 마이데이터산업 도입방안¹⁾ 발표 이후 법제화가 본격적으로 추진²⁾ 되었다. 입법과정에서 빅데이터 활성화를 위한 가명/익명정보 도입과 개인정보 거버넌스 확립 등을 위해 데이터3법 개정이 일괄적으로 추진됨에 따라 시민단체들의 반대가 매우 많았다. 이에 따라 조정과 설득에 상당한 기간이 소요되어 2020년 1월말이 되어서야 개정안이 국회를 통과할 수 있었다.

한편, 법제화와는 별도로 2019년 5월부터는 금융당국을 중심으로 지원기관들과 각 업권별 대표 금융회사 등 70여 기관이 참여한 워킹그룹³⁾이 구성되어 마이데이터 서비스에 필요한 다양한 사안에 대해 논의하기 시작하였다.



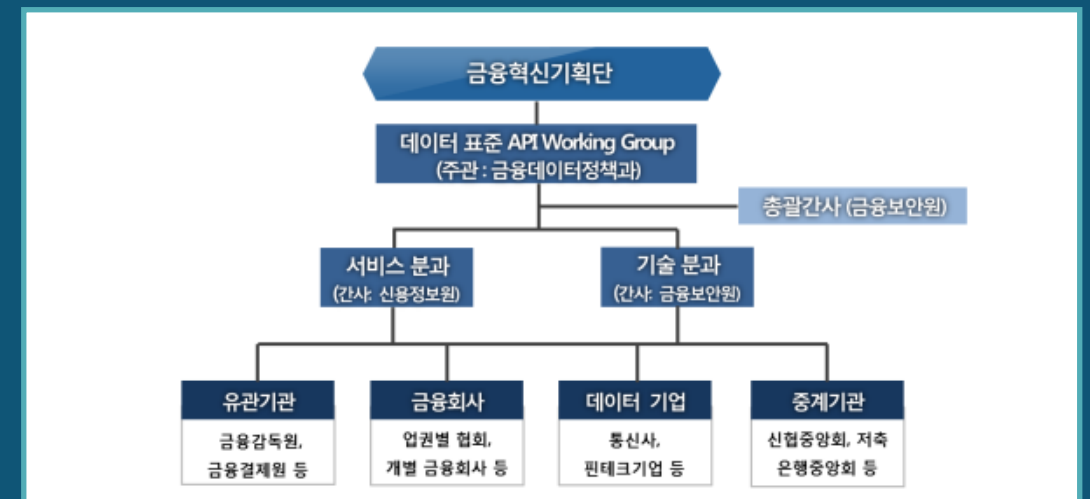
서비스 분과에서는 한국신용정보원을 간사로 하여 전송요구권 행사 대상정보와 항목을 어떤 기준으로 선정할 것인지를 논의한 후, 업권별 소분과를 구성하여 세부 정보항목을 선별하고 데이터 표준화와 구조화를 진행하였다. 또한 해외 마이데이터 사례와 과금 모델, 손해배상 등 관련 사항들에 대해 폭넓은 자료조사를 병행하였다.

기술분과에서는 금융보안원을 간사로 하여 서비스 분과에서 도출한 정보제공항목 표준화 결과물에 대해 데이터별 전산화된 처리규격(데이터명, 타입, 길이, 필수여부 등)을 마련하였고, API 세부규격, 접근

토큰 규격, 전송요구 최소단위, 시스템 부하를 고려한 API설계방안 등을 검토하였다.

이러한 워킹그룹의 활동은 2021년 1월까지 약 19개월간 지속되었으며, 준비과정에서 예기치 못한 어려움이 곳곳에서 발생하였다. 이러한 어려움들은 금융당국을 비롯한 유관기관들의 헌신적인 노력과 공동목표를 위한 이해관계자들의 양보로 하나씩 해결해 왔으며, 그 결과 금융마이데이터는 2021년 12월 시범서비스 운영후 2022년 1월 5일부터 본격적인 서비스가 개시되었다.

데이터 표준 API Working Group 구성도



1) 「금융분야 마이데이터산업 도입방안」 (2018. 7월, 금융위원회)

2) 2018. 11월, 신용정보법 일부 개정법률안 국회 제출(김병욱 의원 대표발의)

3) 실무준비를 위해 금융위 주관으로 데이터 표준API 워킹그룹(2019. 5월~8월(1차), 2019.10월~2021.1월(2차))을 구성하였으며, 서비스분과(간사:한국신용정보원)와 기술분과(간사:금융보안원)로 나누어 운영

금융분야 마이데이터에서 논의되는 주요 이슈

우리나라의 금융마이데이터 산업은 수천여 개의⁴⁾ 금융회사 등이 정보제공의무를 지고 있으며, 모든 정보제공자가 동시에 서비스에 참여하는 것으로, 규모와 속도 면에서 전 세계적으로도 유례를 찾아보기 어렵다. 이 과정에서 다양한 이슈들이 발생하였는데, 그 내용을 간략하게 살펴보기로 한다.

전송대상 데이터 확대

금융분야 마이데이터 산업은 신용정보의 이용 및 보호에 관한 법률(이하 “신용정보법”)에 근거를 두고 있는데, 「신용정보법」의 전송요구권은 GDPR⁵⁾의 데이터이동권을 참고로 하였다. GDPR은 이동권 행사 대상 데이터에 대해서는 대원칙만 정할 뿐 세부적인 규정이 없으나, 우리나라에서는 제도의 조기 안착과 산업활성화를 위해 전송대상 데이터의 개요를 법률에서 규정하고 세부내역은 워킹그룹을 통해 각 이해관계자의 의견을 광범위하게 수렴⁶⁾하여 시행령에 반영하였다. 그러나 이러한 의견수렴 과정을 거쳤음에도 금융마이데이터 출범이 가시화되면서 기존 스크래핑 기반 서비스와 API 제공대상 정보항목에 차이가 있어 제공데이터 확대에 대한 요구가 지속적으로 발생하였다. 이에 대해 정보제공자는 프라이버시 침해 우려

또는 제3자 정보 제공근거 불명확 등 으로 어려움을 호소하는 등 제공대상 데이터를 둘러싸고 상당한 진통을 겪었다. 게다가 법률에서 정한 시행일정이 촉박함에 따라 제공데이터 확대 시 정보제공자의 전산개발이 어려운 점도 많았다.

이에 따라 금융당국을 중심으로 수차례 업권 간 협의를 거친 결과, 당초 제공대상이 아니었던 은행계좌적요와 보험보장내역 및 카드가맹점정보 등도 소비자보호장치⁷⁾를 전제로 제공대상에 포함하였다. 향후에도 법적 쟁점⁸⁾ 해소와 해당 업권의 제공 참여 유도 등을 통해 전송데이터는 지속적으로 확대될 것이다.

식별 key(CI)

마이데이터는 다수의 정보제공자로부터 특정 정보주체의 데이터를 일괄조회하여 서비스를 제공하는 것으로, 동일한 정보주체임을 식별할 수 있는 key가 반드시 필요하다. 현재 정보주체를 식별하는 key는 정보보유자별로 다양⁹⁾하게 이용되고 있는데, 그 중 주민등록번호는 법령에 특별한 규정이 없는 한 사용이 금지¹⁰⁾되어 범용 식별 key로 활용하기 어려운 단점이 있다. 이에 따라 금융마이데이터에서는 범용성을 고려하여 CI¹¹⁾를 식별 key로 선정하였다.

CI는 주민등록번호 대신 본인을 확인할 수 있는 대

체수단¹²⁾으로 「정보통신망법」에 따라 지정된 본인확인기관이 발급할 수 있으나, 발급 및 이용 등에 대해 법령상 상세한 규정이 없어 금융마이데이터에 CI를 활용하는 데에 실무상 어려움이 많이 발생하였다. CI를 사용하지 않거나, 사용하더라도 DB에 저장하지 않는 정보제공자가 상당수 있었으며, 이에 따라 금융당국에서 관계부처 간 수차례 협의를 통해 정보제공자가 보유한 기존 고객의 주민등록번호를 CI로 일괄변환할 수 있도록 조치¹³⁾하였다. 이 외에도 활용목적 상이로 인한 재변환과 공공마이데이터 연계를 위한 CI 적용 등 이슈도 발생하여 이의 해결에 많은 노력과 시간이 필요했다.

통합인증

정보제공자는 안전한 개인신용정보 전송을 위하여 정보주체가 개인신용정보 전송을 요구할 경우 해당 주체가 맞는지 확인하여야 한다.¹⁴⁾ 확인방법은 본인인증을 이용하고 있는데, 이에는 개별인증방식¹⁵⁾과 통합인증방식¹⁶⁾이 있다. 금융마이데이터에서는 정보제공자(중계기관 이용기관 제외)가 고객이 통합인증 및 개별인증 중 하나를 선택하여 인증을 수행

할 수 있도록 하고 있다. 또한 금융마이데이터에서는 인증결과에 CI를 제공할 수 있는 다중요소 공개키인증서¹⁷⁾를 통합인증수단으로 정하였는데, 「정보통신망법」상 지정받은 인증서 본인확인기관¹⁸⁾이 통합인증서를 발급·관리하는 인증기관으로 참여할 수 있다. 그런데, 당초 마이데이터 서비스 개시를 목표로 했던 ‘21.8월 초에는 이용이 간편한 전자서명인증사업자가 「전자서명법」에 따른 인정절차를 통과하지 못하여 사용자의 불편이 매우 우려되었다. 이에 따라 관계부처 간 긴밀한 협의로 관련 절차를 가속화한 결과, 8월 말부터는 전자서명인증사업자가 속속 등장하여 마이데이터서비스 이용자들이 다양한 통합인증수단을 편리하게 이용할 수 있게 되었다.

’22년 56개 사업자, 45개 서비스 출시

금융마이데이터는 올해 1월부터 서비스가 본격화된 이후 안정화 과정에 있다. 수많은 정보제공자와 다양한 정보가 전송되기에 초기에는 시스템 안정성

[표1] 제공데이터 확대요구 주요 대상

업권	제공데이터	주관기관
① 전자금융	주문내역정보*	프라이버시 침해 가능성, 신용정보의 범위 확대 우려 등
② 은행 등	적요·거래내역메모	제삼자정보 제공의 적법성 등
③ 보험	보장내역, 계약자-피보험자 상이건 등	제삼자정보 제공의 적법성 등

※ 주문내역정보는 13개 카테고리만 제공
(가전/전자, 도서/문구, 패션/의류, 스포츠, 화장품, 아동/유아, 식품, 생활/가구, 여행/교통, 문화/레저, 음식, e쿠폰/기타서비스 등)

4) 금융업권, 전자금융업자 및 공공기관 등 5천여개 법인(농·수협조합, 새마을금고, 신탁 등 개별 법인 기준시)
5) General Data Protection Regulation(EU의 일반개인정보보호규칙. 2018.5.25.부터 적용)
6) 핀테크사의 스크래핑 기반 통합조회 서비스 데이터 항목과 각 금융회사 홈페이지에서 고객이 조회할 수 있는 정보항목들을 수집하여 이를 대상으로 신용정보 해당 여부, 민감/가공정보여부 등을 검토하여 제공대상 정보항목을 선별하였음
7) 해당 정보제공에 대한 별도 위험고지 및 별도 동의, 제삼자 제공 및 마케팅 목적 활용금지 등
8) 금융업은 다양한 법령에 따라 사업을 운영하고 있으며, 이 과정에서 정보제공근거가 없거나 불명확한 경우가 많이 발견되고 있음
9) 주민등록번호, CI, DID, 고객번호, 주소, 전화번호, 카드번호 등
10) 개인정보 보호법 제24조의2
11) Connecting Information, 연계정보
12) 정보통신망법 제23조의2 개정(‘12.2월 시행)
13) ’21.5월말 혁신금융서비스 지정으로 일괄변환 근거 마련(마이데이터 통합인증을 위해 주민등록번호를 CI로 일괄 변환 가능)
14) 신용정보법 제33조의2 제8항, 금융분야 마이데이터 기술가이드라인 제4장
15) 고객이 개별 정보제공자가 제공 또는 인정하는 인증수단을 이용하여 각 정보제공자별로 개인신용정보 전송요구 및 인증을 수행하는 방식
16) 고객이 통합 인증기관이 발급한 인증수단을 이용하여 1회 인증만으로 다수의 정보제공자에 개인신용정보 전송요구 및 인증을 수행하는 방식
17) 안전하게 생성·보호된 개인키 및 공개키 인증서로서, 인증요구를 위한 전자서명을 생성하기 위해 개인키 인증정보(비밀번호, 생체정보 등)를 요구하는 방식을 말함(금융분야 마이데이터 기술가이드라인 제4장 참조)
18) 2021년 12월초 현재 인증사업자는 ① (공동인증서 발급기관) 금융결제원(범용/은행용), 코스콤(범용/증권용), 한국정보인증(범용/은행용), 한국전자인증(범용), ② (전자서명법상 인정받은 전자서명인증사업자) NHN페이코, 신한은행, 국민은행, 네이버, 금융결제원, 토스, 뱅크샐러드 등

이 우려되기도 하였으나, 약 3개월간 운영상황을 살펴본 결과 별다른 문제없이 서비스가 안정적으로 제공되고 있다.

2022년 4월 현재 금융위원회로부터 허가를 받은 마이데이터 사업자는 56개사이며, 앞으로도 계속 증가할 전망이다. 사업자별로 허가받은 시기가 달라 서비스 준비에 차이가 있어 '22.4.13. 현재 45개 마이데이터 서비스가 출시되었다.

출시된 서비스를 살펴보면 대다수 사업자가 종합자산관리와 맞춤형 상품추천 등의 서비스를 제공하고 있으며, 일부 사업자는 보험상품, 금융투자, 대출비교 등에 특화된 서비스를 제공하고 있다. 고객이 거래하는 금융회사의 카드·출금 거래정보를 통합조회하여 기간별·결제수단별·카테고리별로 지출내역을 분석하고, 목표설정·달성도 등 다양한 편의기능을 제공하고 있다. 또한 고객이 본인의 자산부채 현황을 쉽게 파악할 수 있도록 하는 재무상태 진단서비스도 있으며, 또한 개인별 소비패턴을 바탕으로 카드 상품을 추천하거나, 대출상품별 최저금리, 최대 한도 대출조건 등을 비교하여 본인에게 적합한 대출을 추천하고 비대면대출계약 서비스를 제공하는 사업자도 있다.

보험분야에서는 보험가입 내역 통합조회서비스, 개인 데이터 기반의 상품 추천 및 모바일 기반의 보험금 간편청구 서비스도 제공하고 있다. 은행, 카드, 보험, 증권, 핀테크 등 다양한 업권의 마이데이터 사

업자가 각자의 강점을 가지고 있으며, 소비자들은 이러한 업권별 특성을 고려하여 본인에게 가장 적합한 마이데이터 사업자를 선택하면 유용한 서비스를 받을 수 있다.

이와 함께 금융정보를 중심으로 타 산업정보를 융합한 특화된 서비스를 준비하는 사업자들도 있는데, 건강·의료데이터를 기반으로 건강관리서비스와 보험상품 추천을 하거나, 부동산도 자산에 통합관리하는 등 다양한 서비스가 있다.

한편, 이러한 서비스는 API기반의 금융마이데이터가 출범하기 이전 스크래핑 방식으로 제공하던 마이데이터 서비스와 별다른 차이가 없다는 시각도 있다. 그러나 스크래핑 기반의 마이데이터가 서비스 안정성과 정보보호에 대한 우려가 상당했던 점, API기반의 마이데이터가 출범 초기인 점 등을 고려하면 지속적인 정보확대와 사업자간 경쟁 본격화를 바탕으로 조만간 서비스 내용이 급속히 확장될 것으로 기대한다.

이용자는 1,840만명으로 서비스 활성화 기대

금융마이데이터 출범 이후 서비스 이용자는 '22.4.13. 현재 1,840만명(마이데이터 사업자 약관 동의 기준, 사업자별 중복집계)이며, 출범 약 2달만에 125억건(누적)의 API가 전송되는 등 매우 활발한 이용실적을 보이고 있다.¹⁹⁾ 출범 초기부터 시장 선점을 위해 사업자별로 가입자 유치경쟁이 매우 치열하였으며, 일정 수준의 경품을 제공하는 등 각종 이벤트도 활발하게 진행하였다. 소비자들은 마



이데이터를 통하여 본인에게 최적화된 맞춤형 금융 상품·서비스 이용에 대한 기대가 높다. 다양하고 초개인화된 서비스로 금융 편의성이 한층 높아질 것이며, 현재 진행중인 개인정보보호법 개정으로 마이데이터가 모든 산업으로 확산된다면 산업간 다양한 데이터가 연결되면서 서비스 내용이 무한대로 확장되어 국민들의 생활은 더욱 더 편리해질 것으로 보인다.

존을 위해 치열한 경쟁을 할 것이며, 또한 금융회사별로 운영되는 폐쇄적 영업방식을 탈피하여 핀테크 등 새로운 플레이어들과의 협업을 통한 새로운 상품, 서비스 및 채널을 만들어 내는 개방형 혁신이 촉진될 것이다.

금융분야 마이데이터 발전을 위한 추진과제

산업적으로는 마이데이터 사업자가 기존 대형 금융회사들이 보유하고 있던 고객 데이터에 접근하기 용이해짐에 따라 기존 금융회사들의 시장지배력이 약화되고 경쟁이 치열해지면서 혁신적인 상품 개발이 기대되고 있다. 마이데이터 서비스가 고객접점으로 정착된다면 금융상품의 제조·판매채널 분리가 가속화될 것으로 전망된다. 금융회사나 핀테크사들은 생

첫째. 서비스 안정화 및 개인정보 보호 · 보안 강화
마이데이터 산업은 수많은 정보제공자와 마이데이터 사업자간 다양한 데이터가 전송되므로 오류 발생 가능성이 곳곳에 산재해 있다. 이에 따라 전송되는 데이터의 정확성 · 최신성 유지가 매우 중요하며, 전송되는 데이터가 부정확할 경우 소비자의 불신으로 산업 자체의 신뢰도가 훼손될 우려가 있다. 서비

[표2] 마이데이터 사업자 허가현황('22.4월말 기준)

업권	은행	여신전문	금융투자	보험, 상호금융, 저축은행	핀테크, CB, IT	전체
사업자 수	10	9	7	4	26	56



스를 개발할 때에는 정확성을 철저하게 검증하여 오류를 방지하고, 만약 오류가 있는 경우 참여자 간 신속한 사후처리를 위해 긴밀한 관리체계를 마련할 필요가 있다.

이와 함께 금융소비자가 마이데이터 서비스를 안심하고 이용할 수 있도록 개인정보 보호·보안체계도 지속적으로 강화할 필요가 있다. 제도적으로는 기능적합성 심사²⁰⁾와 보안취약점 점검²¹⁾ 의무화 등 엄격한 정보보호·보안체계를 마련하여 운영하고 있으나, 기술이 나날이 발전함에 따라 그것만으로는 충분하다고 보기는 어렵다. 생태계 참여자 모두가 각자의 환경에 맞는 보호정책을 마련하고 실행함으로써 소중한 개인정보가 유출되거나 오남용되지 않도록 각별히 유의하여야 할 것이다.

둘째. 전송대상정보 확충 및 서비스 고도화

마이데이터는 표준화된 양식인 API방식으로 정보를 제공하기 때문에 스크래핑방식보다 제공정보 항목을 확대하는 데에 제약이 발생할 수 밖에 없다. API방식이 의무화되면서 스크래핑 기반의 서비스를 제공해 오던 기존의 마이데이터 서비스가 일부 중단되는 등의 불편도 발생하고 있으나, 그렇다고 해서 정보보호에 단점이 많은 스크래핑 방식을 유지하기도 어려운 실정이다. 이에 따라 사업자가 필

요로 하는 정보항목에 대해서는 지속적으로 API를 확대하여야 하나, 정보제공자의 부담도 상당하므로 정보제공을 유도할 다양한 인센티브를 고민할 필요가 있다. 다른 한편으로 정보제공자는 개인정보가 정보주체의 활동에 따라 생성된 것이므로 자신이 보유한 개인정보가 오히려 자신만의 독점적인 자산이 아니라 정보주체에게도 일정한 권리가 있음을 자각할 필요가 있다.

현재 금융마이데이터는 초기단계여서 전송대상정보가 제한적이고 비교·추천 등의 서비스도 규제가 많아 아직까지는 각 사업자가 제공하는 서비스에 특별한 것을 찾아보기 어렵다. 하지만 많은 사업자가 마이데이터 서비스에 참여하고 있는 만큼 정보확대와 규제완화를 바탕으로 창의적인 서비스에 대한 기대가 매우 높다. 아직까지는 제도적으로 해결해야 할 과제도 많고, 이해관계자간 합의를 바탕으로 전송되는 정보의 양과 품질도 지속적으로 확대해 나가야 하는 등 선결과제가 많아 당장에는 기대를 충족시키기 쉽지 않다. 하지만 사업자들은 소비자들의 눈높이에 맞추어 다양한 서비스를 개발할 필요가 있다. 이로써 새로운 산업이 조기에 활성화 되는 것 뿐만 아니라, 금융업의 프로세스를 혁신하여 금융산업 전체를 변화시킬 수 있는 촉매가 되기를 기대한다.

셋째. 합리적 과금체계 마련

마이데이터 산업을 유지하기 위해서는 데이터 전송, 공동인프라 구축, 인증 및 식별 key 활용 등에 다양한 비용이 소요된다. 마이데이터산업의 원활한 운영을 위해서는 이러한 다양한 비용을 누가, 어느 정도로 부담하는 것이 적정한지에 대해 생태계 참여자 간 합의가 필요하다. 정보제공자의 소요비용 보전과 마이데이터사업자의 서비스 운영 부담도 고려하여야겠지만, 산업의 활성화와 정보주체의 권리 보장이란 관점도 고려하여야 할 것이다.

「신용정보법」에서는 ‘정기적인 데이터 전송’에 대해 마이데이터사업자에게 최소한의 비용을 부담하게 할 수 있도록 규정²²⁾하고 있으며, 비용 산정기준 등은 전송요구권 행사 대상 개인신용정보의 특성·처리비용 및 요청한 개인신용정보의 범위·양 등을 고려하여 금융위원회가 정하여 고시²³⁾하도록 하고 있다. 다만, 산업활성화와 적절한 과금체계 마련을 위한 데이터 축적을 위해 산업출범 후 1년간 과금을 유예하였다.

워킹그룹 운영 시 국내외 유사 서비스에 대한 다양한 과금 사례²⁴⁾를 조사하였는데, 현실적으로는 호출 건수 모델²⁵⁾과 정액제 모델²⁶⁾이 가장 이해하기 쉬운 것으로 파악되었다. 금융마이데이터에서는 추후 산업 운영 시 산출되는 통계를 바탕으로 전문기관의 용역 등을 통해 과금모델을 산출할 계획이며, 이해관계자들의 의견을 광범위하게 수렴하여 적절한 과금체계를 마련할 예정이다.

넷째. 타 산업 연계기반 확충

마이데이터는 국민들의 일상생활에 필요한 다양한 서비스를 목표로 하며, 특정 산업의 데이터만 활용

해서는 이러한 목적을 달성하기 매우 어렵다. 공공, 보건, 금융, 유통, 통신 등 여러 분야의 데이터가 연결될 경우 국민들의 다양한 수요를 충족시킬 수 있는 서비스가 한층 수월해질 것이다. 이를 위해 개인 정보보호법에 모든 산업분야에 마이데이터의 근거를 마련²⁷⁾하기 위한 움직임이 있으며, 개인정보보호 위원회에서는 데이터 표준화²⁸⁾도 진행하고 있다.

한편, 마이데이터가 모든 산업에 도입될 경우 정보 제공의무를 지게 될 정보보유자가 얼마나 많을지, 마이데이터 사업자가 얼마나 될지, 지금 단계에서는 상상하기가 어렵다. 이 경우 수많은 사업자간 데이터가 원활하게 전송되고, 정보주체의 권리가 제대로 보장받기 위해서는 표준화를 주관하고 참여자를 인증하는 등의 역할을 수행하는 산업별 지원센터²⁹⁾가 필수적이다. 또한 중소기업 정보보유자의 전송부담을 완화하고, 마이데이터 사업자와 연결을 원활하게 지원할 수 있는 중계기관³⁰⁾도 필요하다.

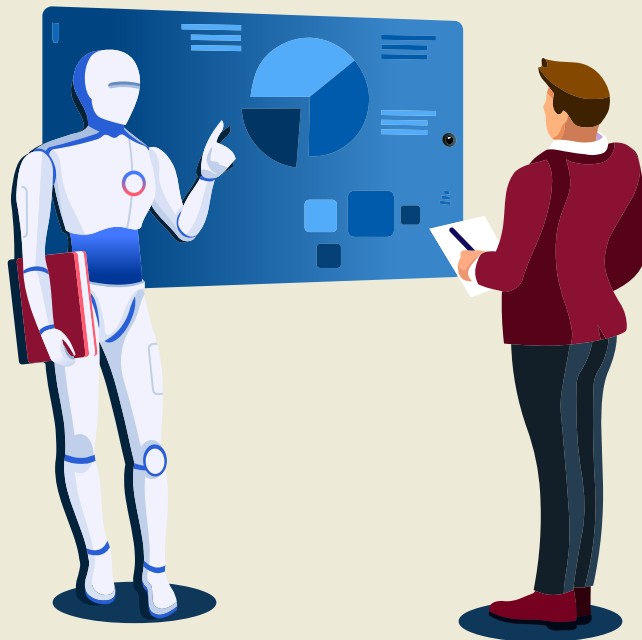
이러한 인프라를 통해 산업간 연계가 원활해지며, 이를 통해 수많은 사업자가 다양한 서비스를 개발할 수 있는 기반이 될 것이다.

19) 금융위 보도자료('22.4.13, 「22년 본인신용정보관리업(마이데이터) 허가심사방향」)
20) 마이데이터 서비스 프로그램의 신용정보법령상 행위규칙 준수 여부, 표준API 규격 적합성 등을 서비스 출시 및 주요 기능 변경 전 사전심사
21) 마이데이터 서비스 관련 시스템·앱 일체에 대해 금융보안원 점검기준에 따라 전금법상 평가전문기관이 연 1회 이상 보안취약점 점검 수행
22) 제22조의9 제6항
23) 신용정보법 시행령 제18조의6 제11항
24) (해외) 월정액, 종량제, 일부무료+추가분 과금 등, (국내) 금융결제원의 오픈뱅킹 수수료
25) API 이용 건수별로 요금 부과(마이데이터 사업자의 불필요한 API 이용을 최소화할 수 있으나, 부담 과도 우려)
26) API 누적 이용 건수를 기준으로 구간별 사용료 책정(API 누적 이용 건수별로 과금액을 유기적으로 조정가능하며, 마이데이터 사업자가 성장하는 정도에 따라 비용 부담 가능)
27) 개인정보보호법 개정안 국회 계류('21.9.28 정부안, '22.1.5 민병덕 의원안 등)
28) 4차산업혁명위원회 데이터특별위원회, 마이데이터 데이터 표준화방안('21.9.28)
29) 금융분야에서는 정보항목 표준화, 과금체계 수립 및 운영, 법·제도 개선제안, 분쟁조정, 기술규격 마련, 참여기관 관리·등록, 회의체 및 Help Desk 운영 등을 위한 지원센터 가동중
30) 금융분야에서는 신용정보원, 금융결제원, 코스콤 등이 업권별로 분담하여 중소기업 정보제공자를 위한 AP전송시스템을 구축함

통계데이터 인공지능활용대회

국 가 데 이 터 ! 인공지능 시대를 연다

오승철 | 통계청 통계데이터기획과 사무관



지금 세계는 인공지능(AI), 빅데이터, 클라우드, 초연결 등으로 일컬어지는 4차 산업혁명 시대를 맞이하고 있다. 18세기 영국을 중심으로 증기기관 기반의 기관차, 증기자동차 등 기계적 동력을 이용하여 노동력을 대체하면서 1차 산업혁명이 일어났고, 1914년 헨리 포드가 자동차 공장에 컨베이어 시스템을 도입하여 전기에너지 기반의 대량생산을 통한 원가절감을 이루면서 2차 산업혁명, 1970년대 엘빈 토플러가 ‘제3의 물결’로 명명한 컴퓨

터나 인터넷 기반의 전자기술 및 IT를 이용한 공장 자동화를 통한 3차 산업혁명, 최근 초연결(hyper connectivity)과 초지능(superintelligence)을 특징으로 하는 4차 산업혁명까지 인류는 끝없이 진화하고 있고 그 주기 또한 점점 빨라지고 있다. 지금은 인공지능이 단순 노동을 대신하는 것뿐만 아니라 인간의 사고까지 대체하고 더 나아가 감정교류까지 하면서 각 분야별로 급속도로 발전하면서 우리의 삶과 산업에 있어 큰 변화를 일으키고 있다.

통계청도 주목한 국가데이터 활용에서 인공지능의 가치

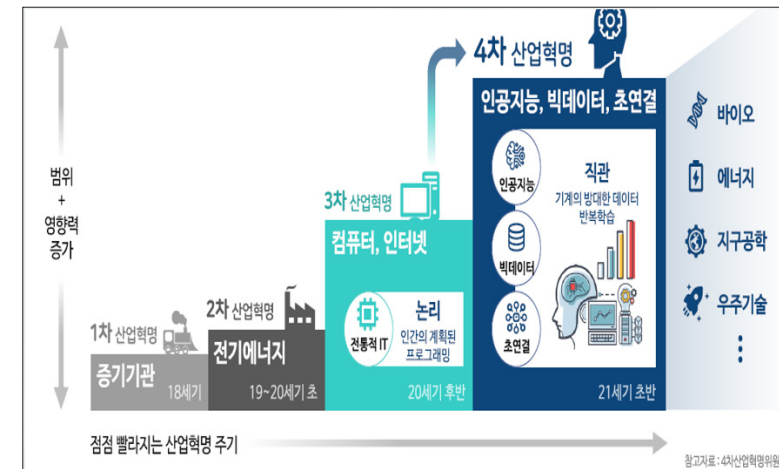
인공지능은 그 기술의 잠재력만으로도 우리 산업에서 근본적인 혁신을 가져오며, 새로운 일자리 창출뿐만 아니라 기존 일자리를 대체하면서 발생하는 실업문제, 재교육, 이직 등 사회적 변화도 일으키고 있다. 이렇게 인공지능에 관한 관심이 본격화되면서 우리 삶의 변화 속도와 폭은 더욱 가속화될 것이고 광범위해질 것으로 예상된다.

세계 주요국들은 인공지능을 활용한 산업의 주도권을 선점하기 위하여 다양한 노력을 기울이고 있다. 데이터 전문기관인 통계청도 세계적인 흐름에 발맞춰 인공지능과 관련된 연구를 진행 중이다. 특히, “국가 디지털 공공서비스 혁신사업”의 일환으로 한국어 기반의 인공지능 분류학습에 필요한 양질의 데이터를 통계데이터센터를 통해 시범적으로 개방하고, 국가통계에 자연어 기반 인공지능 학습 및 활용 가능성을 확인하여 데이터의 잠재적 가치를 제고하기 위해 2022년 처음으로 통계데이터 인공지능 활용대회를 개최하였다.

제1회 대회 결과, 395팀 참여하여 다양한 인공지능 기술력과 아이디어 선보여

올해 처음으로 개최한 인공지능 활용대회는 통계청이 보유한 사업체 관련 자연어(투입물, 산출물, 생산활동 등)와 산업분류의 관계에 인공지능(AI) 기술을 적용하여 사업체 산업분류의 자동화를 도모하는 것으로 통계데이터센터 홈페이지를 통해 참가 신청과 결과물을 제출받았고, 실시간으로 점수와 순위를 공개하는 등 코로나19 확산 방지와 참가자 접근성 및 편의성을 높이기 위해 비대면 방식으로 진행하였다. 대학(원)생, 연구원, 일반 국민 등 395개 팀이 참여하여 총 110만 건의 자료 중 학습용 데이터 100만 건으로 자연어 기술 처리 및 인공지능 기반 산업분류 예측 모델을 구축하고, 이를 이용해 사업체 주요 특성에 관한 별도의 자연어 자료 10만 건을 처리하여 해당 사업체의 산업분류를 예측하였다. 1차 심사에서 정확도(Accuracy)와 F1 점수를 기준으로 정량평가를 통해 2배수를 선발하였고, 2차 심사와 최종 심사에서는 모델의 실용성, 재현성, 독창성, 정확성, 작성력(가독성) 등을 기준으로 전문가들이 평가하여 최종 수상자가 결정되었다. 심사결과, 영예의 대

인공지능과 4차 산업혁명



통계데이터 인공지능 활용대회 개최



상은 “자연어 스테디팀”(김*구, 이*윤, 박*수)이 수상하였다. 이 팀은 HIBERT 모델을 독창적 알고리즘으로 구축하였고, 앙상블 방식 채택, 시각화, 활용 방안 등을 상세히 기술하여 정확성, 실용성, 가독성이 돋보였다. 이외에도 최우수 2팀, 우수 3팀, 특별상 5팀이 최종 선정되었고, 수상자에게는 통계청장상(6팀), 한국통계진흥원장상(5팀)과 함께 총 8백만원 상당의 포상금을 수여하였다.

2022년 통계데이터 인공지능 활용대회 수상 범위

	대상	최우수	우수	특별상
인원	1명/팀	2명/팀	3명/팀	5명/팀
상금	2,500천원	각 1,500천원	각 500천원	각 200천원
표창	통계청장(6팀)/통계진흥원장상(5팀)			

통계데이터 인공지능 활용대회 수상자

참가자 (소속/팀명)	수상
김*구, 이*윤, 박*수 (자연어 스테디)	대상
김*현 (개인)	최우수
이*성, 박*하 (메타몽)	최우수
신*호 (개인)	우수
장*환, 손*은 (경북대학교 데이터사이언스)	우수
김*빈, 강*재, 정*용 (건국대학교/아주대학교)	우수
신*훈 (개인)	특별
서*건, 김*연, 김*영 (광주과학기술원 SCI LAB)	특별
한*찬, 이*, 이*섭 (순천향대학교)	특별
박*연, 장*규, 안*진 (PoolC Bart)	특별
윤*원, 장*건, 신*지 (부스트업)	특별

류근관 통계청장은 인공지능 활용대회 보도자료를 통해 “통계데이터 인공지능 활용대회에 적극적으로 참가해 주신 모든 분께 감사드리고, 수상자에게 축하의 말씀을 전한다.”라면서, “통계청은 4차 산업혁명 확산 및 디지털 플랫폼 구축 등으로 데이터의 중요성이 한층 높아지고 있는 만큼, 통계자료뿐만

아니라 여러 다양한 형태의 데이터에 대해서도 활용 가치를 제고하기 위해 적극적으로 개방해 나갈 계획이다.”라고 밝혔다. “나아가 통계청은 이번 대회를 계기로 데이터 시대를 선도하는 최고의 데이터 과학자를 양성하기 위한 범정부적인 노력에 부응하여 이를 주도해 나가겠다.”라고 포부를 밝혔다.

통계데이터 인공지능 활용대회 시상식



인공지능을 통해 더욱 더 발전하게 하게될 「통계데이터센터」는 어떤 곳인가

인공지능(AI), 사물인터넷, 빅데이터 등 4차 산업혁명 시대 서비스가 폭발적으로 성장하면서 데이터 센터의 중요성이 나날이 높아지고 있다. 통계데이터 센터에서는 이러한 데이터 환경변화에 신속히 대응할 필요가 있어 증거기반 정책지원과 데이터 생태계 조성을 위해 이용자 중심의 서비스를 제공하기 위한 다양한 노력을 기울이고 있고, 그 결과 이용실적이 양적으로 빠르게 성장하면서 이용자 만족도 점수가 매년 지속적으로 상승하고 있다. 대구센터(’21년), 서울상암센터(’21년)를 신규 개소하는 등 다양한 노력을 기울이고 있다. 이와 더불어

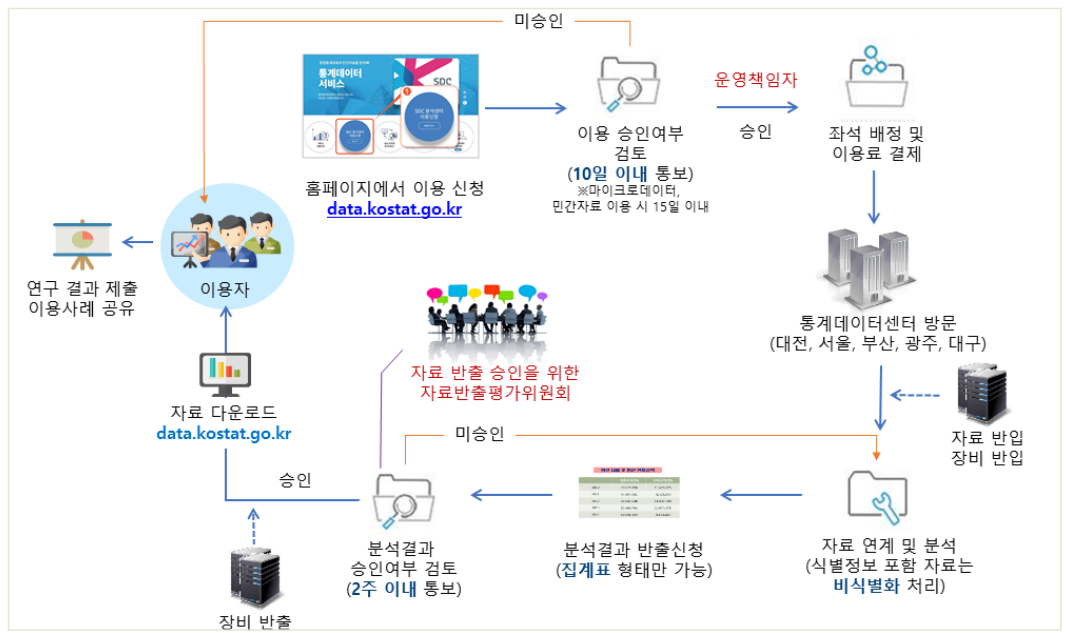
통계데이터센터 연도별 이용현황

(단위: 건, %)

구분	’18년	’19년	’20년(A)	’21년(B)	증감률(B-A)/A
방문 이용 건수(건/일)	46	1,367	1,611	2,315	43.7
방문 이용 인원(명)	175	522	616	801	30.0
주문형 분석 건수	-	15	17	32	88.2
이용자 만족도 점수	3.65	3.86	4.09	4.09	-

2022년에는 한국의 경제 사회 발전을 위한 계획을 수립하고 정책을 제시하기 위해 다양한 연구를 하고 다수의 연구진을 보유하고 있는 한국개발연구원(KDI)과 연계하여 통계데이터센터(SDC)를 추가 설치하였고, 서울강남센터와 서울국회센터도 상반기 중에 오픈할 예정이다.

특히, 한국교육개발원(KDI)은 세종시에 위치하여 KDI 연구진 외에도 세종시에서 근무하고 있는 다수의 정부부처와 연구기관들이 통계데이터센터를 통해 심도있는 연구를 수행할 수 있는 환경이 조성되었다.



통계데이터센터에서는 각종 행정자료를 연계하여 작성된 행정통계DB와 카드 및 통신사로부터 구입한 민간데이터, 통계조사 원자료인 마이크로데이터 등을 서비스하고 있다. 2018년 10종에 불과하던 행정통계DB는 2021년 16종으로 증가하였고, 민간데이터의 경우 2018년 22종이었던 것이 2021년 35종으로 데이터 콘텐츠가 확대되는 등 데이터 인프라 및 서비스가 확대되고 있다.

통계데이터센터에서 제공하는 데이터는 민감한 개인정보가 다수 포함되어 있어 자료의 활용성이 매우 높은 만큼 개인정보를 보호하기 위해 자료보안을 엄격히 운영하고 있다. 통계데이터센터에서는 자료보안을 위해 각종 운영규정과 가이드라인 등을 제정하여 법·제도적으로 정비하였고, 분석공간의 별도 분리, 데이터 반·출입 통제, 가상화 서버를 통한 Zero-Client 단말기 사용 등의 물리적인 통제장치로 자료보안을 철저히 하여 인가된 사람만이 이용할 수 있도록 운영하고 있다.

<통계데이터센터 제공자료 목록(행정통계자료)>

부문	자료명	시계열
경제 · 사회	기업등록부(대표자 기준)	2010~2020
	기업등록부(사업자등록 기준)	2010~2020
	기업등록부_분기(대표자 기준)	2018.4분기~2021.4분기
	기업등록부_분기(사업자등록 기준)	2018.4분기~2021.4분기
	사업장 기초 DB(4대보험)	2009~2019
	사업장 기초 DB(법인)	2008~2019
	종사자-기업체 연계 DB	2015~2019
	육아휴직 사용현황 DB	2015~2019
인구·가구	등록센서스 인구 DB	2015~2020
	등록센서스 가구 DB	2015~2020
	등록센서스 주택 DB	2015~2020
	인구동태 코호트 DB	2019
농림어업	아동가구 통계 DB	2015~2020
	농업 DB	2015~2020
	임업 DB	2015~2020
	어업 DB	2015~2020

공모전을 통해 더욱 더 쉽게 「통계데이터센터」를 활용해보자

통계데이터센터는 인공지능 활용대회뿐만 아니라 데이터 활성화를 위해 국민 누구나 참여할 수 있는 다양한 대회와 공모전을 매년 개최하고 있다. 대표적으로 ① 통계데이터센터 분석·활용대회 ② 통계데이터센터 이용·활용수기 공모전이 있으며, 분석·활용대회는 매년 5~9월, 이용·활용수기는 매년 상·하반기에 신청·심사·시상을 하고 있다. 특히, 통계데이터센터 이용·활용수기 공모전은 데이터에 대한 전문 지식이 없더라도 통계데이터센터를 이용한 경험을 바탕으로 작성 및 제출을 할 수 있기에 부담없이 이용자가 지원할 수 있다. 이렇듯 통계청에서는 각종 대회나 공모전을 통해서 데이터에 관한 관심을 증가시키고 국민들이 쉽게 접근할 수 있는 여러 방안들을 마련하고 있다.



통계데이터센터 제공자료 목록(통계기초자료)

부문	자료명
사업체	광업제조업조사, 경제총조사, 전국사업체조사 등
인 구	인구주택총조사, 사망원인통계, 인구동향조사 등
가 구	가계금융복지조사, 경제활동인구조사 등
농림어업	농림어업총조사, 농가경제조사 등

<통계데이터센터 제공자료 목록(민간자료)>

부문	자료명	시계열	주기
인구	성·연령별, 시간단위, 요일별 유동인구	2015.1~2021.12	월
	성·연령별 유입인구(부산)	2017.7~2018.6	월
	성·연령별 주거/직장 인구(서울, 부산)	2016~2020	년
	가구소득(서울, 부산)	2016~2017	년
	(코로나19) 방문인구정보 (유입지)	2019.1~2020.12	월
경제 · 사회	성 · 연령별, 업종별 매출(현대카드)	2015.1~2018.6	월
	성·연령별 국내/해외 매출(BC카드)	2018.10~2021.11	월
	KCB신용융합통계정보(대출, 소득, 소비)	2018.10~2020.12	월
	KCB평균소득소비(전입, 전출, 통근인구)	2018.10~2020.12	분기
	유통고객 구매상품 정보(롯데멤버스)	2019.12~2021.11	월
	1000대 상권 정보	2018	년
	영업 중 업소	2017.7~2018.6	월
건물, 지가	표준공시지가	2016~2017	년
	주택(단독/다세대/기타)	2016~2017	년
	공동주택정보	2016~2017	년
	건축물정보	2017.7~2018.6	월
	인구집중유발시설	2016.2분기~2018.3분기	분기
차량	도로별 차량통행량	2017	년
	시군구별 차량등록대수	2017	년
기초 정보	행정동/법정동 코드 정보	2018	분기
	소지역정보(1, 2)	1 : ~2018년 2 : 2019년~(개편)	년
기타	SNS데이터	2017	년
	온라인가격정보	2017.1~2022.3	월
	모바일상품권	2018	년

2021년 통계데이터 분석·활용대회

데이터, 새로운 도약을 꿈꾸다

2021년 통계데이터
분석·활용대회

2021. 5. 25. ~ 6. 21.

공모분야

사업분야

참가방법

대상일정

문의사항

SDC 지원금

2021년 통계데이터센터 이용·활용수기 공모전

2021년 3분기
통계데이터센터
이용·활용수기 공모전

공모주제

공모분야

공모자격

공모일정

응모방법

대상

문의처

디즈니랜드 간다고? 난 「KOSIS 통계놀이터」 간다

정승훈 | 통계청 통계서비스기획과 주무관



어린이가 생활 속 통계에 노출되면서 데이터 활용과 해석, 나아가 통찰력을 체험할 수 있도록, 어린이 눈높이로 만든 통계놀이터

통계로 놀이가 가능할까?

4차 산업혁명, AI의 근간은 데이터이다. 데이터 양이 기하급수적으로 증가하는 지식기반 사회가 되면서 복잡한 정보와 지식을 효과적으로 활용할 수 있는 능력, 즉 데이터로부터 유용한 지식과 정보를 탐색하고 창출할 수 있는 데이터 리터러시 교육이 요구되고 있다. 국제사회에서도 미래 사회에 요구되는 역량으로 창의적 사고력, 문제해결 역량, 비판적 사고 역량을 강조하고 있다.

통계청에서 제공하는 국가통계포털서비스는 국내·국제·북한의 통계를 모두 한 곳에 모아 사전 형식으로 신속·정확하게 제공하고 있으나 방대한 자료와 전문용어로 통계 활용 초보자가 친숙하게 접하기는 어려움이 있다.

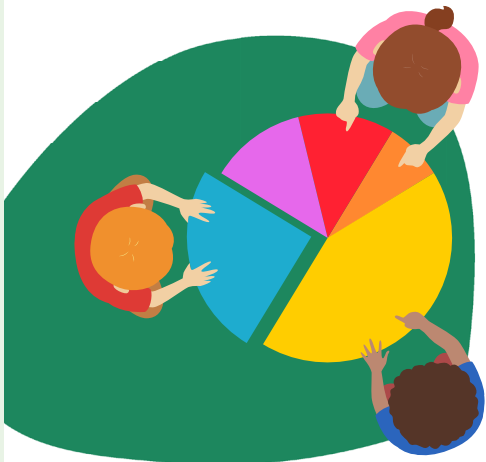
이제는 아이들이 자연스럽게 데이터를 경험할 수 있는 환경을 만들기 위해 많은 통계데이터를 의미 있는 정보로 연결하는 인터페이스 접점을 어떻게 만들어야 어린이들이 그 안에서 뛰어놀고 자연스럽게 데이터를 체득할 수 있을까 고민하였다. 그 결과가 「KOSIS 통계놀이터」이다.

어린이 눈높이에서 다시 들여다본 통계데이터

예전에는 시간적 투입이 가미된 경험에서 인사이트가 생겼지만, 요즘은 데이터를 어떻게 찾고 활용하고 이해하느냐가 바로 인사이트가 된다. 그러나 여전히 데이터에 대한 접근, 데이터를 분석하는 것은 진입 장벽이 높고 어렵다. 특히 우리의 미래인 어린이들은 단어 자체를 어렵게 여기고 여전히 배워야 하는 힘든 공부로 여기고 있다.

학교에서 교과목의 단원이 아니고서는 통계란 그저 분석과 증빙을 위한 수단으로 활용되고 있다. 초등학교 선생님은 혼자서 많은 교과목을 가르치기 때문에, 통계 교육에 대한 의지는 있으나 교과 연구 개발의 한계가 존재한다. 또한 어른들의 관점이 아닌 어린이의 눈높이와 관심사에 맞는 통계데이터도 요구되었다.

이러한 상황에서 어린이가 자연스럽게 통계 경험을 축적할 수 있는 통계서비스 요구가 지속되고 있으며 학교 사회와 교과과정 역시 데이터 리터러시 내용을 추가하였다.



선생님의 생각	“아이들에게 통계를 통해 논리적 사고력을 길러주고 싶어요”
	초등학교5학년의 담임을 하고있는 이해진선생님은 요즘 고민이 크다. 4차산업혁명이 본격적으로 시작되고 체감되는데 이 트렌드의 가장 핵심인 '데이터'에 대한 관심을 학생들이 가졌으면 좋겠지만, 학생들은 관심이 없고 이에 대한 교과 과정 역시 명확하지 않다. 초등학교 담임선생님의 특성상 거의 전 과목을 직접 가르치고 있기에 각 과목이 연결된 내용에서 데이터를 활용하는 통계는 그저 도구로서, 근거의 제시로만 활용될 뿐 직접 데이터를 정리하는 단계를 알려주고 실습할 수는 없었다. 하지만 이러한 마음을 모르는지 학생들은 숫자, 통계에 대한 부정적인 반응을 보이고 있다. 결국 학생들의 흥미를 끄는 동영상으로 대체하여 수업을 진행하였으나, 학생들의 통계에 대한 이해라는 수업목표 달성에는 부족하다. 이에 답답함을 느껴 초등학교사 커뮤니티인 '인디 스쿨'에 접속해 찾아보았으나 수확단원에 대한 정보만이 있을 뿐 통계 교육 카테고리 자체가 없었다. 통계청 홈페이지를 들어가 보니 말 그대로 엄청난 데이터가 있었으나, 활용하기 좋고 보기 좋게 정리된 데이터는 없었다.
어린이의 생각	“어려운 얘기 말고 나와 관련 있는 정보가 필요해”
	12살 초등학교 5학년인 김선호 학생은 평소 활동적이고 활발한 학생이다. 이런 선호는 흥미와 재미있는 것에는 집중력을 발휘하여 집중하나, 그 밖의 것들에는 큰 관심을 두지 않는다. 외향적인 선호는 주말마다 아빠와 근교에 나가 노는 것이 제일 좋고 기다려 진다. 하지만 지난 5주간 매 주말마다 비가 왔고, 이에 실망한 선호는 아빠와 놀러 나갈 비가오지 않는 지역을 찾고 싶었다. 하지만, 일기예보는 우리 동네만을 알려주지 않았고, 인터넷을 통해 찾아본 내용은 틀리기 일쑤였다. 이런 고민을 선생님께 여쭙보니 기사를 찾아보라고 조언해 주셨으나, 같은 지역의 날씨를 정리한 특정한 기사를 찾을 수 없었다. 이어서 선생님은 통계청의 'KOSIS'를 알려주셨는데 들어가자마자 머리가 아팠다. '날씨' 라는 키워드 검색을 통해 무엇인가 나왔지만, 알아들을 수 없는 용어와 그래프는 보자마자 창을 닫게 만들었다. 통계도 모르고 알고 싶지도 않지만 내가 원하는 것을 내 눈높이에 맞춰서 얘기해줄지 두덜거리며 잠을 청했다.

<p>놀이를 통해 통계에 대한 지식과 활용성을 배우게 하자</p> <p>교육유관기관, 교육전문가, 초등학교 선생님 등으로 TF를 구성하고, 수요자-전문가-관련기관과 협업하였다. 또한 서비스를 이용할 어린이가 직접 참여하여 아이디어를 제시하는 등 다양한 국민 참여 채널을 운영하여 서비스 개발의 전 과정에 국민의 의견을 반영하였다.</p>	<p>‘21년 의견수렴, 협업 및 참여 채널 운영 사항</p> <ul style="list-style-type: none"> • 브레인스토밍(5회, 2~3월, 일반인·전문가 22명) • 모니터단(20명, 4~5월, 4회) 의견수렴(45건) • 교수·기자·학생 등 767명(대규모) 의견수렴(3월) • 교사 간담회(6월, 12명), 어린이 온라인 워크숍(21명, 8월, 2회) • 교육유관기관·교수·교사(13명)로 KOSIS-EDU TFT 구성, 상·하반기 회의(4, 11월) • 캐치프레이즈·캐릭터 대국민 공모전(6~8월) • 국민정책디자인 운영(5~9월), 주 중앙부처·지자체 중 우수상 수상
---	---

아이디어	목표
①	<p>소통 모두가 연결되는 공간</p> <p>소통을 통해 스스로 성장할 수 있는 어린이로 만들어라!</p> <p>통계관련 모든 것이 소통되는 공간</p> <ul style="list-style-type: none"> • 어린이-선생님 (수업과제 및 통계활동 모니터링) • 어린이-어린이 (설문지, 통계 포스터 등 자신의 활동을 공유) • 선생님-선생님 (교안 공유) • 어린이-전문가 (통계 관련 질문과 답변)
②	<p>놀이와 배움 통계 놀이를 통한 데이터 리터러시 함양</p> <p>놀이를 통해 통계에 대한 지식과 활용성을 배워라!</p> <p>소극적 → 적극적</p> <p>비주얼 통계 - 통계 마을 - 통계 게임 - 주제선정 - 통계 릴레이 (통계소통)</p> <p>데이터 리터러시</p>
③	<p>흥미 초등학생의 관심 주제와 환경 안에서 통계 놀이터 구성</p> <p>초등학생의 관심 주제를 중심으로 내용을 구성하라! 초등학생의 눈높이에 맞는 용어 선정과 웹 환경을 구축하라!</p> <p>초등학생이 관심 주제를 직접 추천하는 제도 도입</p> <ul style="list-style-type: none"> * (예시) 옆 초등5학년 친구는 몇 명?국가별 초등학생 남녀비율? *통계 마을 등 가상공간을 활용, 실생활 문제해결 중심의 데이터 리터러시 함양 * (예시) 우리 동네 편의점 수는 인구 대비 적절한가?

<p>이러한 수요자 의견을 서비스에 적용하여 설정한 어린이용 통계서비스의 핵심 아이디어는 ‘소통’, ‘놀이와 배움’, ‘흥미’이다. 선생님 대 선생님, 선생님과 어린이, 어린이와 어린이, 어린이와 전문가와의 소통으로, 어렵고 딱딱한 교육이 아니라 놀이 속에서 자연스럽게 체득되는 학습 내용을 바탕으로 한다.</p> <p>「통계놀이터」 안에서 어린이가 놀면서 자연스럽게 통계를 접할 수 있도록 하고, 여기에서 나와 떨어져 있는 어려운 통계가 아닌, 나와 밀접한 통계를 보고 질문을 올리고 전문가가 답을 알려주고, 참여자들과 공유할 수 있게 만들었다.</p> <p>통계로 보고·통계로 말하고·통계로 생각하게 하는 「어린이 통계놀이터」</p> <p>「KOSIS 통계놀이터」는 어린이들이 선생님과의</p>	<p>소통과 보호 속에서 어린이 눈높이의 다양한 통계 데이터를 제공받으면서 다른 학생들과 함께 나누고, 놀이를 통해 자연스럽게 통계를 익히며 데이터 리터러시를 함양할 수 있게 통계청이 제공하는 어린이용 통계활용서비스이다.</p> <p>공신력 있는 통계데이터와 콘텐츠로 통계청만의 차별화된 서비스 제공한다. 핵심서비스는 ‘비주얼 통계’, ‘데이터로 말하기’, ‘프로젝트형 통계 학습’이다.</p> <p>‘비주얼통계’에서는 초등 4~6학년 국정·검정교과서에 수록된 국가승인통계를 어린이가 이해하기 쉽게 단순화하여 직관적으로 표현하였다. 모든 국가승인통계가 수록된 국가통계포털(KOSIS)와 연계하여, 교과서에 활자화된 통계를 최신의 데이터로 업데이트하여 보여준다.</p>
---	---

비주얼 통계

← KOSIS 통계놀이터

통합검색

검색어 입력

로그인

홈

보관함

나의 활동

통계놀이터 활용방법

비주얼 통계

주제별

학년별

인구

교육

문화

교통

환경

기타

데이터로 말하기

프로젝트형 통계 학습

만화로 보는 통계

게시판

Home > 비주얼 통계 > 주제별 > 인구

비주얼 통계

교과서 선택 보기

우리나라 노인인구

댓글 0 조회 228 좋아요 0

우리나라 총인구

댓글 0 조회 167 좋아요 1

우리나라 인구 밀도

다문화 가구 수

교과서 선택 보기

과학

사회

수학

수학익힘

교육부

동아출판

아이스크림미디어

교육사

미래엔

지학사

금성출판사

비상교과서

전재교과서

김영사

비상교육

전재교육

실질관찰

초기화

적용

비주얼 통계에서는 교과서에 수록된 통계를 제공

주제별/학년별로 분류

교과목과 교과서의 출판사도 선택 가능

비주얼 통계(예시) ※ 우리나라 출생아 수(4학년 2학기 사회)



그래프와 통계표로 구성

㉑ 항목과 시점을 선택하여 조회 가능

㉔ 관련통계표에서는 상세한 통계표를,

㉕ 연계교과에서는 통계가 수록된 교과서 정보를 제공

㉖ 이용자는 통계를 공유하거나 보관할 수 있고,

㉘ 댓글로 질문 등 가능

데이터로 말하기(예시)



프로젝트형 통계 학습(예시)



이전에는...

통계놀이터를 통해!

- 어린이는 통계를 복잡하고 어렵게 생각
- 직관적인 인포그래픽 도입으로 정보 전달력 및 이해도 향상
- 데이터 능력을 어디에서 키워야 할지 모름
- 데이터에 기반한 토론으로 데이터 리터러시 능력 향상



‘데이터로 말하기’는 전문가가 선정한 데이터 관련 주제에 대해 어린이가 스스로 해석하고 친구들과 토론으로 참여하는 플랫폼이며, ‘프로젝트형 통계 학습’은 초등학교 선생님이 학교에서 바로 활용할 수 있는 수업자료를 서비스한다.

단계별로 대상과 수록내용을 확대하고 있으며, 초등4~6학년 어린이를 대상으로 한 1단계 구축 결과를 ‘22년 2월부터 서비스하고 있다. 국가통계포털 하위 메뉴(<https://kosis.kr/edu>)에서 이용할 수 있다.

「KOSIS 통계놀이터」는 수많은 데이터를 의미 있는 정보로 바꾸는 징검다리 역할을 수행하면서, 통계적 사고의 조기 개발과 데이터 이해·분석 능력 함양을 지원하는 등 미래인재를 육성하는 기반이 되는 서비스이다.

올해는 2단계 사업으로 선생님용 서비스를 개발하고 콘텐츠와 통계데이터를 확대하고 있다. 앞으로도 지속적으로 노력하여 국민이 원하는 서비스로 발전할 예정이다.



통계와 사회가 만나다, 융합 교과서 「통계와 사회」

정문성 | 경인교육대학교 사회과교육과 교수

「사회」 과목은 어떤 학문인지 궁금합니다

‘사회’ 과목의 목표는 ‘훌륭한 민주시민(good citizen)의 양성’입니다. 인간은 사회 속의 개인으로 살아가므로 사회와 개인은 운명공동체입니다. 살기 좋은 사회를 만드는 것은 곧 사회구성원의 자질에 달려있다고 볼 수 있지요. 특히 민주 사회에서는 모든 개인이 사회의 주인이므로 민주 시민의 자질이 더욱 중요하지요. 사회 과목은 민주시민으로서 갖추어야 할 우리 사회의 법, 정치, 경제, 문화, 역사, 지리 등에 대한 기본적 소양을 배웁니다. 더불어 이러한 지식으로 사회적 삶에서 직면하는 사회현상을 이해하고 사회문제를 해결하는 역량을 기르는 교과입니다.

사회생활에서 통계 활용 능력이 필요한 이유는 무엇인지요

산다는 것은 어떤 행동을 한다는 것이고, 행동한다는 것은 어떤 것을 선택했다는 것이죠. 과거에는 주로 고집, 권위, 직관에 의존한 선택이었습니다. 사회가 발전하면서 좀더 합리적인 의사결정 방법이 필요해졌습니다. 바로 과학적 사고 방법이죠. 과학은 객관적 증거를 사용하므로 사회구성원이 합의하기도 쉽고 무엇이 최선인지 판단하기 쉽지요. 특히 사회생활에서는 통계가 많이 사용되지요. 정치에서는 여론조사, 경제생활에서는 물가, 환율과 주식, 문화생활에서 시청률과 각종 평점 등 일상생활에서 합리적 선택을 하기 위해서는 통계를 활용하는 능력은 필수적입니다.

그럼 사회가 형성되면서 자연히 통계도 생겨났다고 볼 수 있겠네요

수의 개념이 생길 때부터 간단한 통계는 자연스럽게 사용되었다고 봅니다. 역사기록을 보면 기원전 3600년경 고대 바빌로니아 시대나 이집트에서 인구조사를 했고요, 우리나라도 고조선 때부터 인구조사를 했습니다. 특히 농경사회에서는 자연스럽게 농사에 필요한 기상현상을 관측하고 기록한 것으로 보입니다. 우리나라에서는 과학기술을 중시했던 세종대왕이 측우기를 개발하여 강우량을 측정했고요, 최초로 새로운 조세 제도에 대해 여론조사도 했지요. 과반수는 찬성했지만 반대 의견도 상당해서 몇 년간 시범 실시를 하고 전국적으로 실시한 지혜를 발휘한 모습도 흥미롭습니다. 그러나 근대에 이르기까지 통계자료는 일부가 독점하는 자료로 그쳐서 일반인들은 수동적으로 받아들이기만 하는 능동적으로 접근하기에는 어려움이 있었지요.



파피루스에 기록된 고대 이집트의 주택 관련 자료



세종대왕 공법 시행을 앞두고 지역을 고려하여 172,806명을 대상으로 한 여론조사 결과

찬성	반대	계
97,912명	73,384명	172,806명
57.1%	42.9%	100.0%

오늘날 통계가 우리 사회에서 어떤 역할을 하는 건가요

통계 지식이 일상화된 것이죠. 예를 들어 최근 몇 년간 우리는 코로나 확진자 수와 백신 접종 현황 통계를 매일 뉴스로 보고 있습니다. 게다가 통계가 사회문제를 해결하는데 얼마나 중요한지도 체험하게 되었지요. 정확한 통계에 기초하여 방역 정책을 결정하므로 국민이 적극적으로 협조하는 것도 체험했습니다. 지방자치제도의 발달과 함께 잇달아 치러지는 선거 과정에서 여러 차례의 여론조사 발표와 함께 조사 방법과 오차범위 등 구체적 통계정보와 용어에도 익숙하게 되었습니다. 게다가 정보공개정책으

로 대부분 공공기관이나 기업이 통계정보를 공개하고, 인터넷의 발달로 거의 모든 통계자료를 언제든지 누구나 접근할 수 있는 환경 등이 갖추어졌습니다. 이제는 문제를 정확하게 인식하기 위해 통계자료부터 먼저 찾는 것이 습관처럼 되었습니다. 그만큼 통계는 시민 생활에서 문제 해결의 중요한 자료가 되었지요. 이제는 학생들이 통계를 생활 속에서 자주 접하다 보니 어렵다고 생각하기 이전에 일상생활에 반드시 필요한 지식이라는 것을 받아들이고 있는 것으로 보입니다.

사회 과목에서의 통계와 수학 과목에서의 통계의 차이가 있다면 무엇인지요

크게 본다면 수학이 학문적인 이론 중심이라면 사회는 사회생활을 잘하는 방법을 가르친다고 생각하면 됩니다. 예를 들어 수학에서는 표본의 신뢰성이나 자료 해석 방법에 주안점을 두어 가르친다면, 사회에서는 전화 면접이나 ARS나 하는 조사 방법이나,



역사기록을 보면 기원전 3600년경
고대 바빌로니아 시대나 이집트에서
인구조사를 했고요, 우리나라도
고조선 때부터 인구조사를 했습니다.

질문 방법에 따라 조사 결과가 달라질 수 있다는 것에 관심을 두지요. 해석의 관점을 사회적 측면으로 확대해나간 겁니다. 통계를 활용하되 정확한 사회적 의미를 찾는 데 더 관심이 많습니다.

사회 과목은 융합학문입니다. 우리가 사는 세상은 단 한 가지만으로 이루어진 것이 아니라 여러 가지 상황이 종합되어 나타납니다. 사회 과목은 실제 사회와 같이 다양한 요소가 복합적으로 융합된 종합적인 상황에서 합리적인 선택을 할 수 있도록 훈련하는 것이죠.

사회가 발전할수록 점점 더 통계에 대한 요구가 높아지겠군요

사회가 발전하고 복잡해지고, 통계기법도 발전하면서 주로 정부나 공공기관이 정책효과나 정책대안을 마련하기 위해 여러 영역에서 통계를 많이 활용했습니다. 또한 기업들은 매출을 올리기 위해, 이익단체들은 사회문제를 해결하기 위해 여러 영역에서 더욱 구체적인 통계를 활용하게 되었습니다. 특히 대중매체의 발달과 인터넷의 발달이 누구나 쉽게 통계자료에 접할 수 있는 시대가 되었지요.



「통계와 사회」 교과서를 만들게 된 계기가 있다면요

국가교육 과정에서는 사회 과목을 규정할 때 ‘사회현상을 정확히 인식하고’라는 표현과 ‘사회적 문제 해결 능력’의 표현이 나옵니다. 일상생활 속에서 직면하는 사회현상을 정확히 인식하기 위해서 과학적 접근을 해야 하고, 대표적인 자료로 사회통계를 사용하지요. 통계는 연도별로 계속 달라지기 때문에 직접 교과서에 실지 않더라도 수업에서 탐구활동을 할 때는 반드시 통계를 사용합니다. 그리고 이것은 문제 해결 방안을 찾는 단계로 연결됩니다. 따라서 사회 수업에서 적절한 통계자료를 찾아 분석하고 해석하는 것이 중요한 학습활동 중 하나입니다. 그러나 쉬운 일은 아니지요. 아무래도

통계를 어려워하는 학생이 많은데, 이를 능수능란하게 사회현상을 이해하고 설명하는데 적용한다는 것은 더욱 어렵지요. 그래서 사회현상을 이해하는 데에 필요한 기본적인 통계 지식을 공부하고, 우리의 일상에서 접하는 사회현상을 이해하고 분석하는데 어떻게 활용하는지를 집중적으로 배우는 교과서가 필요하겠다는 생각에서 집필하게 되었습니다.

왜 이런 융합 교과서가 필요한지요

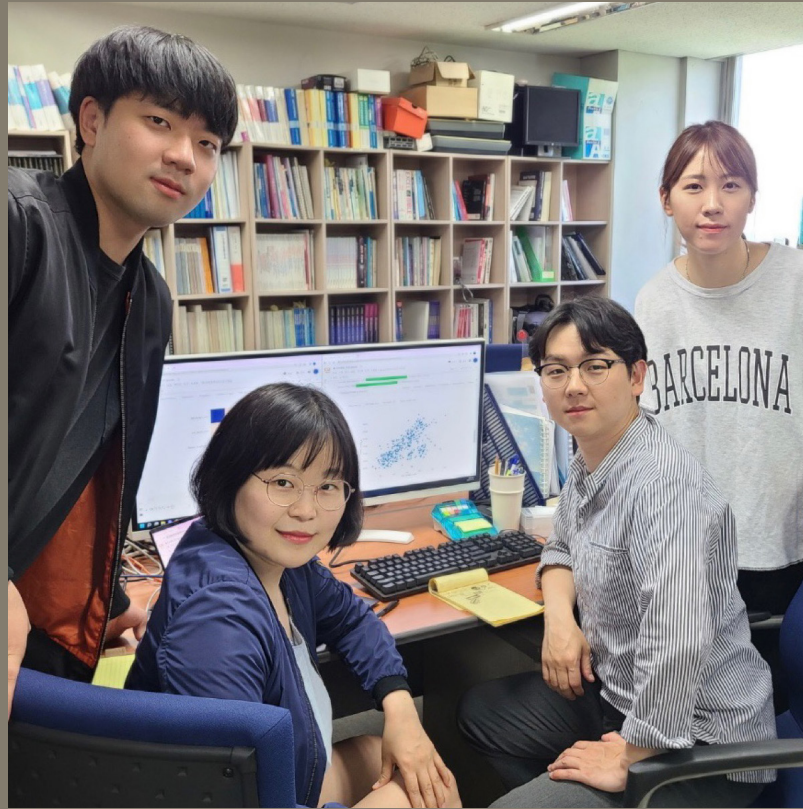
융합에 관심을 두게 된 것은 스티브 잡스의 아이폰 디자인이 화제가 되면서였지요. 창의성의 근간이 서로 다른 영역의 융합에서 출발한 것인데 그것이 상업적인 성공으로 이어지면서 갑자기 주목받았습니다. 그래서 인문학이 주목을 받기도 했습니다. 사실 우리의 삶은 복합적이어서 구분이 따로 존재하지 않습니다. 근대 분석과학이 발달하면서 수많은 학문이 분화되고 더 깊이 있게 발전한 것은 사실이지요. ‘과학’이라는 용어 자체가 분리해서 연구한다는 뜻이니깐요. 우리나라도 합리성을 강조하던 산업화 시대에는 이과와 문과를 가르고, 예체능을 분리하여 인재를 양성했지요. 그러나 포스트모더니즘이 강조되고 창의력이 강조되면서 이과와 문과, 이성과 감성, 정신과 육체 등 융합을 통해 더 현실적이고 현명한 결정과 행동이 요구되는 시대가 되었습니다. 그래서 2015 개정 교육과정과 2022 개정 교육과정 모두 융합 교육과정을 표방하고 있습니다. 사회와 통계도 이러한 세계적 교육 흐름을 반영한 융합 교과서입니다.

「통계와 사회」 교과서의 개발은 어떤 의미를 지니는지요

최초의 사회와 수학의 통합교과서이라는 데에 큰 의미가 있어요. 이러한 임무를 수행하기 위해 사회교육학자, 통계교육학자, 사회 교사가 함께 여러 차례 협의하면서 집필을 했습니다. 쉬운 일이 아니었지만 같은 동기를 가지고 기대하는 바가 있었기에 모두가 즐거운 작업이었습니다. 사회를 이해하고 문제를 해결하기 위한 분명한 목적을 가지고 통계의 기초원리와 방법을 배운다면, 이러한 기초지식으로 우리의 사회문화 생활에 통계가 어떻게 스며들어 있는지를 확인하고, 통계를 통해 어떻게 사회를 이해하고 표현하고 해결하는지를 많은 사례를 통해 배울 수 있습니다.

「통계와 사회」 교과서를 활용하게 될 학생에게 하고 싶은 말이 있으시다면

이 교과서를 통해 사회에서 얼마나 통계가 많이 활용되는지 깨닫고, 통계를 정확히 이해하는 능력을 익힐 수 있었으면 합니다. 배우는 학생들이 우리 삶에 스며든 통계가 어려운 것이 아니라 친숙한 우리 삶의 파트너임을 인식하고, 통계를 통해 현명한 사회생활인으로 성장하기를 기대합니다. 아울러 교과서 자유발행제도가 확산되면 향후 중학교, 초등학교 사회와 통계 교과서도 제작되기를 기대합니다. 끝으로 통계청에서 이 교과서가 잘 활용될 수 있도록 수업에 활용 가능한 인포그래픽이나 자료를 지속적으로 개발·보급해 주면 좋을 것 같습니다.



각본 없는 드라마 속에 숨겨진 비밀을 캐낸다

조선미, 강지연, 이한울, 최윤철 | 명지대학교 스포츠기록분석연구센터

요즘 스포츠 관련 다양한 통계를 보긴하는데요.

스포츠기록분석연구센터하면 아직 낯설게 느끼는 사람들이 많은 것 같습니다.

명지대학교 스포츠기록분석연구센터는 2002년 설립되어, 스포츠 과학을 연구하고 교육하는 기관으로 자리 잡았습니다. 조금 더 구체적으로 말씀드리면 스포츠 경기를 통해 발생하는 다양한 종류의 경기기록 또는 데이터를 체계적이고 일관된 방법으로 수집하고, 객관적이고 과학적인 분석을 통해 의미있는 결과를 도출하는 기관이라고 할 수 있습니다. 이 과정에서 필연적으로 통계가 주요한 요소가 되고 있지요.

스포츠 분야의 데이터는 어떻게 수집되고 어떤 특징이 있나요.

스포츠 경기 분석 혹은 스포츠 데이터 분석은 대부분 기획의 성격으로 이루어집니다. 분석의 목적부터 설계가 과정 중심으로 이루어지고 있죠. 그래서 목적에 따라 다루지는 데이터의 종류와 수준은 매우 다양하게 존재합니다. 형태상으로 분류를 한다면, 영상이나 숫자, 문자로 된 데이터도 될 수 있고, 주제로 분류를 한다면 경기 퍼포먼스와 관련이 있는 선수 데이터 및 팀 데이터, 경기 외적인 경기 정보로 분류할 수 있습니다. 최근에는 빅데이터 분석 방법론이 활용되면서 경기 외적인 미디어 데이터 또는 SNS 등의 비정형 데이터도 스포츠과학 영역에 포함되고 있는 실정입니다.

최근에는 IT기술이 스포츠 영역에서 매우 유용하게 활용되고 있는데요. 그로 인해 스포츠 경기분석을 위한 데이터 수집 방법의 메커니즘에 혁신이 이루어지고 있습니다. 예를 들어 과거 데이터 수집은 부호화를 바탕으로 하는 수기록, 즉, 경기 퍼포먼스에 대한 내용을 분석자가 직접 수집을 하는 경우가 대부분이었습니다. 수집되는 데이터의 양과 수준에 어느 정도 제한이 있던 것도 사실이죠. 그러나 최근에는 AI 또는 IOT 기반의 센서, 레이더 등을 이용한 선수 퍼포먼스 데이터를 전사적으로 수집할 수 있는 환경이 조성되었습니다. 한 동작 또는 한 선수, 한 팀에 대하여 테라바이트 급의 빅데이터들이 자동 수집되고 있습니다. 따라서 스포츠 데이터 분석의 영역의 범위와 내용이 다양해지고 넓어졌습니다.

이런 스포츠 데이터는 어떻게 활용되고 있는지요

경기 분석된 스포츠 통계 데이터의 활용영역은 크게 세가지 부문으로 분류 됩니다. 첫 번째는 스포츠 현장입니다. 경기의 승리 또는 더 나은 경기력 향상을 위한 분석자료로 활용되기도 하고 스포츠 선수의 연봉 협상의 기초자료로 활용되기도 합니다. 두 번째는 스포츠 연구 영역이죠. 스포츠 통계데이터를 보다 쉽게 분석하고 가치를 창출하기 위한 방법론을 개발하고, 학술지 등을 통해 연구 결과를 공유함으로써 스포츠 경기분석의 학문적 발전에 이바지하고 있는 것이죠. 세 번째는 스포츠 산업 분야입니다. 최근 ICT 기술의 발전과 더불어 더욱 활용 가치가 높아지고 있는 부분입니다. 최근에는 운동과 건강에 대한 사회적 관심이 고조되고, 그 저변이 확대되고 있습니다. 소위 ‘보는 스포츠’에서 ‘하는 스포츠’로 생활체육을 즐기는 분들이 많아지고 있고, 엘리트 선수가 아니더라도 개인이 자신의 기록을 확인하고 서비스받을 수 있는 제품들이 생산되고 있습니다. 종목에 따라 어떤 통계 데이터를 어떠한 방식으로 서비스하느냐가 굉장히 중요하게 되었죠. 뿐만 아니라 미디어에서도 전문적인 스포츠 통계 서비스가 활용 되고 있습니다. 단순히, 경기 결과를 제공하는 서비스 보다 좀더 전문적인 스포츠 통계 데이터의 수요가 증가하고 있죠. 예를 들어 연구센터에서도 축구 경기 중계가 있을 때 실시간으로 분석데이터 서비스를 제공하고 있습니다. 실시간 분석을 통해서 슈팅, 코너킥, 파울 등의 빈도 데이터 뿐만 아니라 볼 점유율, 공격점유율, 패스패턴 및 개인선수의 히트맵 등 통계데이터를 상황에 맞춰서 서비스하고 있죠.

이 글을 읽으시는 많은 분들이 스포츠연구센터에서 어떤 분들이 일하시고 계시나 궁금할 것 같네요.

연구센터는 기본적으로 연구를 기반으로 하고 있는 기관이기 때문에, 석사 및 박사생으로 구성되어 있습니다. 무엇보다도 학부 전공의 다양성을 가지고 있다는 것이 첫 번째 특징이라고 할 수 있습니다. 현재 우리 센터 연구원들의 전공은 정보학, 컴퓨터공학, 수학, 경영학, 체육학 등 다양한 배경지식을 가진 멤버들이 모여있습니다. 그리고 두 번째 특징은 스포츠를 사랑하는 사람들이 모여있다는 것입니다. 단순히 즐기는 정도가 아니라 소위 매니아라고 할 수 있을 정도로 한 팀이나 한 종목에 정성을 다하는 사람들이라고 할 수 있습니다.

스포츠 데이터를 다루는데 갖춰야 하는 역량은 무엇인지요.

스포츠 데이터를 다루고, 스포츠 경기를 분석하는데 반드시 필요한 역량이 무엇이라고 명확히 규정하지는 않습니다. 하지만 방대한 양의 스포츠 데이터에서 핵심적이고 의미있는 통계데이터를 발견 또는 분석하기 위한 통찰적인 시각이 중요합니다. 단순히, 경기를 보고 또는 스포츠 데이터를 수집하는 것이 아니라 전체적인 경기 흐름 속에서 인사이트를 찾아내는 것이 중요하고, 그 과정은 오랜시간 동안 트레이닝이 필요한 부분입니다.

아울러 최근에는 통계를 기반으로 하는 빅데이터가 사회적으로 중요한 이슈이기 때문에 AI나 통계적인 역량을 갖추는 것이 스포츠 데이터를 다루는데 효율적이라고 생각합니다.

스포츠 데이터를 다루는데 주로 사용하는 분석툴이 있나요.

사용하고 있는 분석 프로그램은 분석 목적에 따라 다양하게 사용하고 있습니다. 영상 분석을 목적으로 한다면, 영상분석에 최적화된 분석도구를 사용하고, 데이터 분석을 목적으로 한다면 데이터 분석에 최적화 된 도구를 상황에 따라 사용하고 있습니다. 상용제품을 사용하기도 하지만, 축구 종목의 실시간 분석 프로그램은 우리 센터에서 자체적으로 개발한 분석 프로그램을 사용하고 있습니다.

아울러, 스포츠 경기분석 이외에 AI 관련 알고리즘 연구도 병행하고 있습니다. 머신러닝이나 딥러닝 기반 기술을 통해 스포츠 영역에서 예측과 통찰 그리고 의사결정의 피드백을 실현하고 있다고 할 수 있는데요. Pose Estimation 등과 같은 기술을 활용하여 동작기반 데이터를 수집하고, 역학적인 움직임을 분석하거나, 딥러닝 알고리즘을 활용하여 경기의 승패예측 등과 같은 연구를 진행하고 있습니다.



스포츠 데이터가 효과적으로 사용된 국내외 사례가 있나요.

빌 제임스는 단지 야구를 좋아하는 통조림 공장 경비원이었습니다. 매일 전 경기를 자신만의 데이터베이스로 남기는 등 당시 관점에서 보자면 전혀 쓸모없어 보이던 일을 되풀이하였다고 합니다. 어떻게 하면 야구라는 경기의 진면목을 더 효과적으로 숫자로 나타낼 수 있을지 고민하며 새로운 통계수치들을 개발해내었죠. 지금 세이버매트릭스 지표들 중 상당수가 빌 제임스에 의해 개발된 것들이라고 합니다. 이후 빌 제임스를 중심으로 만들어진 미국야구연구협회(SABR)는 세이버매트릭스라는 방법론을 창시하게 되는데 이 세이버매트릭스는 야구를 통계학적, 수학적으로 분석하는 방법론입니다.

90년대 중반 “머니볼”이라는 소설로 유명한 오클랜드 애슬레틱이라는 팀은 세이버매트릭스를 활용하여 스포츠 데이터를 사용하는 새로운 길이 열렸고 보스턴 레드삭스라는 팀도 1918년 이후 월드 시리즈 우승이 없었지만 ‘세이버매트리션’을 대거 기용한 후 2000년대 3번의 우승을 달성한 사례를 통해 점차 많은 구단이 세이버매트릭스 개념을 받아들이면서 빌 제임스도 이름을 널리 알리게 되는 계기가 되었습니다.

이밖에도 2014 브라질 월드컵에서 독일 월드컵 대표팀은 SAP의 인메모리 플랫폼인 HANA 기술에 기반 한 SAP 매치 인사이트를 도입함으로써 독일 대표팀의 경기력 향상과 월드컵 우승을 일조한 사례까지 스포츠 데이터를 효과적으로 사용한 예시로 들 수 있겠네요.

요즘 스포츠 연구센터에서 중점을 두고 진행하는 일은 무엇인지요.

잘 아시겠지만, 올해는 FIFA월드컵과 아시안게임이 개최될 예정입니다. 아시안게임은 연기를 것 같네요. 이런 대규모의 스포츠 이벤트는 전 세계인이 주목하는 축제이기도 하고, 우리 센터에게도 다른 의미의 축제라고 할 수 있습니다. 데이터 수집과 분석을 위한 장이 펼쳐진다고 할 수 있죠. 대회의 분석을 위해 현재 각종 정보들을 수집하고, 지역에선 등과 같은 선행 경기 결과를 분석하고 있습니다.

스포츠 데이터 세계를 꿈꾸는 분들에게 들려주고 싶으신 얘기가 있다면요.

현재는 데이터가 굉장히 중요한 시대입니다. 특히, 스포츠 통계데이터는 어떠한 경기의 내용 또는 경기력을 설명하고 묘사하는데 강력한 파워를 가지고 있죠. 빅데이터의 발전과 더불어 스포츠과학영역의 스포츠 통계는 더욱 발전할 것이라고 예상됩니다. 하지만 다양한 통계 데이터의 생산과 더불어 함께 중요한 것은 그것에 대한 가치와 해석, 통찰적 시각으로 접근해야한다는 것입니다. 목적과 상황에 맞는 스포츠 통계 데이터를 적재적소에 빠르게 배치하는 것이 매우 중요하죠. 마치 흙속에서 진주를 찾아내는 것과 같다고 할 수 있습니다. 기하급수적으로 증가하는 스포츠 데이터에서 가치 있는 진주를 찾아내는 것이 빅데이터 시대에 스포츠 통계를 맞이하는 우리의 자세라고 할 수 있습니다.

쏟아지는 여론조사, 왜 들쭉날쭉해 보일까

신정인 | 넥스트리서치 차장



20대 대통령선거, 2달간 283건 선거여론조사 진행

여론(輿論, public opinion)은 “일반적으로 중요한 쟁점에 관해 상당수의 사람이 표현한 선호의 복합(complex of preference)”이라고 할 수 있다. 사적인 주제에 대한 사적인 의견이 아니라, 한 사회의 구성원으로 공공의 상황과 관련된 문제에 대한 개인들의 의견의 합이 여론이다. 이러한 여론을 파악하기 위해 대표성 있는 표본(sample)을 선정하여 질문을 하고 응답을 받아 정보를 수집하여 조사 대상인 모집단(population)의 특성을 파악하는 것이 바로 여론조사다.

선거제도는 민주주의 사회에서 개인의 의견을 표현하는 가장 대표적인 방법이다. 후보들의 정치적 성향이나 공약 등을 평가하여 자신의 선호를 가장 잘 대변하여 정책에 반영해 줄 대리인을 뽑는 과정이기 때문이다. 민주주의 국가에서 선거만큼 단기간에 공공의 관심이 집중되고 이해가 나뉘는 이슈도 없을 것이다. 그래서 선거기간이면 유권자들의 의견을 물어보는 각종 여론조사가 쏟아지는 것은 자연스러운 현상이다.

우리나라 대통령 선거에서는 여론조사 결과를 단순히 후보 간 현재 지지도를 확인하는 것을 넘어, 후보자 간 단일화에 활용하기도 한다. 2002년 노무현-정몽준 간 단일화에서 노무현 당시 후보로 단일화되는 과정에서 처음 여론조사 방식을 도입했다. 이번 20대 대통령선거에서도 국민의힘 윤석열 후보와 국민의당 안철수 후보도 단일화 방식에 대해 협의할 때 안 후보 측은 계속해서 여론조사에 의한 단일화 방안을 고수하기도 했다.

올해 3월 치러진 20대 대통령선거를 앞두고 2022년 1월~2월 두 달 동안 중앙선거여론조사심의위원회에 등록된 대통령선거 관련 여론조사는 총 283건이었다. 하루 평균 4.8건의 여론조사가 진행된 것이다. 2017년 5월에 있었던 19대 대선 직전 두 달 동안 174건의 대선 관련 여론조사가 있었던 것에 비하면 이번 대선 전 2달간 진행된 조사는 약 1.6배 늘었다.

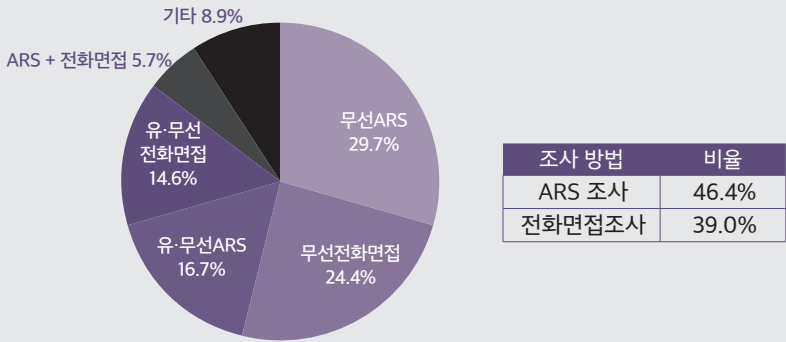
여론조사 이해! 조사방법에 따라 차이가 있다

그러나 선거기간이면 난무하는 여론조사 결과에 대한 보도는 단순 여론조사 피로도를 넘어, 신뢰도에까지 의문이 제기되기도 한다. 후보별 지지도가 짧은 시간에 뒤집히기도 하고, 같은 날 발표된 조사에서 두 후보 간 지지도가 상반된 결과가 나오기도 하기 때문이다. 그러나 이는 조사방법에 따른 차이를 이해하지 못하고 단순 수치의 등락에 주목하기 때문에 발생하는 오류일 가능성이 높다.

조사 방법에는 전화면접조사와 ARS가 있다

대부분의 선거여론조사는 전화조사를 활용한다. 전화조사는 전화조사원에 의한 전화면접조사와 ARS(Automatic Response System, 자동응답시스템) 두 가지 방법으로 나뉜다. 전화면접조사는 조사원이 직접 전화를 걸어 응답자와 통화를 통해 응답받는 방식이고, 자동응답시스템(ARS)은 미리 녹음된 기계음을 듣고 응답자가 전화기의 키패드를 눌러 응답을 하는 방식이다.

올해 3월 치러진 20대 대선을 앞두고 1~2월 두 달간 진행된 246개 여론조사(지역: 전국, 선거구분 : 20대 대통령 선거 기준) 조사 방법별 비율을 살펴보면 아래와 같다. 무선ARS가 29.7%로 가장 많고 다음으로 무선전화면접이 24.4%, 유·무선ARS가 16.7%, 유·무선전화면접이 14.6%였고, ARS와 전화면접을 섞은 방식도 5.7% 있었다.



[그래프1] 2022년 1~2월 진행된 조사 방법별 비율 (중앙여론조사심의위원회 참조)

그래프1에서 보듯, ARS는 46.3%, 전화면접방식은 39.0%로 ARS 방식으로 진행되는 조사가 더 많다. 이는 전화면접조사에 비해 ARS조사가 상대적으로 저렴하고 빨리 조사를 진행할 수 있기 때문이다. 조사원이 진행하는 전화면접방식은 인건비가 들기 때문에 ARS에 비해 상대적으로 조사 단가가 높다. 예를 들어, 전국 1,000개 표본 전화면접조사를 진행하기 위해서는 숙련된 조사원 20명 이상이 하루 8시간 (혹은 그 이상) 이들은 근무해야 조사가 완료된다. 또한 전화면접조사의 경우 준비단계에 시간이 좀 더 소요된다. 조사원도 미리 모집해야 하고, 전화조사원이 앉아서 조사를 진행해야 하는 좌석 개수의 한계로 회사 내 다른 전화조사 스케줄이 미리 잡혀 있을 경우 조사가 어려운 점도 있다. 그러나 ARS의 경우 그런 제약이 적기 때문에 상대적으로 신속하다.

조사방법에 따른 응답률 차이를 이해하자

언론에 가장 자주 보도되는 여론조사 해석 시 유의점은 아마 ARS와 전화면접조사 결과를 단순 비교하면 안 된다는 내용일 것이다. 그 이유로 가장 많이 언급되는 것은 응답률이다. 현재 우리나라 여론조사심의위원회에서 사용하는 ‘응답률’ 개념은 여론조사에 참여한 사람 중 조사 참여를 거절하거나 중간에 전화를 끊지 않고 끝까지 응답한 사람의 비율이다. 예를 들어, “표본 1,000명(응답률 10%)”란 10,000명의 사람과 통화에 성공하고, 그중 1,000명만 마지막 문항까지 조사에 응했다는 것을 뜻한다. 아래 표1에서 보면 조사방법별 응답률 평균값을 확인할 수 있다. 무선전화면접과 유·무선전화면접조사의 응답률 평균은 각각 19.40%, 15.50%로 나타났지만, 유·무선ARS, 무선ARS는 각각 8.40%, 7.10%로 전화면접조사 대비 낮게 나타났다.

No	조사 방법	비율
1	무선전화면접	19.40%
2	유·무선전화면접	15.50%
3	ARS + 전화면접	8.60%
4	유·무선ARS	8.40%
5	무선ARS	7.10%

[표1] 2022년 1~2월 진행된 조사 방법별 응답률 평균 (중앙여론조사심의위원회 참조)

전화면접조사와 ARS 조사 응답률의 차이는 조사 특성에 기인할 것이라는 설명이 가장 일반적이다. ARS는 기계음이다 보니 상대적으로 해당 이슈에 관심이 없는 사람이라면, 끝까지 기계음을 듣고 응답하지 않고 중도 이탈 혹은 거절할 확률이 높아진다. 반대로 해당 이슈에 대해 관심이 많은 사람이라면 기계음을 끝까지 듣고 응답할 경향이 높을 것이다. 그러다 보니 ARS 조사에는 “정치 고관여층이 많이 응답한다”는 이야기들을 한다. 전화면접조사의 경우, 숙련된 조사원들이 일단 접촉된 표본의 응답 성공률을 높이기 위해 총력을 다한다. 전화를 끊으려는 응답자라도 전문조사원이 조사 진행을 유도하게 되면 마지막까지 응답하게 되는 경우가 ARS에 비해서는 높을 수 있게 되는 것이다. 이 과정에서 정치 등 전화조사가 진행되는 이슈에 평소에 큰 관심이 없는 응답자라하더라도 표본에 포함되어 조사가 진행될 확률이 높아진다. 이 과정에서 표본 대체가 더 적게 발생하기 때문에 “만 18세 이상 일반 국민”이라고 하는 표본의 대표성이 조금 더 확보될 수 있다.

이는 실제로 유사한 문항이라 하더라도 ARS 조사에서 “모름/무응답”을 응답하는 비율보다 전화면접조사에서 “모름/무응답” 응답 비율이 더 높은 경향성으로 확인할 수 있다. 다음은 2021년 12월(20대 대선 3개월 전) 한 달 동안 진행된 2개의 ARS(무선ARS 100%) 조사와 전화면접조사(무선전화면접조사 100%)의 조사 결과를 비교한 표이다. 지지하는 대통령 후보를 묻는 질문에서 전화면접조사의 경우, “없다”, “모름/무응답”의 합, 즉 부동층 비율이 한 달 평균 21.5%로 나타났지만, 무선ARS에서는 이 비율이 9.1%로 나타났다. 전화면접조사에서는 마음을 정하지 못했다는 응답자가 21.%가량인 반면, ARS 조사 응답자의 분포는 그 비율이 9.1%로 상대적으로 낮은 편이고 이는 정치에 관심이 상대적으로 많은 사람들이 더 많이 응답했다고 볼 수 있는 것이다. 그렇기 때문에 같은 시기에 조사된 결과라고 해도, ARS 조사와 전화면접조사 결과를 단순히 비교하는 것은 적절하지 않을 수 있다.

[표2] 무선ARS와 무선전화면접조사간 부동층(없음/모름/무응답) 차이 (중앙여론조사심의위원회 참조)

조사 방법					조사 결과
조사기간	조사기관	조사의뢰자	조사방법	피조사자선정방법	부동층(없음/모름/무응답)
11/29~12/01	엠브레인퍼블릭, 코리아리서치	한국리서치, 엠브레인퍼블릭, 코리아리서치, 케이스탯	무선 전화면접(100%)	휴대전화가상번호	23.0
12/06~12/08	케이스탯, 한국리서치				17.0
12/20~12/22	한국리서치, 엠브레인퍼블릭				25.0
12/27~12/29	코리아리서치, 케이스탯				19.0
2021. 12. 전화면접조사 없음/모름/무응답 평균					21.0
12/03~12/04	(주)한국사회여론연구소(KSOI)	TBS	무선ARS (100%)	휴대전화가상번호	9.4
12/10~12/11					8.5
12/17~12/18					8.4
12/24~12/25					10.0
2021. 12. ARS 없음/모름/무응답 평균					9.1

여론조사 이해! 표본추출 방법을 알아

여론조사를 위한 전화조사에서는 할당추출법(quota sampling)이 주로 사용된다. 할당추출법은 비확률 표본추출로 모집단(population)을 상호 배타적인 부모집단(예를 들어, 성, 연령, 지역과 같은 할당 층)으로 나누고 이 부모집단별 구성 비율과 표본의 부모집단별 구성비율이 유사하도록 표본을 작위적으로 추출하는 방법을 말한다. 대부분의 전화 여론조사의 경우, 전국을 성별, 연령대별, 권역별 모집단 분포를 이용해 표본 각 층의 할당된 표본 수가 결정되고, 배분된 할당 수만큼 표본을 비확률적으로 추출하여 채우는 방식으로 조사를 진행한다.

좀 더 쉽게 설명하자면 다음과 같다. 아래 표 4와 같이 전체 조사 모집단(전국 만 18세 이상 남녀)을 성별, 연령대별, 권역별로 나눈 후, 표5와 같이 1,000명 기준으로 성 * 연령대 * 권역별 비례배분한다. 실제 조사가 진행될 때는 추출된 전화번호에 전화를 걸어, 표4의 할당표를 기준으로 각 셀 별에 할당된 숫자를 채우는 방식으로 조사가 진행된다.

[표3] 전국 만 18세 이상 남녀 모집단 분포(2022년 3월 말 행정자치부 주민등록 인구통계 기준)

구분	남자					여자				
	만 18 ~ 29세	만 30 ~ 39세	만 40 ~ 49세	만 50 ~ 59세	만 60세 이상	만 18 ~ 29세	만 30 ~ 39세	만 40 ~ 49세	만 50 ~ 59세	만 60세 이상
서울	17	16	16	17	24	19	16	17	17	29
인천/경기	30	27	32	32	38	27	25	31	31	44
대전/충청/세종	10	8	10	10	15	8	7	9	10	18
광주/전라	9	6	8	10	15	7	6	8	9	19
대구/경북	8	7	9	10	15	7	6	8	10	18
부산/울산/경남	13	11	14	15	22	11	10	13	15	27
강원/제주	4	3	4	4	7	3	3	4	4	8

[표4] 전국 성·연령대·권역별 1,000명 할당(2022년 3월 말 행정자치부 주민등록 인구통계 기준)

구분	남자					여자				
	만 18 ~ 29세	만 30 ~ 39세	만 40 ~ 49세	만 50 ~ 59세	만 60세 이상	만 18 ~ 29세	만 30 ~ 39세	만 40 ~ 49세	만 50 ~ 59세	만 60세 이상
서울	17	16	16	17	24	19	16	17	17	29
인천/경기	30	27	32	32	38	27	25	31	31	44
대전/충청/세종	10	8	10	10	15	8	7	9	10	18
광주/전라	9	6	8	10	15	7	6	8	9	19
대구/경북	8	7	9	10	15	7	6	8	10	18
부산/울산/경남	13	11	14	15	22	11	10	13	15	27
강원/제주	4	3	4	4	7	3	3	4	4	8

전화조사에서는 “전화 걸기”를 통해 표본을 추출한다. 그러므로 유선인지, 무선인지, RDD(Random Digit Dialing)인지, 가상번호인지, 어떤 표본추출틀을 활용해 추출된 전화번호에게 전화를 거는지가 중요하다. 최근 여론조사에서 일반적으로 사용되는 방식은 RDD와 가상번호이다. 2022년 1~2월 동안 공표된 20대 대통령선거 여론조사에서 가상번호는 33.5%, 무선 RDD 25.9%, 유무선 RDD 27.2%, 유선RDD + 가상번호는 13.4%가 사용됐다.

휴대전화 가상번호

이용자의 이동전화번호가 노출되지 않도록 여론조사에 사용할 전화번호를 가상의 번호(예: 050-XXX-XXXX) 형태로 변환하여 여론조사 기관에 제공하는 것으로, 여론조사 응답자의 개인정보 노출 없이 지역별·성별·연령대별로 대표성을 확보하여 여론조사 결과의 신뢰성을 높이기 위해 2016년 최초 도입된 제도이다.

이는 국내 통신 3사(SK텔레콤, KT, LG U+)를 통해 받는 것으로 지역·성·연령대 특성에 따라 번호를 추출하여 총화추출이 가능하다는 장점이 있다. 그러나 가상번호 신청을 위해서는 조사 시작 10일 전에 신청해야 하고, 비용이 발생한다는 점에서 이용에 일정 부분 제약이 있을 수 있다. 또한 2022년 2월 말 기준 무선통신서비스 가입현황에 따르면 MVNO(Mobile Virtual Network Operator, 알뜰폰) 가입자 현황이 1,000만 명을 넘었다. 가상번호에는 이들 알뜰폰 가입자가 배제된다는 한계가 있다.

무선RDD

무선 RDD는 010 뒤에 있는 번호를 무작위로 생성하여 만든 전화번호를 투입하여 피조사자를 선정하는 방식이다. 그렇기 때문에 전화번호에 대한 어떠한 정보(성·연령·지역 등)도 가지고 있지 않다. 이는 곧 무선전화 이용자(통신 3자 가입자 + 알뜰폰 가입자)라면 누구나 같은 확률로 추출될 수 있다는 뜻이기도 하여 가상번호보다 대표성이 높을 수 있다. 그러나 동시에 정보가 없어 모르는 전화번호에서 걸려온 전화를 잘 받지 않는 특성이 있는 계층(여성, 고령자 등)에 대한 접근이 어렵다는 단점이 있다.

유선RDD

유선RDD는 지역번호 뒤에 있는 국번을 무작위로 생성하여 만들어낸 전화번호를 통해 응답자를 찾는 방식이다. 유선RDD가 활성화되기 전, 즉 2011년 이전까지 대부분의 여론조사기관에서는 KT 가구전화 등재번호부에서 전화번호를 무작위로 추출하는 방식을 사용했다. 그러나 이 방식은 KT이외의 가구 전화나 전화번호 공개를 원하지 않는 가구번호는 표집 틀(Sampling frame)에서 원천적으로 배제되는 등의 한계로 인해 유선RDD가 등장했다. 유선RDD는 지역번호와 국번이 있어 시/군/구 단위에 대한 지역 정보는 포함하고 있다. 그러나 집전화 보유가 계속해서 줄어들고 있고, 전화조사가 진행되는 시간대에 집에 사람이 없는 경우 등이 있어 최근에는 대부분의 여론조사에서 무선전화 표본추출틀의 한계를 보완하는 차원에서 20% 이내로 사용하고 있다.

이렇듯 어떤 표본추출틀을 사용했는지가 곧 어떻게 조사 대상자를 선정했는지를 결정하기 때문에 표본추출틀이 다른 조사 결과를 단순히 비교하는 것은 지양해야 한다.

여론조사의 한계를 알아야 제대로 활용할 수 있다

한국기자협회의 선거여론조사보도준칙 제3조 “여론조사의 한계”에는 다음과 같이 규정했다.

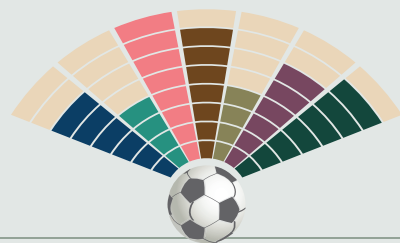
“여론조사는 여론을 탐색하는 많은 방법 중의 하나이며, 여론조사의 수치는 여론을 이해하는 데 도움이 되는 하나의 자료이다. 여론조사를 통해 얻은 수치가 곧 여론 그 자체는 아니므로 미디어는 여론조사 결과를 여론과 동일시해서는 안 되며, 수치에 과도한 의미를 부여해서는 안 된다.”

여론조사는 직관적으로 이해하기 좋은 “숫자”의 형태로 그 결과가 나타난다. 그렇기 때문에 선거기간이면 조사 방법이나 표본추출 방법이 다른 여론조사 결과를 비교해 후보들 간의 우위를 확인하고 지나친 의미를 부여하는 기사들을 종종 볼 수 있다. 그러나 위에서 언급한 조사 방법이나 표본추출틀에 대한 이해 없이 단순히 지지도 결과를 비교하는 것은 여론을 잘못 이해할 소지를 제공하고, 여론조사에 대한 불신을 가져올 수 있다는 측면에서 조심스러울 필요가 있다.

선거여론조사보도준칙 제16조 “오차범위 내 결과의 보도”에서는 신뢰구간에 따른 표본오차를 감안하고, 오차범위 내 지지도 격차가 발생한 경우, 순위를 매기거나 서열화하지 않고, 수치만을 나열하여 헤드라인으로 쓰지 못하도록 규정했다. 또한 제17조 “조사 결과의 비교”에서는 서로 다른 시점에서 실시된 여론조사는 그 조사방법이 동일한 경우에만 상호비교가 가능하고, 지지도의 상승과 하락에 대한 표현을 할 경우에도 기존 여론조사 결과의 추세, 최대 표본오차, 표본 수 등을 종합적으로 고려해 신중하게 보도해야 한다고 되어있다. 더 설명할 것도 없이 이 두 개 조항은 언론인뿐 아니라 여론조사 결과를 제대로 이해하는데 가장 핵심이 되는 준칙이라고 할 수 있다.

스포츠 경기! 승리에 앞서 무엇이 중요한가

김정윤 | 웨슬리퀘스트 상무



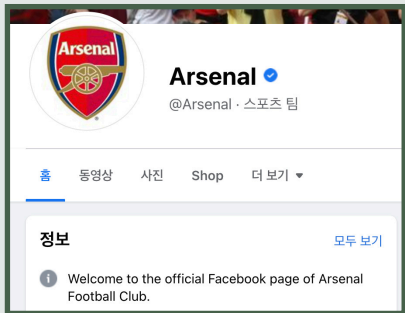
프로 스포츠 구단 운영 목적은 스포츠적인 가치(Sports Value, 승리, 우승 등)와 사업적 가치(Business Value, 매출 증대 등) 2가지 영역에서 모두 탁월한 성과를 창출하는 것이다. 그리고 데이터의 활용은 두 가지 영역에서 모두 매우 활발하게 이루어지고 있다. 본 글에서는 비즈니스 가치(Sports Business Value) 영역에서 데이터를 사용하는 이야기를 해보려고 한다.

Win is Not Enough! 승리만으로 충분하지 않다.

프로 스포츠 관련 컨설팅을 하면서 EPL(영국프로축구) 명문 구단인 아스날의 CEO를 만나 2시간 이상 대화를 하며 성공적인 구단 운영에 대한 이야기를 들었던 소중한 기회가 있었다. 그때 들었던 가장 기억에 남는 이야기는 “Win is not Enough!(승리만으로는 충분하지 않다)”였다. 세계 최고의 명문 구단이 되기 위해서는 승리하고 우승하는 것 이외에 팬들에게 영감을 줄 수 있는 구단의 비전이 필요하고 팬들이 구단으로부터 지속적인 관심과 사랑을 받고 있다고 느끼고 경험하게 하는 것이 중요하다고 이야기해 주었다.

아스날의 핵심 가치 중에는 “함께 만드는 승리”가 있다. 이 핵심가치를 어떻게 고객에게 전달하는지 페이스북 사례로 설명이 이어졌다. 아스날은 페이스북에서 3,700만명이 넘는 팔로워를 보유하고 있다. 이 중 실제로 런던을 방문하여 아스날 경기 티켓을 살 수 있는 팬은 5% 이하라고 한다. 그래서 아스날은 “Contact Globally, Engage Personally”라는 팬 서비스 방향을 설정하여 전세계에 있는 모든 아스날 팬이 경기장에 직접 방문하지 못해도 항상 함께하면서 승리를 만들어내는 경험을 전달하기 위해 노력한다고 했다. 그리고 이 모든 과정은 데이터를 기반으로 고객을 이해하는 것이 출발점이라고 강조했다.

미팅의 마지막 단계에서 아스날 CEO는 현재 구단이 새로운 CRM도입을 전사적인 프로젝트로 진행하고 있는데 100년 이상 된 구단 역사에서 가장 흥분되는 프로젝트라고 자랑하기도 했다. 고객을 알아가고 맞춤형으로 서비스하면서 더 많은 관심과 사랑을 받는 미래를 만들기 시작했기 때문이라고 했다.



<페이스북 아스널 계정, Follower가 3,700만명이 넘습니다>

팬을 연구해서 의미 있는 서비스를 창출해야 한다

프로 스포츠 경기를 관람하는 고객은 매우 다양한 니즈(Needs)로 구성되어 있다. 친구와 놀러가서 응원을 하며 스트레스를 해소하는 목적일 수도 있고, 가족과 즐거운 시간을 보내는 수단일 수도 있으며, 좋아하는 선수를 직접 보고 만나서 사진도 찍고 사인 받는 경험을 하기 위한 이유로 경기장을 방문하기도 한다. 그러면 구단은 각기 다른 목적을 가진 고객에게 각기 다른 서비스를 제공하는 것이 필요하다. 예를 들어서 코트에서 가장 가까운 자리를 구매해서 선수를 보다 생동감 있게 보고 싶어하는 고객에게 무료로 입장하는 서비스를 제공하는데 그 자리가 3층이라면 이 서비스는 아무런 의미를 갖지 못하는 것이다. 그래서 스포츠에서 고객 데이터 분석은 먼저 방문의 목적을 알아가는 것이 필요하다.

하지만 아직도 한국의 많은 구단은 팬들이 경기장을 방문하는 목적을 세부적으로 구분하여 맞춤형으로 서비스하기 보다는 승리와 우승이 가장 큰 목적으로 운영되고 있다. 처음에 이야기한 스포츠적 가치(Sports Value)와 비즈니스적 가치(Business Value)가 조화롭게 성과를 창출해야 하는데 스포츠적 가치(Sports Value)쪽에서만 치우친 상태의 운영이 대부분이다. 한국 프로 배구에서 가장 강한 팀이었던 삼성화재가 성적이 떨어지면서 팬들에게 외면 받고 있는 것과 한 때 아이돌 이상의 인기를 보유했던 프로 농구가 사람들의 관심에서 멀어진 것 모두 성적만 중요시하고 팬들을 연구하거나 새롭고 의미 있는 서비스를 지속적으로 창출하지 못해 만들어진 결과다.

프로 스포츠가 상업적으로 가장 성공한 미국으로 눈을 돌려보자. 미국 프로 야구의 가장 명문 구단으로 뽑히는 뉴욕 양키스의 데이터 마케팅팀을 만나 많은 대화를 나눈 적이 있다. 미국의 경우 팬을 관리하는 가장 중요한 방법 중 하나가 시즌 티켓(1개 시즌의 모든 경기를 관람할 수 있는 티켓 상품)을 판매하는 것이다. 인기 구단의 경우 80% 이상의 좌석이 모두 시즌 티켓으로 판매되고 한번 구매한 시즌 티켓은 대를 이어 물려주기도 한다. 그만큼 충성심이 높은 팬을 많이 보유하고 있다는 반증이기도 하다. 그러면 어떻게 이렇게 충성심이 높은 팬을 보유하는 것일까? 뉴욕 양키스는 팬들에 대한 철저한 데이터 분석이 핵심요소라고 설명했다.

한 예로 시즌 티켓 구매 고객의 좋아하는 와인 브랜드까지 알고 있을 정도다. 그래서 시즌티켓 구매 고객의 생일 이면 경기장의 좋은 자리로 업그레이드 서비스를 제공하고, 친구 또는 가족과 함께 경기장을 방문할 수 있도록 무료로 추가 티켓을 제공해주며, 초청된 팬이 좋아하는 와인을 자리에 세팅해 놓는다고 한다. 시즌 티켓 구매자 가 구단으로 전화를 걸어 문의할 때 그 고객을 담당하는 전담 직원에게 연결되고 통화 연결과 함께 고객을 관리 하는 화면에 핵심적인 고객정보가 동시에 제공된다. 그래서 전화에 응대하는 구단 직원은 팬에 대해 많이 알고 있는 상황에서 팬을 응대하기 때문에 팬 입장에서는 구단이 자신을 알아봐 주고 잘 관리해 준다는 경험을 하게 된다고 설명했다.

팬 데이터를 수집하고 분석하여 맞춤형 서비스를 제공하는 사례는 책 한권으로도 부족할 만큼 많이 있다. 사실 스포츠 마케팅 하면 많은 이들이 화려한 이벤트(올림픽 개회식이나 슈퍼볼 하프타임쇼 등)와 연계된 다양한 광고 정도를 떠올리는 경우가 많은데, 조금 더 깊이 들어가보면 고객 정보를 수집하기 위한 치열한 노력과 고민이 있고 체계적으로 분석된 결과를 통해 상품과 서비스로 이어지는 매우 정교한 프로세스가 있음을 발견하게 된다. 마케팅은 상품이나 서비스를 팔기 위한 이벤트가 아니라 고객을 알아가고 이를 통해 고객에게 기대 이상의 만족 을 주는 긴 여정(Journey)라는 것을 먼저 이해하는 것이 매우 중요하다.

팬을 이해하기 위한 노력은 바로 데이터 수집·축적·분석

현재 미국과 유럽 세계 최고 구단들의 데이터 마케팅 성공 사례를 한국에 적용하는 프로젝트를 하고 있다. 프로젝트 파트너는 한국프로배구연맹이고 모든 구단의 참여로 프로젝트가 진행되고 있다.

프로젝트는 가장 첫 단계로 구단들을 만나 현재까지 축적된 고객 정보를 확인하는 작업부터 시작했는데 사실상 고객과 관련된 정보가 전무한 상태였다. 그래서 비용만 사용하면서 효과가 없는 소모성 이벤트가 많이 있다는 것을 발견했다. 예를 들면 관중을 증가시키기 위해 최신 핸드폰을 경기 후 후첨하는 이벤트를 했는데, 이 핸드폰 을 받아간 고객은 원정팀을 응원하기 위해 방문한 분이였다. 핸드폰을 주는 이벤트는 이를 통해 우리 구단을 좋아하고 방문을 증가시키기 위한 수단인데 엉뚱한 원정팀 고객이 받아가는 일이 벌어진 것이다. 이와 같은 문제가 발생하는 것은 고객에 대한 데이터 관리가 전혀 되지 않은 결과의 한 사례다. 그래서 우리 팀은 팬 데이터를 수집하는 것부터 시작해야 했다. 먼저 데이터베이스는 티켓 판매 분석, 팬 성향 분석 그리고 선수 성과 평가 분석 3가지 분야로 설계하고, 데이터를 축적하기 위한 핵심 질문(Key Question)을 도출했다

<데이터 축적을 위한 핵심 질문>

1. 티켓 판매 데이터베이스(Ticket Sales Database) : Ticket을 구매하는 고객은 누구이며 얼마나 자주 방문하며, 어떤 자리를 선호하는가?

2. 팬 성향 데이터베이스(Fan Behavior Database) : 팬들은 어떤 생각을 가지고 있고, 티켓 구매를 결정하는 중요한 변수는 무엇인가?
3. 선수 성과 평가 데이터베이스(Player Evaluation Database) : 프로 스포츠의 가장 중요한 상품인 선수를 어떻게 평가하고 영향력을 측정할 것인가?

이와 같은 구조에서 3개 시즌을 운영한 결과 Ticket Sales 데이터는 약 400,000건(전 구단 전체 티켓 구매 결과 데이터 축적, 1개의 구매 데이터는 구매자 정보, 구매 시간, 구매 좌석 등 약 20가지 항목의 구매 내역 포함) 축적, 팬 성향 데이터는 약 200,000건(팬 참여 설문조사 데이터 중심, 200,000건은 티켓 구매 고객 2명 중 1명의 매우 상세한 성향을 알고 있는 것을 의미한다. 1건의 팬 설문조사 데이터는 평균 20개의 세부 항목 포함(번호 구단 및 선수, 좋아하는 음식, 경기장을 방문하는 이유, 방문하지 못하는 이유 등)) 축적 그리고 선수 성과 평가 데이터는 10개 시즌 동안 출전한 전체 선수의 경기적 성과(On-Court Performance)와 2021/22시즌 약 57,000건의 고객 참여 데이터로 분석한 선수별 마케팅 영향력 평가 데이터를 보유하게 되었다.

이와 같이 데이터를 축적하여 팬을 분석하고 새로운 상품을 개발한 결과 코로나 사태로 무관중 및 매우 제한된 좌석 판매 상황에서도 오히려 구단의 매출은 증가하였다. 실제 A구단의 경우 작년 시즌(2020/21시즌) 판매 좌석수는 재작년 시즌에 비해 약 5% 정도만 가능했는데 매출은 50%수준을 유지했고 올해(2021/22시즌)는 코로나 이전 시즌에 비해 약 30% 수준의 좌석 판매가 가능했지만 매출은 300%가 성장하는 성과를 창출했다.



<3개 시즌간 축적된 데이터를 분석한 사례 : 팬 연령대/성별/주소 분석>

이제부터는 우리가 데이터로 발견한 핵심 내용을 설명하려고 한다. 단지 프로젝트 특성상 모든 부분에서 구체적인 수치를 제공하지 못한다는 한계는 있다.

(1)팬들은 경기를 보는 것이 아닌 선수를 보기 위해 경기장을 방문한다

입장 가능 비율이 30% 수준인 경우 티켓은 당연히 매진될 것으로 예측했다. 이유는 코로나이전 시즌에도 전구단의 최소 관중 비율이 판매 가능 전체 좌석 대비 30%를 넘기고 있었기 때문이다. 하지만 매진이 되지 않는 경기가 지속적으로 발생하였고, 티켓 판매 데이터 분석 결과 코트와 먼 자리(2층 또는 3층 맨 위에 위치한 자리)가 판매되지 않는 것을 발견했다.

그래서 팬들이 경기장을 직접 방문하는 이유와 선호하는 좌석에 대한 분석을 하게 되었다. 데이터를 수집하여 분석한 결과 팬들은 “선수를 직접 그리고 가까이 보기 위해” 경기장에 방문하며 이 목적을 이루기 위해 “코트에서 가장 가까운 좌석”을 선호한다고 가장 많이 응답하였다. 또한 티켓을 구매하지 않는 이유는 “내가 원하는 좌석이 없어서”가 가장 높은 비율의 응답이 나왔다. 위 데이터를 해석하면 왜 입장 비율이 30% 상황에서 매진이 되지 않는지 설명이 된다. 팬들은 선수를 가까이 보기 위해서 코트에서 가까운 좌석을 구매하는데 만약 이 좌석이 구매가 가능하지 않다면 경기장을 방문하지 않는 결정 즉 티켓을 구매하지 않는 결정을 행동으로 옮기고 있었던 것이다.



<코트에서 멀어질 수록 빈자리가 많이 나타나고 있다>

이와 같은 데이터 분석을 결과로 코로나 상황에 부합한 새로운 상품을 설계했는데, 핵심 요소가 선예매권(일반 고객들보다 먼저 예매할 수 있는 권리를 부여, 선예매권을 구매한 고객은 자신이 원하는 좌석을 구매할 가능성이 높아지는 서비스)을 포함시키는 것이다. 이 선예매권을 따로 판매하는 것이 아니라 멤버십 상품을 만들어 포함시켰다. 구단에 소속되었다는 느낌을 주는 멤버십 상품에 선예매권을 포함한 것이다. 이 상품에는 많은 분들이 호응해 주었고, 대부분의 구단들이 5분 안에 준비했던 상품을 모두 매진 시켰으며 이 때 매출은 우리 팀이 프로젝트를 시작하기 전 시즌별 구단 전체 매출의 50%이상을 만들어 내기도 했다.

(2)팬들은 구단을 좋아하는 것이 아니라 선수를 좋아하는 것

경기장 방문 목적이 선수를 가까이에서 보는 것이 가장 중요한 이유인 것에서 확장하여 팬들의 행동 분석을 좀 더 심도 있게 하기 위해 구단에 대한 선호도와 선수에 대한 선호도를 별도로 분석하였다. 그 결과 인기가 높은 선수는 구단을 이적하는 경우 응원하는 구단도 바꾸겠다고 하는 팬의 규모가 높은 비율로 나타나는 것을 발견했다. 구단에 대한 충성도가 아닌 선수에 대한 충성도가 높은 구조로 시장 현황이 분석된 것이다.

필자는 유명 연예 기획사의 신사업 전략을 컨설팅한 경험이 있다. 그 때 발견한 것이 아티스트(연예인)는 좋아하지만 소속사는 싫어하는 팬들의 행동을 발견하였다. 그래서 연예 기획사가 아티스트의 본원적 활동 이외에 신사업을 진행할 때 대부분 실패하는 사례가 많다는 것을 고객 데이터 분석으로 찾아내어 향후 전략 방향을 제시하였다.

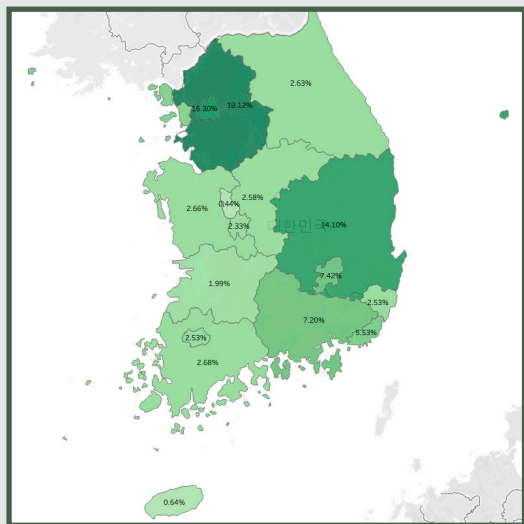
배구 팬 데이터 분석 결과 현재 한국 프로 배구 시장에서 팬들의 행동 방식은 연예인을 좋아하는 것과 동일한 패턴이 나오는 것을 알 수 있었다. 특히 10~20대 여성 팬들이 전체 시장에서 50% 이상을 차지하고 있는 것도 아이돌을 좋아하는, 소위 덕질 하는 문화가 프로 스포츠 시장으로 확대되고 있음을 파악하게 된 것이다. 특히 작년 여름 도쿄 올림픽에서 엄청난 감동을 주었던 여자배구 국가대표 선수들을 중심으로 많은 팬들이 몰리고 있다는 것도 발견하게 되었다.

그래서 우리 팀은 새로운 서비스와 상품을 구성하는 벤치마킹 대상을 미국, 유럽의 프로 스포츠 구단이 아닌 BTS와 같은 아이돌 그룹으로 변경하여 설정하였고, 현재 아이돌 그룹 상품(MD, 굿즈 등) 중 가장 활발하게 판매되는 상품이 무엇인지 분석하여 프로배구 시장에 접목하기 시작했다. 그 결과 위에서 설명했던 매출의 증가가 이루어진 것이다.

(3)암 체어 팬(Arm Chair Fan)을 관리해야 한다

우리가 데이터 분석을 통해 발견한 또 다른 중요한 요소는 티켓을 구매하여 경기장을 찾는 팬만이 팬이 아니라 경기장에 방문하지 않지만(또는 방문할 수 없지만) 여전히 관심을 가지고 지속적으로 구단의 이야기에 귀 기울이는 팬들도 체계적으로 관리해야 한다는 것이다. 프로 스포츠 시장에서는 이 팬들을 집에서 편히 앉아 경기를 관람하는 팬이라는 의미로 “암 체어 팬(Arm Chair Fan)”이라는 용어를 사용하기도 한다.

김천을 연고지로 사용하고 있는 한국도로공사 팬의 지리적 분포를 분석하면 김천이 있는 경북지역보다 많은 팬들이 서울/경기 지역에 분포하고 있는 것을 볼 수 있다. 하지만 서울/경기지역 팬들은 김천에서 열리는 도로공사 홈경기에 모두 참여하기란 사실 상 불가능하다.(평일 7시에 시작하는 홈경기를 보기 위해서는 직장인은 휴가를 내야한다). 하지만 거리적으로 떨어져 있는 팬들은 비록 경기장에는 자주 방문할 수 없지만 매우 높은 관심을 가지고 구단의 이야기를 경청하고 있으며, 언제든지 매력적인 상품이 나오는 경우 경기장 방문과 관계없이 구매할 준비가 되어 있다는 것을 알게 되었다.



<김천을 연고지로 한 한국도로공사 지역적 팬 분포 분석, 구단의 연고지는 경북 김천이지만 서울, 경기 지역의 팬 분포 수준이 더 높게 나타난다.>

그래서 프로배구의 타이틀 스폰서인 도드람 그리고 한국도로공사와 협력을 통해 도드람의 상품과 선수 포토카드(아이돌 그룹의 핵심 상품 중 하나)가 포함된 상품을 설계했다. 이 상품은 경기장 방문과 관계없이 돼지고기와 선수카드를 받을 수 있는 상품인데, 출시 후 역시 5분만에 모두 매진되었고, 너무 많은 구매자가 몰려들어 도드람 온라인 몰 사상 처음으로 서버가 마비되는 사건이 발생하기도 했다.(도드람에 따르면 서버가 마비되지 않았다면 1분만에 매진되었을 것이라고 한다.)

이 사례를 통해 스포츠 스폰서가 단순히 브랜드만 알리는 것이 아니라 실제 매출을 발생시킬 수 있다는 성공적인 의미를 알 수 있지만, 이 글의 주제인 팬 중심 운영으로 돌아오면, 팬들이 가지고 싶은 상품을 데이터 분석을 통해 설계하고 제공했을 때 매우 폭발적인 반응이 있다는 것을 증명하는 사례이기도 하다. 경기장에 직접 방문하는 것과 상관없이 말이다.



<프로배구 타이틀 스폰서인 도드람과 한국도로공사 배구단의 협력으로 만들어진 상품 인스타그램 홍보 사진>



스포츠 마케팅 분야에서 데이터 분석이 성공하기 위한 6가지 방법

팬 데이터 분석은 스포츠 마케팅에서 심장과 같은 역할을 한다. 꼭 스포츠 마케팅뿐만 아니라 모든 마케팅 영역에서 동일한 중요성이 있다고 생각한다. 데이터 분석을 통한 마케팅은 한 때의 유행이나 캐치프레이즈가 아니라 조직 운영의 핵심 DNA가 되어야 한다고 생각한다.

다양한 프로젝트를 하면서 느낀 데이터를 잘 활용하기 위한 방법을 소개한다.

첫째, 평균에서는 안 보이는 고객이 있다. 하지만 시장을 매우 세분화하면 이 작은 그룹들이 보이기 시작한다.

세분화로 밝혀진 작은 그룹은 개별화된 맞춤형 서비스를 제공해야 한다. 좀 특별한 요구사항들이 이 그룹 내에는 존재하기 때문이다. 하지만 맞춤형 서비스는 비용이 좀 더 발생한다. 그래서 단기적으로 볼 것이 아니라 고객 생애 가치(Life Time Value)를 계산하여 접근해야 한다.

하지만 단기적으로 발생하는 많은 비용만 보이지 고객이 시간을 두고 만들어내는 좀 긴 시간에서의 가치는 잘 보이기 않기 때문에 비용 사용을 주저하는 경우를 자주 보게 된다.

고객은 인정받기 원하고 자신을 알아주기를 원한다. 그래서 고객에게 시간도 돈도 투자할 줄 알아야 한다. “의미 있는 일에는 돈을 낙엽처럼 태울 줄 아는 용기”가 있는 리더가 마케팅 프로젝트를 리드해야 한다. 그렇지 않으면 장사치에 머물게 된다.

다시 말하면 고객은 자신을 알아주는 곳으로 이동한다. 고객을 알아준다는 것은 꼭 이름을 외우고 인사하고 납작 엎드려 ‘주인님 말씀하세요.’하는 방식을 의미하는 것이 아니다. 고객은 자신이 원하는 것 이상, 생각하는 것 이상의 상품, 서비스, 솔루션 등을 제공받았을 때 인정받았다고 느끼게 된다. 고객은 이런 인정을 받는 곳을 찾아 이동한다. 이것을 “발로 하는 투표”라고 표현하기도 한다.

고객에게 투자하는 용기가 있어야 한다. 그러면 고객은 우리에게 더 가치 있는 데이터를 기꺼이 제공해 준다.

둘째, 데이터는 방향을 알려주는 것이지 고객과 관계를 형성하는 것은 사람의 노력이 필요하다.

따뜻한 마음으로 고객과 함께하는 것은 좋은 인격을 품은 사람만이 할 수 있다. 그런데 이 좋은 인격을 품은 사람도 비용이 들어가야 경쟁력과 지속 가능성이 생긴다. 근데 “인건비”라는 이름으로 이것을 제한하고 있다. 그래서

데이터 분석에서 아무리 인사이트 있는 결론을 도출해도 경쟁력은 잘 생기지 않는 경우가 많다. 정작 실행해야 할 사람에 투자하지 않기 때문이다.

고객과 좋은 관계가 형성되어야 새로운 제안을 할 수 있다. 관계가 형성되지 않았으면 아무리 좋은 제안도 스팸에 일일 뿐이다. 관계가 잘 형성된 고객은 더 품질 높은 데이터를 제공할 뿐 아니라 가장 뛰어난 마케터로 변신한다. “추천”이라는 방법을 사용해서 말이다.

셋째, 데이터는 결정하라고 있는 것이지 분석된 그래프를 만들라고 있는 것이 아니다.

특히 위험한 결정을 도와주기 위해 데이터를 쓰는 것이다. 그런데 위험한 일에는 결정을 하지 않고 시간을 흘려 보내고만 있는 경우가 많다. 사실 결정을 하지 않기 보다는 그냥 예전의 가장 안전한 방법을 그대로 사용하는 결정을 하는 경우로 봐야 할 것 같다. 하지만 그 과거의 안전한 방법이 사실은 침묵의 암살자(Silent Killer)로 서서히 시장에서 자신이 속한 조직을 질식시키고 있는지는 모르는 채로 시간만 낭비하고 있는 것이다.

결정을 안 하면 속도가 느려진다. 경쟁자들이 변하면서 경쟁력을 갖추어 나갈 때 속도가 느리다는 것은 고객들이 떠나가는 이유를 제공하는 것이다. 빠르게 고객의 문제를 해결하는 방법을 제시하는 곳에서 고객은 인정받았다고 느끼게 된다. 속도가 느리다는 것은 문제해결 능력 방법 제시가 느린 것이고 이는 “내 문제를 풀 수 있는 실력이 없다.”는 고객의 평가를 받게 될 가능성이 높다. 그러면 고객은 떠나 간다. 반면 자신의 문제를 풀어주는 경험을 한 고객은 한 단계 높은 미션을 제시하며 또 다른 데이터를 축적할 방향을 제시해 준다.

위험한 결정을 데이터를 활용해 위험을 최소화하며 고객에게 혁신적이며 새로운 해결책을 제시하는 것을 주저해서는 안 된다.

넷째, 분석은 유연해야 한다. 그래서 데이터를 정형화된 시스템 속에 집어넣고 자동화 시키려 하면 돈과 시간만 낭비하게 된다.

보고서 형식을 너무 이쁘게 만드는 것과 자동화라는 환상에 사로 잡혀서 IT시스템 중심으로 움직이면 안 된다. 시장이 변하면서 질문도 변하고 그러면 답을 만들기 위한 방법도 변해야 하며 보고를 하는 방식도 변하게 되는데 고정된 방식의 자동화는 이 결과를 어렵게 만들기 때문이다.

시장의 변화와 관련해서 질문하는 것이 실력이지만 이미 질문 된 것에 정답을 만드는 것이 실력은 아니라고 본다.

정형화된 질문에 대한 답은 이제 AI가 하면 한다.

아직도 CRM을 IT로만 접근하는 오류에 빠진 실무진과 이 말을 믿는 사람들이 리더 자리에 있는 것을 종종 보는데 시간이 지나면 이 잘못된 결정에 대한 책임을 져야하는 어려움을 겪어야 할 것이다. 눈에 보이는 번쩍이는 대시보드에 현혹돼서 시스템을 유연성 없이 만들면 1년 지나면 버리고 싶은 욕망이 생기고 3년안에 대부분 버릴 수밖에 없기 때문이다.

좋은 질문이 좋은 데이터를 모은다.

다섯째, 실패해야 전진이 있다.

아무도 가지 않은 길을 지름길로 갈 수는 없다. 공식이 없기 때문에 실험해야 하고 실패해야 하는 것이다. 가다 돌아오고 이 산이 아니라고 말해야 하고 그 과정에 오해 받고 욕먹고 하는 것이 새로운 도전을 하는 사람들의宿命이다.

좀 미친 것 같아 보이는 아이디어에 “한 번 해봐”라고 용기를 주는 리더가 필요하다. 그리고 실패하면 “다음에 더 미친 짓 해봐”해야 한다. 그래야 변할 수 있다. 이런 미친 도전이 새로운 데이터를 모으는 원동력이 된다.

남들이 성공한 안전한 길로 가면서 데이터를 모으는 것은 중복 투자일 뿐이다. 새로운 데이터가 새로운 경쟁력을 보유하도록 도와준다.

마지막으로 데이터 분석은 지위고하를 막론하고 직접해야 한다.

요약된 결과만 보면 자신의 경험 범위에서 결론을 도출하기 때문이다. 데이터가 들어가 있는 셀 하나 하나 값을 수정해 가면서, 필드 값 하나 하나 수정해 가면서 숨어 있는 데이터를 느끼고 경험해야 새로운 질문과 도전을 할 수 있다.

그래서 나이가 들어도 데이터 다루는 기술을 익혀야 한다고 본다. 짜증나도록 반복적인 데이터 수정 작업까지도 직접 하는 것이 필요하다. 그래야 분석을 위한 질문과 새로운 데이터 수집의 니즈(Needs)가 생기기 때문이다.

자신의 경험 범위에 있는 질문으로는 새로움을 만들 수 없다. 직접 현장에서 부딪치고 문제를 눈으로 봐야 새로운 질문이 생기고 그 새로운 질문이 새로운 데이터를 축적할 수 있는 기회를 부여한다.

만약 이 과정을 게을리하게 될 때 정보는 많지만 인사이트는 부족한(Too much information, Too little insight) 상황이 발생되고 데이터 기반의 변화와 발전 프로젝트가 비용대비 효과가 없는 결과로 귀결될 가능성이 높다.



통계를 활용한 지능화된 보험사기의 대응

김주환 | SAS Korea 이사



보험사기 적발 인원 18년 대비 20년 25% 증가

보험사에 보험계약자(또는 가입자)가 보험금을 부당하게 청구하여 보험금을 받으려 보험사를 속이는 행위를 보험사기라고 한다. 예를 들면 부당한 목적으로 특정 기간 내 너무 많은 청구를 하거나 특정 질병으로 과다하게 청구하는 경우 등의 유형이다. 보험사에서는 매년 꾸준히 증가하고 있는 보험사기에 대응하기 위해 데이터 기반으로 보험사기 유형을 분석하고 이를 시스템으로 구축하고 있으며 이 시스템을 보험사기방지시스템(IFDS, Insurance Fraud Detection System)이라고 한다. 대부분의 보험사는 금융감독원의 권고에 따라 보험사기방지시스템을 갖추고 있으며 이를 활용하여 보험사기에 체계적인 대응을 하고 있다. 하지만 보험사의 대응에도 불구하고 보험사기는 매년 지속적으로 증가하는 추세이다. 금융감독원의 통계를 보면 보험사기 적발 금액은 2018년 7,982억원에서 2020년 8,986억원으로 13% 증가, 적발인원은 2018년 79,179명에서 2020년 98,826명으로 25% 증가하였고 특히 적발금액 증가율에 비해 적발인원 증가율이 더 높게 나타나고 있다.(2020, 금융감독원 보험사기적발통계) 이는 보험사기에 가담하는 인원이 더 증가하고 있고 그 유형도 복잡화되어 가고 있음을 예상해 볼 수 있다.

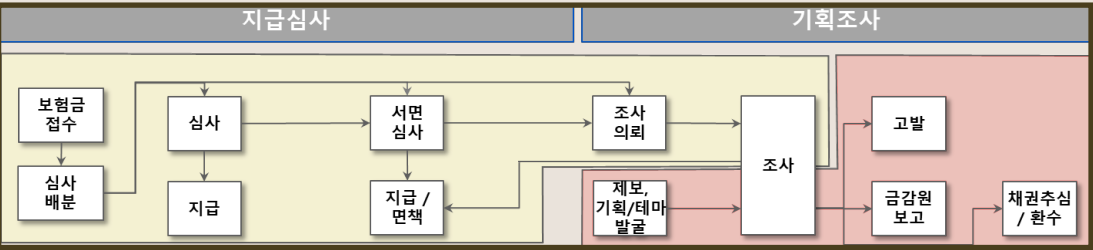


보험사기 적발통계 (2020, 금융감독원)

주) 보험 종류는 생명보험(생보; 일반보장성 보험), 손해보험(손보; 자동차보험, 장기보험)로 구분됨

보험 심사 업무 프로세스(지급심사, 기획조사 업무)

보험 심사에서의 통계 활용을 설명하기 전에 보험 지급 심사 프로세스를 우선 소개하고자 한다. 보험금을 청구하면 보험금을 지급하기 위한 보험금 지급 심사 단계, 보험금 지급 후에는 상세 분석을 통한 기획조사 단계가 진행된다. 지급심사 단계는 보험가입자가 보험사에 보험금을 접수매체(모바일, 인터넷, 설계사, 고객센터 등)를 통해 청구하면 보험청구는 접수되며 청구심사하여 보험금이 지급되고 접수건 중 의심이 되는 건은 전문 심사로 이관되어 정밀 심사 및 조사를 통해 보험금 지급 여부가 결정된다. 기획조사는 보험금 지급이 후 단계로 제보접수 또는 지급된 데이터를 분석하여 사기 유형을 분류하고 수사 의뢰 및 조사를 통해 고발 및 추심 단계로 진행된다. 참고로 기획조사를 위한 유형 분석 시 장기적인 관점의 접근이 필요하여 데이터는 최대 5~7년 지급데이터를 활용하여 분석한다.



보험 심사 업무 프로세스 (SAS 업무 자료)

지급심사 단계에서의 통계 모델 적용 및 활용

보험 지급 심사 단계에서 통계적 방법은 보험금이 청구 접수되면 이후 심사배분 시점에 적용되어 보험금 지급의 위험도 정보를 제공하여 위험도가 일정 수준 이하인 경우는 보험금 지급 절차를 통해 지급되며 위험도가 일정 수준 이상인 경우 전문 심사 단계로 이관되어 정밀 심사 및 조사 단계를 진행하도록 관련 정보를 제공한다. 위험도 정보 제공을 위한 통계적 방법은 모델링 방법이 주로 활용되는데 이때 적용되는 모델을 보험사기방지모델 (Insurance Fraud Detection Model)이라고 한다.

보험사기방지모델은 모델링 대상(Target) 정의, 모델링 기법 정의, 후보 Factor 구성, 모델링 및 최종 모델 구성의 단계로 개발된다.



보험사기방지모델 개발 단계 (SAS 업무 자료)

[모델링 대상(Target) 정의]

보험사기방지 모델링을 위해 모델 대상을 정의하며 이를 Target이라고 한다. Target은 보험사 업무 담당자와 협의하여 대상을 확정하며 주로 보험사기여부(Y/N, 0/1)로 정의하여 모델을 개발한다. Target 정의 후에는 과거 사기 적발 데이터를 수집하여 보험사기유형의 발생 건수를 산출하여 모델의 단위를 정의한다. 사기 건수가 충분히 확보되는 정도를 고려하여 단위를 결정하며 이에 따라 개발할 모델의 개수가 정의된다. 예를 들면, 수술모델/장해모델/위장사고모델 등의 단위를 의미한다.

[모델링 기법 정의]

모델 개발을 위한 기법은 보험사의 요구사항 및 데이터 현황을 고려하여 기법을 선정 및 설계한다. 모델링 기법은 주로 로지스틱회귀분석(Logistics Regression), 의사결정나무(Decision Tree)의 기법이 활용되고 있으며, 최근에는 보험사기 분류 정확도를 높이기 위해 기계학습(Machine Learning) 기법을 검토 및 적용하고 있다. 참고로 통계적 모델링을 통해 보험사기여부를 분류하는 모델을 모델룰(Model Rule)이라고 표현한다.

[후보 Factor 구성]

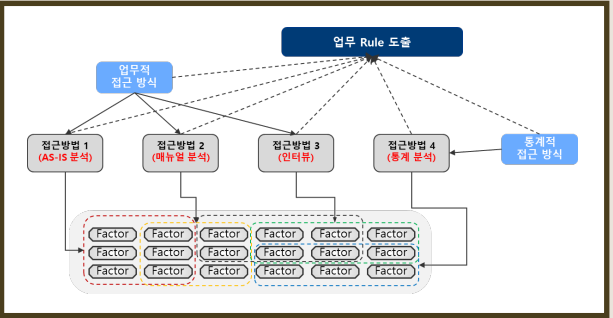
모델링을 위한 후보 변수는 업무지식의 정보화, 통계적 분석의 방법을 통해 구성한다. 업무지식의 정보화는 보험심사 업무 및 조사 부서 업무 담당자들 대상으로 인터뷰하여 보험사기를 판단하는 다양한 업무 지식을 수집하고 이를 로직으로 구성하여 정보화하며 통계적 분석 기법을 적용하여 보험사기 발생패턴 분석 및 정보화하기도 한다. 가능한 정보를 최대한 구성하여 모델링 시 적용하며 초기 후보는 상황에 따라 다를 수 있으나 2,000~3,000여개 규모까지도 생성하기도 한다.

[최종 모델 구성]

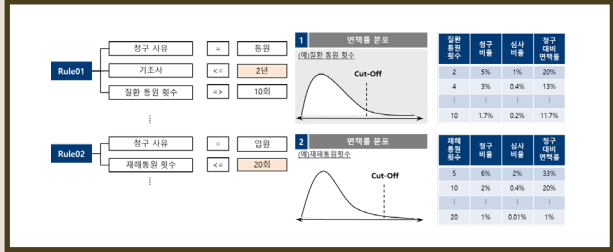
모델링 대상(Target, 보험사기여부), 모델링 기법 정의, 후보 Factor 구성 이후 모델링을 적용하여 유의한 Factor를 선정하고 모델의 성능이 가장 높은 모델을 최종 모델로 선정한다. 선정된 모델을 적용하여 보험사기 발생 위험도를 확률값인 Score로 산출하고 이를 지급심사단계에 적용한다.

업무지식을 반영한 보험사기 예측 보강

보험사기여부 예측 시 통계 모델 만으로는 한계가 있으며 이를 보완하기 위해 심사 및 조사업무 담당자의 보험사기 적발 경험을 로직으로 구성하여 보험사기예측의 보강에 활용한다. 참고로 보험사기 적발 경험을 기반으로 보험사기여부를 분류하는 방법을 업무룰(Business Rule)이라 표현한다. 업무룰(Business Rule)은 업무 담당자의 보험사기 적발 경험을 로직화하고 이를 모델 대상(Target)과 비교하여 업무룰별 분류기준치(임계치)를 정의하여 구성한다.



업무룰(Business Rule) 도출 방법 (SAS 업무 자료)



업무룰의 분류 기준치 정의 (SAS 업무 자료)

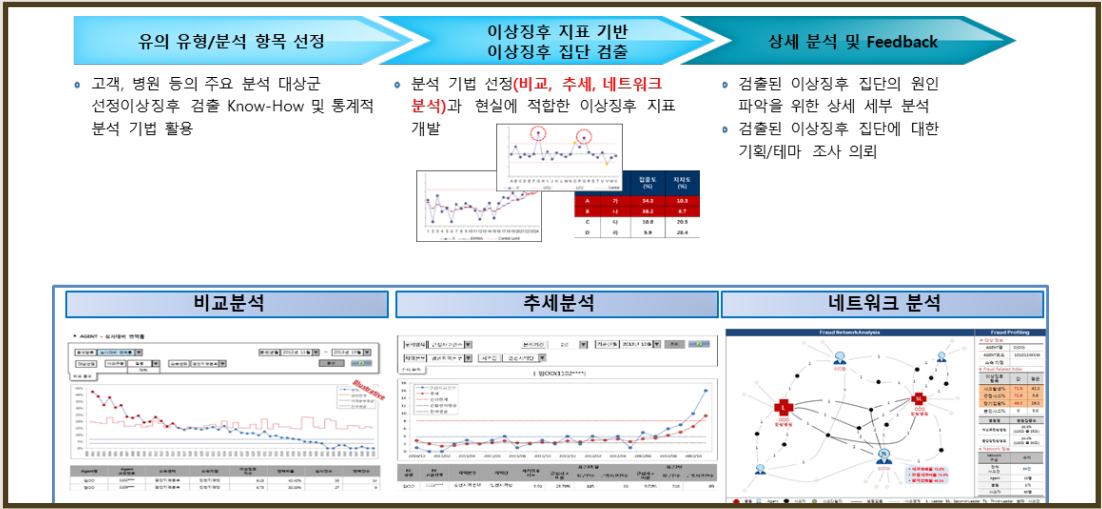
업무룰별 분류 기준치 정의 시 청구비용, 심사비용, 면책비용 등 주요한 보험 정보를 활용한다.

심사단계에서의 보험사기 분류를 위한 모델은 최종적으로 통계모델을 기반한 통계룰(Model Rule)과 업무에 기반한 업무룰(Business Rule)을 혼합하여 보험사기 위험도인 Score를 산출하여 정보로 제공한다.

기획조사 단계에서의 통계적 방법 적용

보험금 지급이 후 장기적인 지급 유형을 분석하여 보험사기를 유형화하고 이를 기반으로 조사 및 수사하여 적발을 통해 보험금 환수하는 단계를 진행한다. 단계는 제보접수 또는 보험사기 패턴을 분석하여 유형화하고 이를 기반으로 조사하는 단계로 진행하며 그 중 보험사기 유형화는 장기간 데이터를 활용하여 유형화하며 분석 시 보험사기관련자도 고려한다. 관련자는 생명보험인 경우 보험가입자, 설계사, 병원, 지점 등이고 손해보험인 경우 가입자, 설계사, 정비업체 등을 의미한다. 보험사기유형화 분석 방법은 병원, 설계사, 정비업체 등 특정 그룹별 분포를 비교 분석하여 유의 그룹을 도출하거나 병원-설계사-가입자, 정비업체-설계사-가입자별 청구유형 및 금액 비교 등 연관된 그룹별 관계를 고려하여 유형을 도출하고 이를 지표화 한다. 분석방법은 주로 추이분석, 연관성분석, 관리도, 의사결정나무 등을 활용한다.

하지만 다양한 보험사기 유형화, 적극적인 보험사의 대응에도 불구하고 보험사기 유형은 지능화 되고 있으며 보험사기 유형도 복잡한 형태로 확산되고 있다. 이에 복잡한 보험사기 관계를 모니터링하고 보험사기 의심 집단을 도출 및 대응할 수 있는 분석 방법의 필요성이 대두되어 최근 네트워크분석(Fraud Network Analysis)과 같은 방법의 접근이 시도되고 있다. 실제 최근 보험사에 적용되어 향후 다양한 보험사에 확산 적용될 것으로 보이며, 이와 같이 보험사에서는 보험사기 지능화에 대응하기 위해 많은 연구와 데이터적인 접근을 시도하고 있다.



보험사기 유형 분석 (SAS 업무 자료)

네트워크분석에 기반한 지능화된 보험사기 대응

최근 보험사기는 그 유형이 지능화되고 가담하는 인원도 증가하여 대체로 2인 이상이 공모하는 형태로 복잡한 구조를 보이고 있으며 보험사기 적발을 위해서는 복잡한 구조에서 패턴을 분석하고 시각화할 수 있는 접근이 필요하게 되었다. 이와 같은 필요성에 따라 네트워크분석을 통한 접근이 고려되었고 최근 실제 보험사에 적용되어 효과가 나타나고 있다. 상세히 네트워크분석에 기반한 보험사기 분석의 단계에 대해 소개하고자 한다.

네트워크는 하나 이상의 관계 유형에 의해 연결된 구성원의 집합을 의미하며, 상호작용하는 개체 간 규칙적인 관계 패턴을 분석하는 것을 네트워크분석이라고 한다. 보험사기분석에 적용하면 네트워크는 피보험자, 설계사, 병원, 정비업체, 렌트업체 등 다양한 개체와 치료, 사고, 계약 등 다양한 관계 유형으로 구성되어 있으며 이와 같이 규칙을 분석하는 것이 네트워크분석이다. 네트워크 분석 단계는 분석 테마 및 시나리오 정의, 분석을 위한 관계

네트워크 테마 정의	보험사기 유형 중 네트워크 분석을 통해 적발 가능한 시나리오 정의
분석데이터 구성	각 테마 별 네트워크 분석을 위한 데이터 생성(관계 정의)
네트워크 분석	연관이 높은 개체들로 구성된 혐의집단 및 집중도 도출
혐의집단/개체 프로파일링	집단 및 개체의 혐의도를 알 수 있는 지표 산출
네트워크 시각화	네트워크 정보 시각화를 통한 조사 지원

네트워크 분석 단계 (SAS 업무 자료)

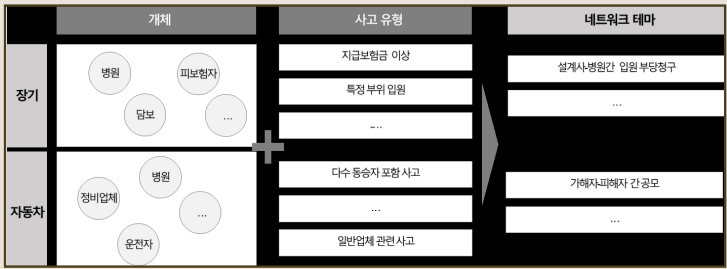
데이터 구성, 분석을 통한 혐의 집단 도출, 프로파일 분석 및 지표 산출, 분석 결과의 시각화단계로 진행된다.

[네트워크 테마 정의]

네트워크분석은 분석을 위한 테마를 정의하고 테마에 대해 개체 간 관계 유형을 분석한다. 네트워크 테마는 위험 판단에 사용되는 혐의지표를 도출하고 보험군별 개체와 사고 유형을 고려하여 네트워크 사기 시나리오를 정의하여 네트워크 테마를 도출하는 단계로 진행된다. 예를 들면 장기보험은 설계사-병원간 부당 공모 청구, 자동차보험인 경우 가해자-피해자 부당 공모 청구 등과 같이 테마를 도출할 수 있다.

[분석데이터 구성]

네트워크 테마 정의 후 분석을 위해 관계 데이터를 구성한다. 관계 데이터는 청구건을 기반으로 From-To 관계를 구성하는 것으로 건별 설계사→사고자, 사고자→병원, 가해자→피해자와 같이 관계를 구성하고 건수를 고려하여 가중치를 구성하는 형태로 데이터를 구성한다. 구성된 관계 데이터가 네트워크 분석 시 적용된다.



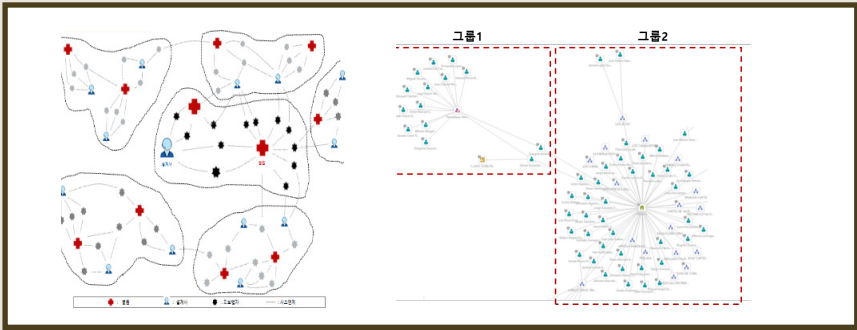
네트워크 테마 정의 (SAS 업무 자료)

No.	From	To	Weight
1	설계사1	사고자1	3
2	사고자1	병원1	2
3	사고자1	병원2	1
4	설계사2	사고자3	4
5	사고자3	병원2	4
6	설계사3	사고자4	3
7	사고자4	병원1	2

관계 데이터 구성 형태 (SAS 업무 자료)

[네트워크 분석]

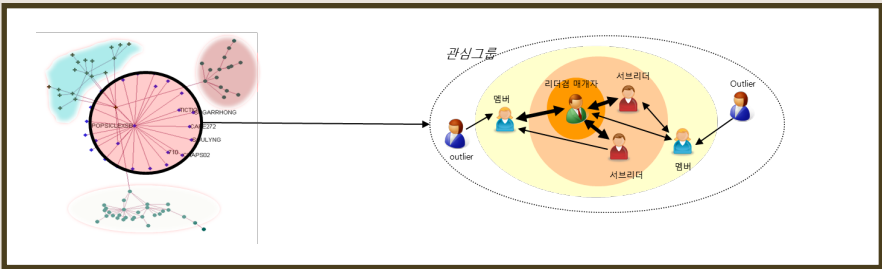
관계데이터를 기반으로 네트워크분석을 통해 보험사기 혐의집단을 도출한다. 1차로 개체간 네트워크를 구성하고 2차로 집중도 및 위험도 기반으로 혐의그룹을 구성하여 3차로 혐의그룹별로 상세 구조화한다. (혐의그룹 내에서는 관계의 밀집도는 높고 혐의그룹 간의 차이는 큰 형태)



네트워크분석 및 혐의그룹 도출 (SAS 업무 자료)

[혐의그룹 프로파일 분석]

도출된 혐의그룹에 대해서는 프로파일 분석한다. 혐의 그룹 내 해당 관계의 영향도를 고려하여 지표(중심성, 연관도 등)를 측정하여 역할(리더, 서브리더, 멤버, Outlier)을 설정하고 상세 프로파일을 분석하고 혐의도를 측정할 수 있는 지표를 산출한다..



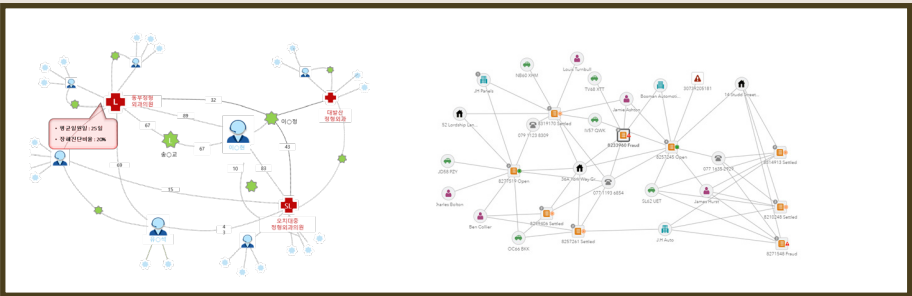
네트워크분석 및 혐의그룹 도출 (SAS 업무 자료)

리더, 서브리더, 멤버는 다음과 같은 기준으로 역할을 구분하며 역할의 특성을 고려하여 공모형태를 분석한다.

- 리더 : 커뮤니티 내 구성원과 직접 접촉 높은 경우(20% 이상), 양방향 소통 비율이 높은 경우 직접 접촉수가 최대인 경우
- 서브리더 : 리더와 유사한 성향이나 리더보다는 영향력이 약한 경우
- 멤버 : 양방향 의사 소통은 존재하나 리더군에 속하지 않는 경우
- Outlier : 의사소통이 단방향으로 이루어진 경우

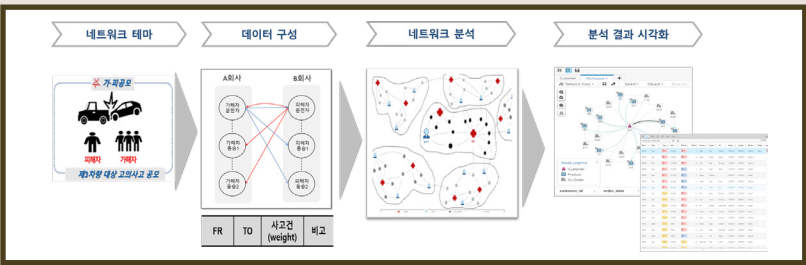
[네트워크 시각화]

네트워크 분석을 통해 혐의그룹을 도출하고 상세 프로파일을 분석한 후 보험사기 조사에 활용하기 위해 네트워크 정보를 시각화한다. 시각화 시에는 네트워크 정보와 상세 참고정보(청구건수, 금액, 위험도 등)를 제공하여 조사 시 활용하도록 제공한다.



네트워크분석 시각화 (SAS 업무 자료)

보험사에 적용하면 네트워크 테마 중 자동차보험의 가해자-피해자 공모인 경우 분석을 위해 가해자, 피해자, 동승자의 청구 정보를 활용하여 관계데이터를 구성한다. 이를 기반으로 네트워크분석 하고 위험지표(피해자수, 보험금지급액 등)를 활용하여 혐의그룹을 도출하며 이를 가해자-피해자 공모군 도출에 적용하는 단계로 진행한다.



네트워크분석 사례(가피공모) (SAS 업무 자료)

네트워크분석 도입으로 빠르게 복잡한 공모 관계 분석 및 대응 가능

D사에서는 네트워크분석을 도입하여 보험청구와 연루된 병원-환자 간 거미줄처럼 복잡하게 얽힌 공모형 사기를 빠르게 파악하고 과잉청구 및 조작사고에 대응 및 예방할 수 있는 기반을 마련하였다. D사는 네트워크분석 및 관련 시각화를 구성하여 현장의 공모에 대한 제보가 있는 경우 시스템에서 빠르게 조회하여 제보된 대상자의 네트워크관계를 분석하여 공모된 대상자를 조사할 수 있는 체계를 구성하였다. 기존은 제보 후 관계 분석 시 5일 이상 소요되었으나 현재는 2분 이내로 관계를 분석하여 대응할 수 있는 개선 효과가 있었다. 또한 복잡한 사고 처리 과정도 보험사기를 부추기는 요소 중 하나인데 해당 처리과정도 분석 가능한 환경까지 구성할 수 있었다. 예를 들면 자동차보험인 경우 정비업체와 부품사, 레커 등의 관계를 파악하기가 쉽지 않았으나 네트워크 분석 기반시스템에서 모 정비공장의 관계도를 검색한 결과 부품사, 모 렌터카 업체 등과의 연관성이 높게 나타났고 부품은 가까운 곳에서 가져오는 것이 일반적인데 A지역 모 정비공장에는 B지역 모 레커 업체와 실적이 월등히 많은 점이 나타나 조사할 수 있었다. 이와같이 D사는 네트워크분석 구축으로 빠른 공모 관계 분석이 가능하였고 복잡한 보험사기 유형도 검출할 수 있는 환경이 구축되는 효과가 있었고 이를 지속적으로 대응할 수 있는 효과를 기대할 수 있게 되었다.

핵심 기능	활용 데이터
AI가 빅데이터를 분석해 의심 거래 탐지	실손 차보험 청구건 중 의심 사례
사회적 관계 분석(SNA)	보험거래처 고객들 중 특이 사례
5일 걸리던 데이터 분석을 2분 만에 해결	특정 의료기관 과다 청구 데이터 등

보험사 네트워크분석 효과 (매일경제신문 기사, 2022.02)

통계적 방법을 도입을 통해 보험사기에 효과적 대응 가능

보험사기는 그 유형이 복잡하고 다양한 형태를 보이고 있고 매년 증가하는 추세이다. 이러한 보험사기에 효과적으로 대응하기 위해서는 통계적 방법을 통한 과학적 접근이 필요하다. 지급심사 단계에서는 보험사기를 감지하기 위한 정확도를 높이기 위해 Machine Learning 등 다양한 기법을 도입하여 모델을 개선하고 있고 기획조사 단계에서는 제보 및 추세, 분포 분석 외에 복잡한 공모에 대응하기 위해 네트워크분석의 방법을 도입하여 검출 및 대응 효과를 보이고 있다. 이와같이 보험사기 에 다양한 통계적 방법의 연구 및 적용으로 효과적인 대응이 가능할 것으로 기대할 수 있다.

통계로 바라보는 세상이야기

신동헌 | 도서출판 지일박스 대표

통계로 본 2021년 한국의 변화, 한국의 사회지표

통계청은 매년 초, 우리나라의 사회상을 종합적으로 살펴보고, 국민 삶과 관련된 전반적인 경제, 사회 변화를 쉽게 파악하기 위해 한국의 사회지표를 발간하고 있습니다. 올해도 2021년의 사회지표를 정리한 2021 한국의 사회지표가 3월 24일에 공개되었어요. 2020년 우리나라의 총 인구는 5,184만 명이었고, 이후 2021년에는 5,175만 명으로 약 10만 명이 감소했죠. 여성이 가임 기간(15~49세) 동안 낳을 것으로 예상되는 평균 출생아수를 뜻하는 ‘합계 출산율’은 작년에도 하향세였는데, 1996년부터 계속해서 하향세를 보이면서 2021년 0.81명으로 역대 최저치를 경신했습니다. 65세 이상 고령 인구도 857만 명으로 전년보다 약 42만 명 증가했고, 이는 전체 인구의 16.6%라고 해요.

한류문화 콘텐츠, 세계인의 마음을 사로잡다!

영화 <미나리>로 지난해 아카데미 여우 조연상을 수상한 윤여정 배우. 얼마 전 미국 3대 음악 시상식인 ‘그래미 어워드’에서 단독 퍼포머로 무대에 올라 기립박수를 받은 BTS. 그래미 어워드 사회자가 오징어 게임의 대사 '무궁화 꽃이 피었습니다'를 한국어로 선보이는 등 한국 문화의 세계적 인기를 다시 한 번 실감했습니다. 한국국제교류재단의 '2021년 지구촌 한류 현황'에 따르면, 2012년에 약 926만 명이었던 해외 한류 동호회 회원 수는 2021년에 약 1억 5,661만 명으로 무려 17배나 증가한 숫자입니다. 또한, 한국은행이 발표한 '2021년 지식 재산권 무역수지 보도자료(잠정)'에 따르면 문화 예술저작권의 수치가 2020년 흑자로 돌아선 이후, 2021년에는 7.5억 달러를 기록했습니다.

어려운 데이터 NO, 통계는 즐거워! KOSIS 통계 놀이터

2022년 2월, 통계청이 누구나 쉽고 편리하게 활용할 수 있는 ‘KOSIS 통계 놀이터’, 어린이가 자연스럽게 데이터를 활용하고 해석할 수 있도록 만든 어린이용 통계 서비스입니다. ‘KOSIS 통계 놀이터’는 통계를 활용하고 해석하는 능력 즉, 통계 리터러시 능력을 키울 수 있습니다. 통계 놀이터의 3가지 핵심 서비스는 비주얼 통계, 데이터로 말하기, 프로젝트형 통계 학습이 있는데요, 특히 ‘비주얼 통계’의 경우 교과서에 수록된 다양한 주제의 통계 자료를 인포그래픽 차트, 간단한 표로 한눈에 보기 쉽게 작성되어 더 편리하게 해석할 수 있습니다. 그 밖에도 통계에 대한 흥미와 관심을 끌 수 있는 ‘통계 웹툰’ 및 ‘카드 뉴스’도 있고, 낱말 퍼즐, 통계 퀴즈와 같은 각종 이벤트도 있습니다.



요즘 인기 있는 술은? 돌아온 전통주의 열풍

최근 힙합 아티스트 박재범이 지역 양조장과 협업해 만든 전통주인 증류식 소주가 오픈런이 벌어질 정도로 많은 사람들의 인기를 끌며 완판 행렬을 이어가고 있습니다. 국내 주류시장 규모는 감소했으나, 전통주 시장은 오히려 상승세입니다. 농림축산식품부·한국농수산식품유통공사, <2021년 주류 시장 트렌드 보고서>에 따르면 전체 주류 시장 출고 금액은 19년도 8조 9,412억 원에서 21년도 8조 7,995억 원으로 1.6% 감소한 반면, 전통주는 531억 원에서 626억 원으로 17.9% 증가했습니다. 출고금액 기준으로 탁주가 28.9%로 가장 큰 비율을 차지하였고, 과실주(23.1%), 약주(14.8%)가 그 뒤를 이었으며, 전통주 음용 상대는 배우자와 함께 마시는 비율이 27.9%로 나타났습니다.

인간은 어디서 삶의 행복을 느낄까요?

통계청에서 발표한 2020년 기준, 우리 국민들의 환경 만족도에서 대기 질은 38.2%, 수질은 37.7%, 토양 환경은 36.7%, 소음은 35.7%, 녹지환경은 58.7%로 대부분 50% 이상의 만족도를 넘지 못했습니다. 도시민의 삶의 질을 결정하는 중요한 요소는 도시공원으로 2020년 1인당 도시공원 면적은 11.0㎡로 2019년 10.5㎡에 비해 0.5㎡ 증가했고, 실면적은 525㎢로 2000년 214㎢에 비하면 2배가량 증가한 수치라고 합니다. 반면에 지역 생활환경의 질적 수준을 평가하는 대표적인 지표는 상수도 보급률인데, 2019년 농어촌 지역의 상수도 보급률은 78.6%로 2001년 29%에 비하면 큰 폭으로 증가하였지만, 도시 지역의 상수도 보급률인 99%에는 못 미치는 수준이라고 합니다.

한국 사회에 장애인에 대한 차별이 있다?없다?

지난 4월 20일은 장애인의 날이었습니다. 서울 지하철역에서는 출근 시간에 전국장애인차별철폐연대 회원들의 기습 시위가 벌어져 시민들의 관심을 끌었습니다. 우리 사회는 장애인에 대한 차별이 있을까요? 한국법제연구원의 국민법의식실태조사에 따르면 “장애인에 대한 차별이 있다”의 질문에 전혀 그렇지 않다는 응답이 1.4%, 별로 그렇지 않다가 12.7%, 보통이다 26.6%, 대체로 그렇다 46.7%, 그리고 매우 그렇다는 대답이 12.6%로 나타났습니다. 보통 이상의 응답이 85.9%로 국민의 대다수가 장애인에 대한 차별이 있다고 느끼는 것으로 조사되었습니다. 2007년에 장애인차별금지 및 권리구제 등에 관한 법률이 제정되었는데, 여전히 장애인들은 근로기준법의 사각지대에 있다고 합니다.

산림이 살아야 나라가 산다, 한국 산림의 변화

한국에서 산림은 전 국토의 62.6%를 차지하고 있으며 6,286천 ha입니다. OECD 국가 중 핀란드, 스웨덴, 일본에 이어 4위 수준이죠. 산림의 웅장한 크기는 단위 면적당 임목축적으로 표현하는데요. 한국 산림의 ha당 임목축적은 165.2㎥입니다. 5.66㎥/ha였던 1953년과 비교하면 거의 30배 가까이 증가하며 웅장함을 유지하고 있다고 해요. 한국 산림의 수관활력도는 92%로 높은 수치라고 하네요. 산림은 활용도가 무궁무진한 천연자원입니다. 온실가스를 흡수하면서 기후 위기에 대응하는 방안이 되기도 하고, 토사 유출을 방지하며 아름다운 경관으로 국민들에게 힐링과 휴양의 경험을 제공하기도 하죠. 2018년 조사에 따르면 산림의 공익적 가치는 총 221조 원이라고 해요.

메가 트렌드를 만드는 X-세대(영포티)

MZ세대 바로 위에는 오늘의 주인공인 X세대가 있습니다. X세대는 캐나다 작가 더글러스 커플랜드의 1991년작 소설 <X세대>에서 유래된 용어입니다. 한국에서는 주로 1980년대 후반~1990년대에 청년 시절을 보낸 1972년 전후에 태어난 이들로 이른바 ‘깁 세대’라 불리는데 2022년 2월 기준 8,166,792 명(만 40~49세)이 있습니다. 이전 40대와 달리, X세대는 시대의 흐름에 발맞춰 젊게 살아 영포티(Young Forty)라고 부르는데, MZ세대보다 구매력이 뛰어나고, 베이비부머 세대보다 트렌드에 민감한 X세대는 오늘날, 새로운 메가 트렌드를 창출해내는 세대로 문화산업의 변화를 온 몸으로 겪었고, 아날로그와 디지털을 모두 경험한 세대로 변화에 익숙하고, 트렌드에 관심이 높습니다.

동네를 접수하다! “하이퍼로컬”

“#지역기반 #로컬컨셉 #슬세권 #동네사람들 #정보공유” 하이퍼로컬은 말 그대로 지역을 기반으로 서비스하는 다양한 플랫폼을 말합니다. 글로벌 시장조사업체인 리서치앤드마켓은 2019년 약 9,730억 달러 규모인 하이퍼로컬 서비스 시장이 2027년까지 약 20% 성장해 약 3조 6343억 달러 규모에 이를 것이라고 전망했습니다. 디지털타임스의 ‘언택트 시대 슬세권 ’하이퍼로컬‘ 인기몰이’ 기사에 따르면 오픈업은 뉴욕에서만 151%, 시카고에서 68%에 달하는 사용자 증가세를 보이며 급성장하고 있으며, 동남아의 캐러셀에서는 생활용품을 넘어서 자동차, 부동산까지 거래되는데, 2012년 설립된 캐러셀의 기업가치는 9,850억 원에 육박하며 열 번째 동남아 유니콘 기업 후보로 꼽혔습니다.

그 시절 기억나니? 통계로 시간여행

KOSIS 국가 통계 포털에서 제공하고 있는 ‘통계로 시간여행’ 서비스에서 2021년 짜장면 한 그릇 가격을 5,000원으로 설정한 후 가격을 비교해 보니, 1990년 1,102원에서 2020년 4,851원으로 30년 동안 짜장면 가격이 약 4배 이상 인상된 것을 알 수 있었습니다. 30년 전에는 4인 가족도 5,000원으로 짜장면 외식을 할 수 있었겠네요. <그땐 그랬지>항목을 통해 그 시절 주요 사건을 알아볼 수 있습니다. ‘그 시절 스케치’에서 주요 사건들의 사 진을, ‘간추린 그 시절’에서는 해당 연도의 주요 사건들을 나열해 제공하고 있습니다. 제가 태어난 1999년에는 고궁에서 밀레니엄 기념 덕수궁 패션쇼, 마이클 잭슨의 내한 공연, 서울 지하철 8호선 개통, 20세기 최후 개기일식 등을 알 수 있었습니다.

비주류에서 주류로, 각광받는 키덜트(어린이)

어른이 되어도 어린 시절의 향수를 자극하는 물건을 만나면 사람들은 자신도 모르게 아이가 되곤 합니다. 바로 키덜트 이야기인데요. 최근 16년 만에 돌아온 포켓몬빵이 품절 사태를 맞이한 걸 보면 키덜트의 힘은 대단해 보입니다. 한국일보 기사에 따르면 영화 개봉 전보다 스파이더맨 키워드 검색 수는 20만 회에서 80만 회로 4배 가까이 증가했고, 피규어 관련 상품도 약 5,500개가 증가했으며, 스파이더맨 피규어 관련 상품 클릭 연령대를 살펴보면 40대가 40%, 30대가 25.5%, 20대가 13.5% 10대가 13.1%였습니다. 한국콘텐츠진흥원의 '2021 캐릭터 이용자 실태조사'의 최근 1년 캐릭터 상품 구매 경향을 살펴보면, 키덜트/취미 제품에 대한 소비는 30대가 37.1%로 가장 높았습니다.

스포츠계의 변화, Z세대를 주목하라!

Z세대는 1990년대 중반에서 2000년대 초반에 걸쳐 태어난 젊은 세대를 이르는 말입니다. 최근 2022 베이징 올림픽에서도 Z세대 선수들의 넘치는 열정과 멋진 활약이 돋보였는데요. 대한체육회 스포츠 지원포털에서 <선수 학년별 현황>을 확인하면 등록된 선수 중 Z세대는 운동부 및 클럽 기준으로 2021년에 중등부는 32,787명, 고등부는 25,610명, 대학부는 13,515명이 등록하여 Z세대에 속하는 선수는 총 71,912명의 선수임을 알 수 있습니다. 트렌드 모니터에서 1,000명을 대상으로 조사한 <2022 베이징 동계올림픽 관련 인식조사>에 따르면 79%의 사람들이 요즘 올림픽에 참가하는 젊은 선수들은 메달을 못 따도 주눅 들지 않고 경기를 즐기는 경향이 뚜렷한 것 같다고 응답했습니다.

1만 건의 정보를 1분 안에 분류! AI 활용 온라인 가격조사

2021년 12월 온라인 쇼핑 거래액은 17조 8,165억 원으로 최대치를 기록했는데요. 코로나19 팬데믹 이후, 비대면 중심의 일상이 보편화되면서 온라인 구매 형식이 더욱 증가하고 있습니다. 통계청에서는 매일 약 250만 건 이상의 온라인 가격정보를 수집하고 있지만, 온라인 가격정보는 일관된 형태나 형식이 일정하지 않고, 그 수가 어마어마하게 많아서 자료를 수집, 관리하는데 어려움이 있어 인공지능(AI)을 활용하는 방법이 논의되고 있습니다. 통계청 빅데이터통계과에서는 AI 알고리즘을 활용한 「온라인 가격정보 정제 시스템」을 개발하고 있습니다. AI 알고리즘을 활용하면 1만 건의 데이터에 대해 약 1분 안에 99.3%의 확률로 품목과 단위 정보를 신속하게 식별할 수 있습니다.

내 가치대로 구매한다! 가치소비

가치소비는 자신이 지향하는 가치나 만족도를 기반으로 제품을 구매하는 합리적 구매 방식입니다. 트렌드 모니터에서 진행한 『착한 소비 활동 및 SNS 기부 캠페인 관련 조사』에 따르면, 가치소비 관련 인식에 ‘물건 하나를 사더라도 개념 있는 소비를 하려는 사람들이 많아진 것 같다’고 답한 응답자가 78.1%나 차지했어요. 또한 성장관리업 ‘그0우’에서 발표한 『가치소비 관련 통계』 자료에 따르면, MZ세대 중 자신이 가치소비자인가를 묻는 문항에서 ‘그렇다’라고 답한 비율은 79%에 달했어요. 특히, 최근 함께 이슈가 되고 있는 ESG 경영에 대해서는 환경(Environment)에 대한 관심이 64.7%로 가장 높았고, 실천하고 있는 친환경 활동은 ‘리사이클링’이 40.7%로 1위를 차지했어요.

캠핑의 꽃은 흔적 지우기! 친환경 캠핑 실천도

“아름다운 사람은 머문 자리도 아름답습니다.” 캠핑의 꽃은 내 흔적 지우기라는 걸 아시나요? 한국관광데이터랩의 2020년 기준 캠핑 이용자 실태 조사에 따르면 항목별 친환경 캠핑 실천도는 분리수거 4.53점, 잔여 음식물 줄이기 4.27점, 일회용품 미사용 3.73점으로 나타났습니다. 국립생태원이 제시하는 친환경 캠핑법에 따르면, 첫째는 음식물 쓰레기를 줄이기 위해 먹을 만큼의 음식재료만 준비하기입니다. 둘째는 일회용품보다는 다회용품을 사용하고, 집에서 사용하는 수저와 접시, 텀블러 등을 사용하는 것이 좋겠죠. 셋째는 야생 동·식물을 보호하고, 주변 환경을 훼손하지 않는 것이 중요합니다. 끝으로 자나 깨나 불조심하세요. 건조한 날씨에는 특히 화재 예방에 신경 써야 합니다.

심판 너머 소년들의 삶, 드라마 <소년심판>과 소년범죄

소년범죄자는 만 19세 미만의 범죄를 저지른 사람을 말하는데, 나이에 따라 범죄소년, 촉법소년, 범법소년으로 나뉩니다. 형법상 만 10세 이상~만 14세 미만은 형사책임 능력이 없는 ‘촉법소년’으로 분류돼 사회봉사, 보호관찰, 소년원 송치 등 보호처분만 가능하며 전과 기록도 남지 않습니다. 한국 형사 정책 연구원 분기별 범죄 동향 리포트에 따르면 최근 3년간 소년범죄자의 수는 연평균 약 6만 5천여 명입니다. 2018년에서 2019년까지 소년범죄자 수는 소폭 증가했고, 2020년 다시 감소했습니다. 소년범죄자 통계만 보면 전체 범법 규모가 크게 느껴지지 않지만, 전체 범죄자 가운데 소년범죄자는 3.9%~4.0% 비중이며, 이 비율은 소폭 증감을 되풀이하고 있다는 것을 알 수 있습니다.

통계를 만들고 분석할 수 있는 분석도구, 「통그라미」 따라하기 (시각화)

정승호 | 영남고등학교 교사




실용통계교육을 위한 필수 학습도구! 통그라미

2016년 전면 도입된 중학교 자유학기제는 기존의 교사 중심의 강의식 수업에서 학생 중심의 참여형 수업으로의 변화를 가져왔으며 평가 역시 결과물이 아닌 과정중심의 평가를 권장하였다. 이로 인해 토의토론학습, 자기주도학습, 프로젝트 학습 등 다양한 교수·학습 방법이 도입되었다. 이와 더불어 2015 개정 수학과 교육과정에서는 통계교육의 내용을 생활 속에서 자료를 수집, 분석, 해석하는 활동 중심으로 개편하는 것을 방향으로 설정하였다. 교육과정의 개정과 교과서 개편으로 근본적인 방향은 제시하였지만 이를 학교 현장에 적용하기에는 많은 과제가 있었다.

이에 교육부는 통계교육원과의 협업을 통해 교사들의 통계적 소양을 함양시킬 수 있는 연수를 개설하였으며, 전국통계교사연구회를 공모하여 실제 현장에서 적용할 수 있는 통계교육 프로그램을 개발, 보급하여 실용통계교육이 이루어질 수 있는 기반을 마련하였다.

여기에 2016년 통계교육원이 개발한 통그라미의 web버전은 기존의 통계교육 프로그램과는 달리 자료의 수집과 분석을 모두 할 수 있었기 때문에 학교 현장에 널리 보급되기 시작했다. 클릭 한 번이면 자료를 분석할 수 있기 때문에 중학교뿐만 아니라 초등학교와 고등학교까지 보급되었다.



통그라미

통계학원

새문서

통계분석

참여문식

한글도루리

교과 활동사례

인쇄

공개 설문지

게시물 목록

설문조사

공개 설문지

설문 제목	공개	공개일			
번호	제목	작성일	등록일	조회수	추천수
24617	2022 하반기 통계학원 연구실 운영	날짜	20220418	5	0
24616	자주 묻는 질문: 통계학원 운영에 필요한 자료는 무엇인가요?	일	20220417	10	1
24615	통계학원 운영	일	20220416	4	0
24614	통계학원 운영	일	20220416	1	0
24613	통계학원 운영	일	20220416	4	0
24612	통계학원 운영	일	20220416	6	0
24611	통계학원 운영	일	20220416	6	0
24610	통계학원 운영	일	20220416	4	0

[그림1] 통그라미 공개 설문지 화면

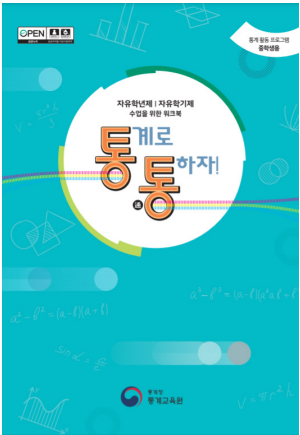
중학교에서 통계프로젝트 수업을 경험한 학생들이 고등학교로 진학하면서 좀 더 심화된 통계교육에 대한 요구로 인해 2019년 고등학교 인정교과서 “실용통계(대구광역시교육감 인정)”를 개발하게 되었다. 점점 늘어난 통계교육에 대한 요구로 통그라미의 기능이 업그레이드 되었다. 업그레이드된 통그라미 프로그램을 이용한 초, 중, 고등학생들을 위한 통계 교수학습 자료인 "샘이랑 통그라미하자(2020년, 통계교육원)"와 자유학기제 수업을 위한 워크북 “통계로 통하자!(2021년)”가 개발되었다.



[그림2] 실용통계 교과서



[그림3] 샘이랑 통그라미 하자



[그림4] 통계로 통하자

더욱 새로워지고 막강해진, 통그라미

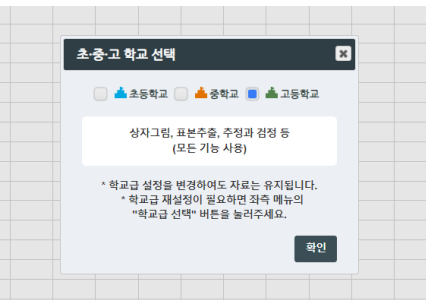
통그라미는 변화된 교육 환경과 사용자 편의를 위해 계속 업데이트되고 있으며, 고등학교 ‘실용통계’ 교과서의 개발로 인해 [표1]과 같이 표본추출, 분포, 추정과 검정 메뉴가 추가되었다.

[표1] 2021년 추가된 통그라미 메뉴

분류	기능	기능설명
그래프	워드 클라우드	분석 및 그룹변수 정의 후 통계분석 정보를 워드 클라우드(단어구름)로 표현함
	모션	분석변수 및 X축, Y축 정의 후 통계분석 정보를 모션차트로 표현함
	버블차트	분석변수 및 X축, Y축 정의 후 통계분석 정보를 버블차트로 표현함
통계	확률의 개념	확률을 정의하는 방법의 이해
표본추출	복원/비복원 추출	자료창에서 복원 또는 비복원추출
	표본평균의 분포	표본평균의 분포를 그래프로 표현
분포	이항분포	이항분포표 및 그래프 조회
	이항분포 정규근사	이항분포와 정규분포와의 관계, 이항분포확률
추정과 검정	모평균의 추정	모평균과 신뢰구간 추정
	모비율의 추정	표본 비율의 분포를 도표로 표시하고 신뢰구간 추정
	모평균 z검정	임계값과 표본평균을 비교하여 기각 및 인용여부 도출
	모비율 가설검정	임계값과 검정통계량을 비교하여 기각 및 인용여부 도출
	독립표본 T검정	독립적으로 존재하는 두 집단간의 평균 차이를 검정
	대응표본 T검정	동일한 표본을 대상으로 표본 차이를 검정

통그라미 시각화 과정 1 : 사용자의 편의성을 위한 학교급 구분

통그라미는 초등학생도 자료를 분석할 수 있도록 개발되었다. 하지만 통계 교육에 대한 수요 확대로 인해 통그라미는 고등학생까지 사용할 수 있도록 다양한 메뉴가 추가되었으며 이로 인해 어린 학생들이 사용에 어려움을 겪게 되었다. 이를 해결하기 위해 [그림5]와 같이 자료분석 시작 단계에서 학교급을 선택하고 이에 따라 [그림6]과 같이 학교급에 맞는 메뉴가 나타나도록 하였다. 학교를 선택하지 않을 경우 고등학교로 자동 선택된다. 학교급을 잘못 선택했을 경우, 왼쪽 메뉴 첫 번째에 있는 ‘초중고 학교 선택’을 클릭하여 변경할 수 있다.



[그림5] 초중고 학교 선택창



[그림6] 학교급에 따른 메뉴 표시

통그라미 시각화 과정 2 : 옵션 선택을 이용한 그래프 변형

기존의 통그라미는 옵션 선택에서 ‘그래프 추가’를 통해 동일한 화면에 다양한 그래프를 나타내어 주어진 자료를 설명하기에 적합한 그래프를 선택할 수 있도록 하였다. 하지만 개선된 통그라미에서는 선택한 그래프를 옵션 선택을 통해 설명하기 가장 좋은 형태로 변형할 수 있게 되었다.



[그림7] 기존의 줄기와 앞 그림

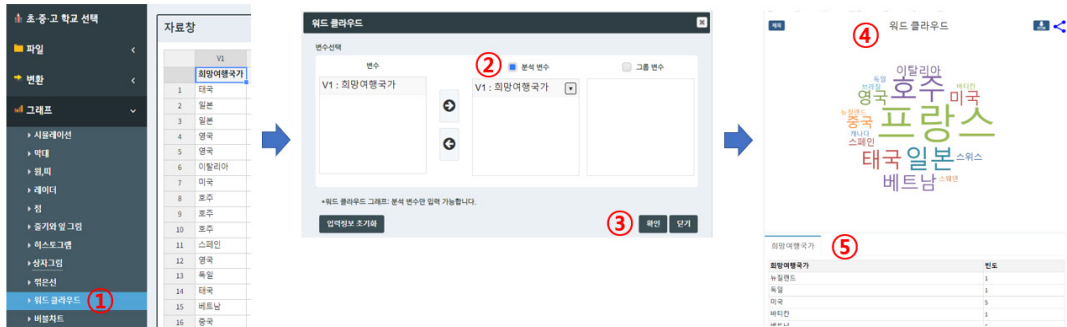


[그림8] 옵션선택에서의 설정

[그림7]은 이전 버전에서 표현할 수 있는 줄기와 앞 그림이다. 개선된 통그라미에서는 [그림8]과 같이 ① 옵션 선택을 클릭한다. ② 자릿수에서 줄기를 ‘2’로 변경한다. ③ 앞을 ‘1’로 변경한다. ④ 보이의 ‘빈 줄기 채우기’를 선택한다. 이와 같은 과정을 통해 자료의 분포를 설명하기에 적합한 줄기와 앞 그림을 만들 수 있다. 또, 가독성을 위해 줄기의 자료를 10개 단위의 다른 색깔로 표시된다.

통그라미 시각화 과정 3 : 시각화를 위한 그래프 추가

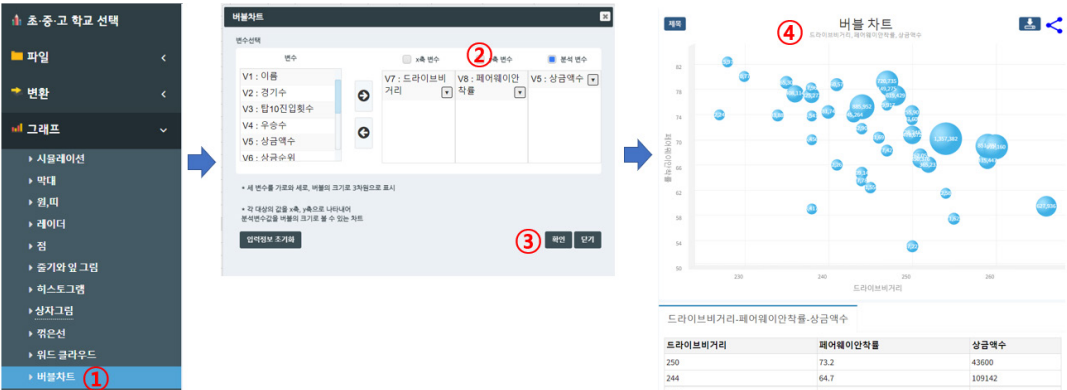
개선된 통그라미는 기존의 그래프에서 워드클라우드, 버블차트, 모션 그래프를 추가하였다. 워드클라우드는 텍스트를 시각화한 그래프로 특정 단어가 많이 나타날수록 단어의 크기가 더 크고 굵게 나타나도록 시각화한 그래프이다.



[그림9] 워드클라우드 만들기

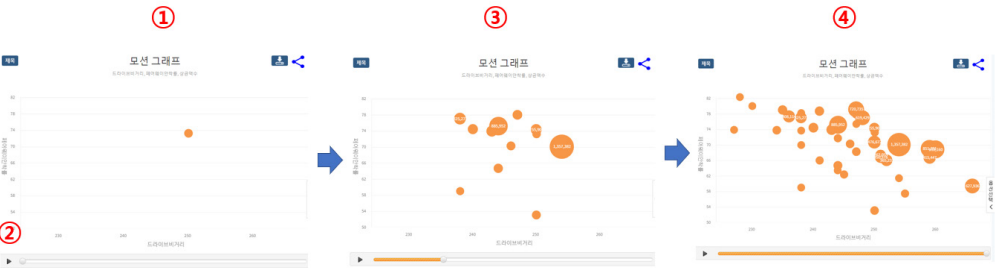
[그림9]와 같이 ① 왼쪽 메뉴의 그래프에서 ‘워드 클라우드’를 선택한다. ② 워드 클라우드 메뉴창이 나타나면 워드 클라우드로 나타내고 싶은 변수를 ‘분석변수’로 이동시킨다. ③ 워드 클라우드 메뉴창의 오른쪽 아래에 있는 확인을 클릭하면 ④와 같은 워드 클라우드가 나타나며 워드 클라우드 아랫부분에는 분석변수에 대한 ⑤와 같은 빈도표가 만들어진다. 워드 클라우드와 빈도표를 비교하면 워드 클라우드의 글자 크기가 단순히 빈도에 비례하여 커지는 것이 아니라는 것을 알 수 있다.

버블 차트와 모션 그래프는 3차원으로 표현하는 그래프로 각 대상에 대한 값을 x축과 y축에 나타낸 후 분석변수의 값을 버블의 크기로 나타내는 그래프이다. 버블 차트가 대상 전체를 하나의 그래프로 나타낸다면, 모션 그래프는 실행버튼을 눌러 각 대상(버블)의 움직임을 볼 수 있는 동적인 그래프이다.



[그림10] 버블차트 만들기

[그림10]과 같이 ① 왼쪽 메뉴의 그래프에서 ‘버블차트’를 선택한다. ② 버블차트 메뉴창이 나타나면 x축 변수, y축 변수, 분석변수에 나타내고 싶은 변수를 이동시킨다. ③ 버블차트 메뉴창의 오른쪽 아래에 있는 확인을 클릭하면 ④와 같은 버블차트와 자료를 정리한 표가 나타난다.

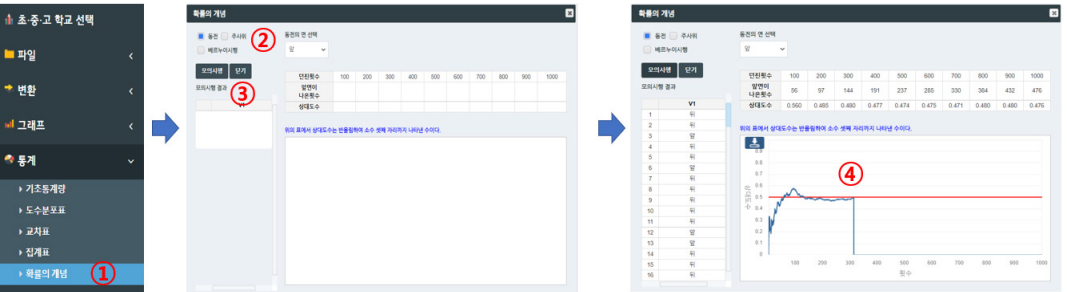


[그림11] 모션 그래프의 실행

모션 그래프는 버블차트의 움직임을 나타낸 그래프이므로 메뉴의 그래프에서 ‘버블차트’ 대신 ‘모션’만 다르게 선택하면 된다. 대신 결과에서 [그림11]와 같이 ① 모션 그래프가 나타난다. ② 왼쪽 아래에 있는 플레이(▶)를 클릭한다. ③ 모션 차트에서 버블이 추가로 나타난다. ④ 모션 그래프가 완성된다. 완성된 모션 그래프는 버블 차트와 같은 것을 확인할 수 있다.

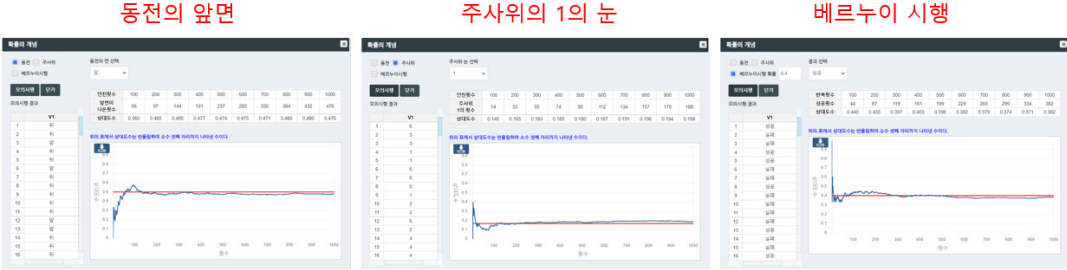
통그라미 시각화 과정 4 : 확률의 개념 이해를 위한 모의실험 추가

개선된 통그라미에서는 확률의 기본 개념인 ‘대수의 법칙(Law of large numbers)’을 설명할 수 있는 모의실험이 추가되었다. 대수의 법칙은 어떤 일이 몇 번이고 되풀이할 경우, 일정한 사건이 일어날 비율은 횟수를 거듭할수록 일정한 값에 가까워진다는 경험적인 법칙이다. 하지만 학생들은 일정한 횟수가 넘어서면 실험에서 나온 비율과 수학적 확률이 같아진다고 이해하고 있었다. 예를 들어 동전을 던질 때 앞면이 나올 확률은 1/2이므로 동전을 일정 횟수 이상 던지면 앞면이 나오는 비율이 0.5로 고정된다고 생각하는 경우가 많다. 이런 오개념을 해결해 줄 수 있는 것이 모의실험이다. 통그라미에서는 동전을 던지는 경우, 주사위를 던지는 경우, 임의의 결과가 ‘성공’ 또는 ‘실패’의 두 가지 중 하나로 나타나는 베르누이 시행의 3가지 경우를 모의실험으로 제공하고 있다.



[그림12] 모의실험

[그림12]와 같이 ① 왼쪽 메뉴에서 통계에서 ‘확률의 개념’을 선택한다. ② 확률의 개념창에서 실험의 종류(동전, 주사위, 베르누이 시행)를 선택한다.(베르누이 선택의 경우 성공할 확률에 0과 1사이의 값을 입력한다.) ③ 모의실험을 선택한다. ④와 같이 시행 횟수에 따른 비율의 변화(움직임)가 나타난다.



[그림13] 모의실험결과

[그림13]은 동전을 던졌을 때 앞면이 나오는 비율, 주사위를 던졌을 때 1의 눈이 나오는 비율, 베르누이 시행에서 성공할 확률이 0.4일 때 성공이 나오는 비율의 모의실험결과이다. 모의실험결과를 통해 실험 결과가 수학적 확률에 가깝지만 정확히 같은 값은 아니라는 것을 이해할 수 있다.

지금까지 개선된 통그라미의 기능 중 자료의 시각화와 관련된 부분에 대해 살펴보았다. 위에 설명된 부분 이외에 변수필터링을 통한 특정 그룹만 시각화하기, 그래프 색상 수정 등이 있다. 이는 굳이 설명을 하지 않더라도 충분히 사용할 수 있다. 이런 개선은 통그라미를 사용하는 모든 사용자 혹은 초등학생과 중학생을 위한 것이었다. 다음 연재에서는 고등학생들의 ‘실용통계’ 교과서 수업을 위한 표본추출, 분포, 추정과 검정을 살펴보고자 한다.

분석은 합의(合意)이다

강양석 | Deep Skill 대표

분석의 시대이다.
직관의 시대는
갔다까지 얘기한다.
그도 그럴 것이 인공지능은
곳곳에서 인간을 넘어서는
성능을 발휘하고 있으니 말이다.
그렇다면, 이 질문에도
답을 해보자.

“동일한 목적과 데이터를 가지면,
동일한 결과가 나올까?”

나름 이 질문이 중요하다고 생각하는 이유는 정말 분석이 직관을 넘어서 문제해결 능력을 가졌다면 이 질문에도 그렇다라고 얘기할 수 있어야 할 것이다. 왜냐하면 중지를 모으는 것은 문제해결의 중요한 필요조건 중 하나이기 때문이다. 마침 이 질문에 매우 적절한 연구 사례가 있어 소개해 보겠다.

전문지식을 공유하는 비영리 단체인 센터 포 오픈 사이언스(Center for Open Science)의 설립자 브라이언 노섹(Brian Nosek)은 2015년 ‘같은 데이터, 다른 결론(Same Data, Different Conclusions)’ 현상에 대한 실험 결과를 발표했다. ‘축구 심판들이 흑인 선수에게 레드카드를 더 자주 준다는 말은 사실인가?’를 밝혀내려는 이 실험에는 총 29개 팀 61명의 데이터 분석가가 참여했고, 이들에게 동일한 데이터가 제공되었다. 이들이 사용한 분석 기법들은 간단한 선형 분석에서 다중회귀 분석, 베이지안(Bayesian) 분석까지 다양했다.

결과는 흥미로웠다. 20개 팀이 ‘흑인 선수들은 레드카드를 더 받는다’는 결론을 내린 반면 9개 팀은 ‘피부색과 레드카드는 관련이 없다’고 결론 지은 것이다. 그저 접근법이 상이했던 것일 뿐인데 말이다.

각 팀이 사용한 분석 기법은 다른 팀의 검증 절차를 거쳤기에 분석 과정에서 오류가 있을 가능성은 없었다. 분석가들의 능력 부족이나 태만이 변수였을 가능성 역시 존재하지 않았다. 또 다른 프로젝트 리더이

동일 데이터, 다른 결과

29개의 리서치 팀에게 모두 동일한 축구 경기 데이터를 주었다. 그런 다음 피부가 검은 선수에게 심판이 레드카드를 더 많이 줄 가능성에 대해 의뢰했다. 각 팀은 서로 다른 통계기법을 사용했으며 피부색과 레드카드 사이에 관계성이 있음을 발견했다.



동일한 데이터로 동일한 문제에 접근해도 결과는 천양지차일 수 있다. 출처: Brian Nosek et al.

자 저명한 심리학자인 인시아드 비즈니스 스쿨(INSEAD Business School) 아시아 캠퍼스의 에릭 루이스 올만(Eric Luis Uhlmann)이 “모든 참가자는 최고 수준의 전문가였고, 답을 찾고자 하는 열정으로 가득 차 있었습니다”라고 못박았으니 말이다. 덧붙여 그는 이렇게 갈무리했다. “하나의 분석으로 궁극의(definitive) 답을 찾는 건 쉽지 않은 일입니다. 모든 결론은 다른 결론이 그것을 뒤엎기 전까지 한시적(temporary)으로 ‘답’이라는 위치에 있을 뿐입니다.”

결국, 아무리 같은 목적과 같은 데이터를 가진 전문가들이라도 서로 다른 접근법을 취하면 충분히 다른(subjective) 결론을 내릴 수 있다는 것이다.

‘한시적인 답’이 곧 분석 무용론을 뜻하는 것은 아니다. ‘한시성’을 강조하는 이유는 분석 결과의 유연성을 인정해야 자신의 문제를 자신의 데이터로 자유자재로 해결할 용기가 생기기 때문이다. 정답과 오답이 정해져 있는 것이 아니니 용기를 가지고 더 많은 설명력을 가진, 더 오래 정답의 지위를 고수할 답을 찾아가봐야 한다는 것이다.

분석 결과에 대한 막연한 맹신은 위험

분석적인 문제해결 방법이 직관적인 분석 대비 우리의 중지를 모으고, 동일한 해결 시도에 대해 답변의 일관성을 유지시키는 것은 당연할지라도, 근본적으로 분석이 ‘답의 다양성과 변화무쌍함’을 해결해주지는 못한다. 즉, 직관과 분석 중 무엇이 우월하냐는 불필요한 질문을 넘어 분석에 대한 막연한 맹신에서 벗어나는 것이 중요하다. 그것이 진짜 분석의 힘을 기르는 시작이라고 생각하고 말이다. 그럼 분석의 정체를 다시 생각해보자. 결론부터 말하면, 분석은 ‘고도의 합의’ 행위이다. 자연과학처럼 ‘필연성’을 가진 답을 주는 것은 아닌 그때 그때 좋은 합의에 의한 인공적인 결과물이라는 것이다. 이를 간접적으로 느껴 보기 위해 아래의 상황을 가정해 보겠다.

각 버스별 도착 예정 안내 시각과 실제 도착 시각 데이터(단위: 분)

각 버스회사가 안내한 버스별 도착 예정 시각

	월	화	수	목	금
1000번 버스	8:00	8:04	8:06	8:10	8:02
2000번 버스	8:10	8:06	8:00	8:02	8:10

실제 조사한 각 버스별 도착 시각

	월	화	수	목	금
1000번 버스	8:03	8:06	8:05	8:10	8:03
2000번 버스	8:15	8:09	8:00	8:00	8:10

모두 오전 시간 의미

예를 들어 당신이 최근에 이사를 해서 새로운 통근 버스를 정해야 하는 상황이라고 가정해보자. 회사에서 집까지의 동일 노선에는 서로 다른 2종의 버스가 있다. 어떤 버스가 좋을지 판단하기 위해 당신은 1주일 동안 집 앞 정류장에 버스가 도착할 예정 시각에 대해 버스 회사가 제공한 정보와 실제 도착 시각을 조사해본 뒤 다음과 같은 데이터를 얻게 되었다.

자, 그럼 당신은 어떤 버스를 주로 이용하는 것이 가장 좋을까? 당연히 미리 안내된 시각에 잘 맞춰 오는 버스일 것이다. 그런데 ‘잘 맞춰 온다’는 것은 너무 일찍 와도 안 되고 너무 늦게 와도 안 된다는 것을 의미한다. 버스가 예정 시각보다 일찍 오든 늦게 오든 당신이 정류장에서 기다려야 하는 건 마찬가지니 말이다. 또 실제 도착하는 시각이 안내 정보와 달리 너무 들쭉날쭉하면 그 버스를 이용해서 출근할 때 마다 매번 마음을 졸여야 한다. 그래서 당신은 각 버스별 예정 시각과 실제 시각의 차를 구한 후 평균을 내보기로 했다.

각 버스별 도착 시간의 차이(단위: 분)

	월	화	수	목	금	평균
1000번 버스	3	2	(1)	0	1	1
2000번 버스	5	3	0	(2)	0	1

	월	화	수	목	금	평균대기시간
1000번 버스	0	4	6	10	2	4
2000번 버스	10	6	0	2	10	6

실제 도착한 시각이 더 늦은 경우는 플러스 값으로, 반대의 경우는 마이너스 값으로 평균을 구했더니…… 이런! 두 버스의 평균이 공교롭게도 동일하게 1이다. 그럼 두 버스 중 어떤 것을 타든 상관없다고

할 수 있을까? 아니다. 중요한 것은 원래의 안내 시각에서 벗어난 정도의 평균값이 아니라 ‘얼마나 많이 벗어나는가’다. 벗어난 정도가 클수록 당신이 정류장에서 기다리는 시간은 길 것이고, 도착 예정 시각에 대한 믿음이 안 생기기 때문이다. 그래서 당신은 또 한 번 가벼운 가공을 해보기로 한다. 바로 각 버스의 도착 예정 시각에서 벗어난 정도, 즉 표준편차를 구하는 것이다. 비록 평균값은 같더라도 ‘평균에서 얼마나 벗어나는가’라는 새로운 기준을 대입하니 1000번 버스가 훨씬 낫다는 것을 금세 알아차릴 수 있다.

각 버스별 요일별 편차의 제곱, 분산, 표준 편차

	월	화	수	목	금	분산	표준편차
1000번 버스	4	1	4	1	0	2.0	1.41
2000번 버스	16	4	1	9	4	6.8	2.61

1000번 버스의 표준편차가 2000번 버스의 그것보다 작는데, 이는 예정 시각에서 벗어난 정도가 작음을 의미하기 때문이다. 이상은 아주 간단하지만 표준편차라는 개념을 적절히 활용하여 문제를 풀어본, 전형적인 예제 풀이식 데이터 기반 의사결정의 연습이었다. 그럼 이제 이 문제는 시원히 해결된걸까? 어떤 사람은 이런 풀이 과정 어딘가에서 석연치 않은 느낌을 받았을 수도 있다. 이미 정답이라는 게 존재하니 그것에 끼워 맞춰져야 한다고 강요 받은 듯한 느낌 말이다.

1000번 버스					변동의 폭 = 3 - (-1) = 4
8:00	8:04	8:06	8:10	8:02	
8:03	8:06	8:05	8:10	8:03	
3	2	-1	0	1	

또는 평균치에서 벗어난 정도를 판단의 기준으로 삼아야 한다는 상황 자체가 와닿지 않았을 수도 있다. 그렇다면 문제의 상황을 충분히 자기화하여 다른 방식으로 자기화한 답들이 나올 수 있을지 살펴보자. 문제 인식, 해결 기준 제시, 최선의 해결 방안 지목 등의 모든 과정은 자기 자신을 중심으로 이뤄질 수밖에 없다. 때문에 개인화된 문제해결 방안이 더 있지 않을지에 대해서도 의식적으로 살펴봐야 한다. 그래야만 유연성과 실천적인 답을 찾아낼 관찰력과 근성이 길러진다.

모범 답안에서 한 걸음 더 나아가 자기화한 경우를 생각해보자. 가령 ‘버스가 빨리 오는 상황과 늦게 오는 상황은 절대 같지 않다’고 얘기할 수 있다. 버스가 예정 시각보다 늦게 올 경우 고작 몇 분만 기다리면 되는 데 반해 빨리 올 경우엔 아예 버스를 놓치고 말기 때문이다. 주어진 데이터에는 나오지 않았으나 만약 그 다음 버스와의 배차 간격이 길다면 충분히 설득력 있는 주장이다. 즉, 어떤 버스를 탈지 결정하는 기준에 대해 ‘늦게 오는 버스는 조금 기다렸다가 탈 수 있지만 일찍 오는 버스는 정말 문제다’라고 충분히 자기화한 뒤 ‘1000번 버스는 1분 차이로 일찍 왔으나 2000번 버스는 2분 일찍 왔으니 1000번 버스가 더 안정적’이라는 결론을 내릴 수 있다. 물론 ‘1000번 버스가 더 낫다’는 결론 자체는 동일하지만

각 버스회사가 안내한 버스별 도착 예정 시각

	월	화	수	목	금
1000번 버스	8:00	8:04	8:06	8:10	8:02
2000번 버스	8:10	8:06	8:00	8:02	8:10

실제 조사한 각 버스별 도착 시각

	월	화	수	목	금
1000번 버스	8:03	8:06	8:05	8:10	8:03
2000번 버스	8:15	8:09	8:00	8:00	8:10

이유가 바뀔 수 있는 것이다. 그런가하면 ‘내가 왜 버스 도착 시각에 영향을 받아야 해? 난 내 시간에 버스를 맞추겠어’라는 또 다른 자기화도 가능하다. 이런 경우 ‘난 항상 아침 8시에 정류장에 도착한다’는 사람이라면 ‘8시를 기준으로 평균 대기 시간이 짧은 1000번 버스를 택하겠다’는 결론도 충분히 설득력 있는 답이 될 수 있다.

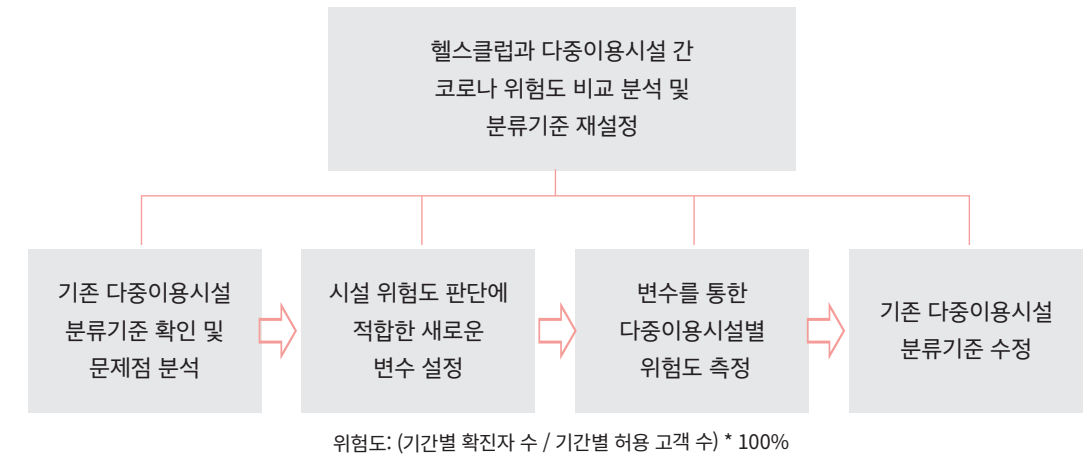
그럼, 각각의 버스가 가장 일찍 오는 시각과 가장 늦게 오는 시각을 파악한 후 그 격차가 작은 버스를 택하는 방법은 어떨까? 1000번 버스는 월요일에 가장 늦게 오고(3분), 수요일에 가장 일찍 오므로(1분) 그 격차는 3-(-1)=4분이 된다. 같은 방식으로 계산하면 2000번 버스의 시간 격차는 7분이 되므로 역시 1000번이 낫다는 결론이다. 표준편차를 활용했던 앞선 방법과 마찬가지로 이 접근법 역시 ‘예정된 시각에서 버스가 벗어나는 경향’을 보겠다는 점은 같다. 하지만 문제가 있다. 실제로는 예정 시각을 더 안 지키는 버스지만 우연히 최대-최소 값의 격차가 적어 더 좋은 버스로 여겨질 수도 있기 때문이다.

이런 식으로 따지면, 얼마든지 다른 방법들을 상상해 볼 수 있다. 예를 들면, 5일 중 정시에 온 횟수가 많은 버스를 ‘성실함’의 기준으로 보는 것이다. 일종의 정시 도착률 같은 것이다. 아니면, 앞서 얘기한 ‘빨리 오는 것과 늦게 오는 것은 같지 않다.’에 착안하여, 5일 중 빨리 오는 경우만을 제외한 날짜수를 세는 방법도 있을 수 있다. 어떤가? 정말 다양하지 않은가?

개념에 대한 가치관에 따라 분석 기법도 달라진다

이렇게 분석의 종류가 다양하게 존재하는 것은 사전적으로 ‘바람직한 버스’에 대한 상(想, 이미지)이 다르기 때문이다. 당연히 이 개념은 매우 주관적일 수 밖에 없고 말이다. 즉, 바람직한이란 개념에 대한 각자의 가치관이 다르기 때문에 그 가치관을 대변하는 분석 기법도 달라지는 것이다. 실제로, 문제를 둘러싼 가치관의 문제는 분석 기법 선정에 매우 중요한 영향을 미친다.

예를 들면, 어떤 사업체의 성과평가를 하려는 경우 그 사업이 어떤 주기(태동-발전-성숙-쇠퇴)에 있느냐



에 따라 그 평가지표는 시장점유율(%), 매출액(원), 매출성장률(%), 이익규모(원), 이익성장률(%)등 매우 여러가지로 나타낼 수 있듯이 말이다.

이 그림은 코로나 시국에 다중이용시설의 위험도를 그 다중시설에서 발생한 확진자 수로만 볼 것이 아니라, 그 다중시설을 이용한 사람 수 대비 확진자 수로 봐야한다는 취지의 분석 기획 내용이다. 언뜻 들으면 그럴싸 해보인다. 하지만, 잘 생각해보면 확진자 발생 위험도는 효율성 관점이 아닌 효과성 중심으로 볼 문제라고 볼 수 있다. 왜냐하면 10명 중 2명이 발생한 다중위험시설보다 1000명 중 5명이 발생한 다중위험시설이 더 위험하다고 볼 수 있는 이유는 다중이용시설을 이용한 사람들의 편의의 크기와 관계없이 무조건 확진자가 발생한 절대 규모 자체가 적은 것이 방역 관점에서 중요하기 때문이다. 즉, 분석을 지배하는 가치관의 힘에 따라 분석의 내용은 강하게 지배 받을 수 밖에 없는 것이다.

이렇듯, 분석은 각자가 가진 가치관과 그 가치관을 대변하는 분석 어프로치에 따라 늘 다른 답을 주게 되어 있다. 그러니, ‘분석은 정답을 구하는 행위이다.’라는 오해는 ‘나는 맞고, 너는 틀리다.’라는 굉장히 위험한 발상으로 이어질 수 있다. 늘 소통과 합의로 답을 찾는 것이 아니라, 서로 빚어내는 행위라는 것을 알아야 하는 것이다. 내가 앞으로 연속으로 진행될 데이터 리터러시 기고의 첫 글로 이 분석의 유연성을 강조하는 이유는 ‘기법과 툴 중심’의 데이터 학습에 대한 강력한 우려임과 동시에 데이터 기반 문제해결에 대한 좀더 입체적인 이해를 해달라는 당부이기도 하다. 그래야 데이터 리터러시에 필요한 생각하는 힘, 감각, 가치관 같은 단어가 데이터, 분석 기법/툴이란 단어와 함께 등장할 때 어색하지 않을 것이기 때문이다. 다시 말해, 분석은 답을 주는 행위가 아니라 고도의 합의의 산물이라는 생각을 견지해야 매 순간 더 나은 답이 있을 수 있다는 마음의 여유와 근성이 생긴다.

교육과정별 운영계획

가. 집합과정

구분	과정명	수준	교육대상	교육 일수	기당 인원	교육 횟수	교육 일정
집합교육(96)							143
A 기본교육(8)							12
1	9급 신규자 기본교육(공채)		9급 공채 신규임용예정자	15	60	2	3.14~4.1. 11.21~12.9.
2	9급 신규자 기본교육(경채)		9급 경채 신규임용예정자	10	-	-	-
3	7급 신규자 실무교육		6~7급 신규자	4	20	1	6.21~6.24.
4	4급 승진후보자 역량향상		4급 승진후보자	5	25	1	2.21~2.25.
5	5급 승진후보자 역량향상		5급 승진후보자	5	24	2	4.11~4.15. 8.22~8.26.
6	6급 승진자 역량향상		6급 승진자	4	30	2	5.10~5.13. 10.25~10.28.
7	현장조사 역량강화		조사담당자	3	30	2	4.27~4.29. 5.25~5.27.
8	지방형 조사관리 역량강화		지방통계청	3	30	2	9.21~9.23. 11.9~11.11.
B 전문교육(85)							124
B-1 국가통계정책(12)							17
B-1-1 통계정책(7)							12
9	국가통계의 이해	초급	제한없음	3	20	2	4.27~4.29. 9.5~9.7.
10	통계와 정책	중급	통계청	3	20	2	4.18~4.20. 7.6~7.8.
11	정책과정과 통계의 역할	초급	제한없음	3	30	1	9.19~9.21.
12	국가승진통계관리	초급	통계작성기 관	3	30	2	5.2~5.4. 9.14~9.16.
13	통계품질관리	중급	제한없음	3	30	2	2.14~2.16. 7.13~7.15.
14	정책지도 작성 방법론	초급	제한없음	2	20	1	9.5~9.6.
15	통계기초 및 활용	초급	제한없음	5	30	2	5.23~5.27. 9.26~9.30.
B-1-2 통계기초(5)							5
16	한국표준산업분류	초급	제한없음	3	30	1	3.2~3.4.
17	사회분류(직업분류 등)의 이해	초급	제한없음	3	30	1	6.8~6.10.
18	한국표준직명·사인분류의 이해	초급	제한없음	1	30	1	4.22.
19	한국표준직명·사인분류의 활용	중급	제한없음	1	30	1	11.11.
20	한국표준건강보험의 이해(신설)	초급	제한없음	1	30	1	5.20.
B-2 국가통계이해(10)							12
B-2-1 경제통계(3)							4
21	경제통계의 이해	중급	제한없음	4	30	1	3.22~3.25.
22	국민계정	중급	제한없음	3	20	1	9.5~9.7.
23	재무계표	중급	제한없음	3	30	2	6.22~6.24. 8.24~8.26.
B-2-2 사회통계(5)							5
24	농어업통계의 이해	중급	제한없음	3	30	1	9.5~9.7.
25	인구통계의 이해	중급	제한없음	3	20	1	4.13~4.15.
26	보건통계의 이해	초급	제한없음	3	30	1	10.5~10.7.
27	소득분배지표의 이해	중급	제한없음	2	20	1	8.4~8.5.
28	고용통계의 이해	중급	제한없음	2	20	1	6.28~6.29.
B-2-3 기타(2)							3
29	빅데이터와 행정자료의 이해	초급	통계청	3	30	2	3.14~3.16. 7.25~7.27.
30	지속가능발전목표(SDGs)의 이해	초급	제한없음	1	20	1	7.6.
B-3 국가통계작성(16)							22
B-3-1 조사기획(4)							6
31	국가통계실무(조사설계 및 조사표설계 등)	초급	통계청	4	40	2	3.14~3.17. 8.30~9.2.
32	국가통계실무(표본설계 및 추정)	초급	통계청	4	40	2	4.11~4.14. 9.27~9.30.
33	조사설계 및 조사표설계	초급	제한없음	3	20	1	4.18~4.20.
34	표본실무	중급	제한없음	3	20	1	7.20~7.22.
B-3-2 자료수집·처리 및 분석(8)							9
35	통계조사관 직무연수	초급	통계청	3	30	1	8.17~8.19.
36	지역통계실무	초급	제한없음	3	20	1	3.16~3.18. 5.17~5.20.
37	국가통계실무(자료수집·처리 및 분석)	초급	통계청	4	40	2	10.24~10.27.
38	자료수집·처리 및 분석	초급	제한없음	3	20	1	8.29~8.31.
39	계절조정실무	중급	통계청	2	30	1	1.10~1.11.
40	데이터 에디팅	중급	제한없음	4	20	1	7.11~7.14.
41	시계열분석	중급	제한없음	4	20	1	7.4~7.7.
42	지수이론	중급	제한없음	3	20	1	7.13~7.15.
B-3-3 공표 및 관리(4)							7
43	국가통계실무(통계작성·공표 등)	초급	통계청	4	40	2	6.14~6.17. 11.22~11.25.
44	국가통계정보의 활용	초급	제한없음	3	20	1	5.2~5.4.
45	통계데이터 비밀보호의 이해	중급	제한없음	3	20	1	6.27~6.29. 3.28~3.30.
46	통계보고서 작성	중급	제한없음	3	30	3	6.8~6.10. 10.12~10.14.
B-4 데이터 분석(16)							28
B-4-1 R(4)							7
47	R 초급 통계분석	초급	제한없음	4	30	2	2.21~2.24. 8.16~8.18.
48	R 데이터 시각화	중급	제한없음	3	30	2	4.6~4.8. 8.31~9.2.

구분	과정명	수준	교육대상	교육 일수	기당 인원	교육 횟수	교육 일정
49	R 중급 통계분석	중급	제한없음	3	30	2	6.27~6.29. 10.4~10.6.
	50 R 고급 통계분석	고급	제한없음	3	30	1	7.11~7.13.
B-4-2 SAS(3)							4
51	SAS 초급 통계실무	초급	통계청	4	20	1	3.28~3.31.
52	SAS 중급 통계분석	중급	제한없음	4	20	2	4.11~4.14. 7.4~7.7.
53	SAS 고급 통계분석	고급	제한없음	4	20	1	10.17~10.20.
B-4-3 SPSS(2)							4
54	SPSS 중급 통계분석	중급	제한없음	4	30	2	5.10~5.13. 9.26~9.29.
55	SPSS 고급 통계분석	고급	제한없음	4	30	2	7.19~7.22. 11.14~11.17.
B-4-4 파이선(2)							4
56	파이선 초급 통계분석	초급	제한없음	3	20	2	4.20~4.22. 8.10~8.12.
57	파이선 중급 통계분석	중급	제한없음	3	20	2	6.20~6.22. 10.24~10.26.
B-4-5 엑셀(2)							5
58	엑셀 초급 데이터 분석 및 활용	초급	제한없음	3	30	3	3.2~3.4. 6.20~6.22. 10.31~11.2.
59	엑셀 중급 통계분석	중급	제한없음	3	30	2	4.25~4.27. 10.11~10.13.
B-4-6 빅데이터(3)							4
60	행정자료 통계작성	중급	제한없음	3	20	1	6.13~6.15. 6.29~7.1.
61	하둡 기반 빅데이터 분석	중급	제한없음	3	20	2	11.2~11.4.
62	빅데이터 프로젝트 수행(SAS기반)	고급	제한없음	4	20	1	7.18~7.21.
B-5 전문가 양성(9)							4
B-5-1 통계 전문가(7)							2
63	경제시계열분석 및 지수이론	고급	통계청	14	10	-	-
64	국민계정	고급	통계청	12	10	-	-
65	인구통계분석	고급	통계청	12	10	-	-
66	통계데이터기획	고급	통계청	13	10	-	-
67	표본설계 및 추정	고급	통계청	13	10	-	-
68	무응답 자료처리 및 분석	고급	통계청	12	10	-	-
69	통계데이터 비밀보호	고급	통계청	12	10	-	-
B-5-2 데이터 사이언스 트랙(2)							2
70	데이터 사이언스 전문교육	고급	통계청	23	30	1	-
71	AI 적용 데이터 사이언스 전문교육(신설)	고급	통계청	23	30	1	-
B-6 기타 통계교육(25)							36
B-6-1 맞춤형 통계교육(기타)							16
72	맞춤형 통계교육(기관)	통계작성기 관		2	25	12	-
B-6-2 통계세미나(1)							3
73	통계세미나	제한없음		1	40	3	-
B-6-3 학생교육(4)							7
74	사람나눔 통계교실	초·중·고 학생		3	25	2	-
75	어린이 통계캠프	초·중·고 학생		3	30	2	-
76	중학생 통계아카데미	중학생		3	30	2	7.20~7.22. 8.17~8.19.
77	고등학교 통계아카데미	고등학생		2	30	1	8.11~8.12.
B-6-4 교사교육(7)							12
78	초등학교 교사 통계연수	초등학교 교사		2	30	2	1.11~1.12. 8.8~8.9.
79	중학교 교사 통계연수	중학교 교사		2	30	2	1.20~1.21. 8.8~8.9.
80	고등학교 교사 통계연수	고등학교 교사		2	30	2	1.13~1.14. 8.1~8.2.
81	중등 교사 통계연수(심화)	중·고등학교 교사		3	30	2	1.10~1.12. 7.25~7.27.
82	실용통계 지도교사 통계연수	고등학교 수학 교사		4	30	2	1.17~1.20. 7.25~7.28.
83	시도교육청 교사 통계연수	초·중고 수학 교사 등		2	30	-	-
B-6-5 외국인교육(1)							2
84	통계를 활용한 통계사회 지도교사 연수	고등학교 사회과 교사		2	30	2	1.24~1.25. 8.1~8.2.
B-6-5 외국민교육(1)							2
85	UNSIAP 공조 통계연수	외국공무원		5	20	2	-
B-7 정보화(2)							5
86	오피스를 활용한 데이터 시각화	중급	제한없음	3	40	4	3.2~3.4. 5.25~5.27. 9.19~9.21. 11.7~11.9.
87	디지털영상 및 이미지활용	초급	제한없음	3	30	1	10.19~10.21.
C 기타교육(3)							7
C-1 사책교육(1)							1
88	소셜미디어 활용	제한없음		3	20	1	6.8~6.10.
C-2 일반소강(2)							6
89	생활속의 음악	제한없음		3	60	2	8.3~8.5. 11.2~11.4.
90	현장조사 스트레스 관리		통계청	3	30	4	5.25~5.27. 6.29~7.1. 9.28~9.30. 10.26~10.28.

나. 이러닝 과정

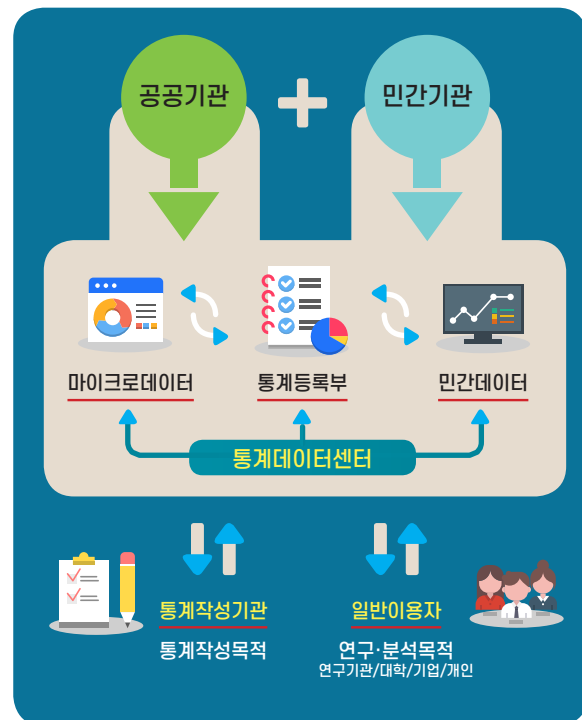
분	과정명	수준	교육대상	인정시간
이러닝 (115)				
A 기본교육(11)				
1	5급 승진후보자 역량평가의 이해	초급	통계청	3
B 전문교육(80)				
B-1 국가통계정책(15)				
B-1-1 통계정책(13)				
2	국가통계 및 제도	초급	제한없음	1
3	성인지통계의 이해와 활용	초급	통계청	1
4	지체통계활용집단 관리	초급	제한없음	4
5	지역사회지표 작성과 활용	초급	제한없음	4
6	지역정책과 통계 활용	초급	제한없음	11
7	통계가 쉬워지는 기초 수학 끝내기	초급	제한없음	2
8	통계기초 및 활용	중급	제한없음	13
9	통계법	초급	제한없음	2
10	통계업무 필수 지식	초급	제한없음	3
11	통계작성기관을 위한 통계DB시스템 사용법	초급	제한없음	2
12	통계적으로 사고하기	중급	제한없음	3
13	통계품질관리	초급	제한없음	1
14	통계학의 이해	초급	제한없음	7
B-1-2 통계기초(2)				
15	한국표준산업분류	초급	제한없음	6
16	한국표준직업분류	초급	제한없음	5
B-2 국가통계이해(24)				
B-2-1 경제통계(5)				
17	광업 제조업동향조사	초급	통계청	4
18	서비스업동향조사	초급	통계청	3
19	소비자물가지수	초급	통계청	5
20	온라인쇼핑동향조사	초급	통계청	2
21	재무계표 이해	중급	제한없음	4
B-2-2 사회통계(19)				
22	가계동향조사	초급	통계청	3
23	가족동향조사	초급	통계청	1
24	경제활동인구조사	초급	통계청	3
25	경제활동인구조사 사례집	초급	통계청	2
26	농가경제조사	초급	통계청	3
27	농가판매 및 구입가격 조사	초급	통계청	1
28	농산물생산비조사	초급	통계청	4
29	농업면적조사(신설)	초급	통계청	3
30	농작물생산조사(생산량부문)	초급	통계청	2
31	사회통계의 이해	초급	제한없음	6
32	신자별입조사	초급	통계청	2
33	양곡소비량조사	초급	통계청	2
34	어가경제조사	초급	통계청	3
35	어류양식동향조사	초급	통계청	2
36	어업생산동향조사	초급	통계청	3
37	어민자 체류실태 및 고용조사	초급	제한없음	3
38	인구동향조사	초급	통계청, 공무원	3
39	지역별고용조사	초급	제한없음	3
40	축산물생산비조사	초급	통계청	3
B-3 국가통계작성(17)				
B-3-1 조사기획(11)				
41	조사방법의 이해	초급	제한없음	7
B-3-2 자료수집 처리 및 분석(7)				
42	표본학식 기초	초급	제한없음	27
43	자료수집	초급	제한없음	2
44	조사원 면접기법교육(신설)	초급	제한없음	1
45	시계열자료의 분석과 실무	초급	제한없음	5
46	표본론 기초	중급	제한없음	6
47	현장조사 인력양성	초급	제한없음	4
48	회귀분석의 이해와 사례	초급	제한없음	7
B-3-3 공표 및 관리(9)				
49	e-나라지표 업무시스템 이용방법	초급	통계청, 공무원	2
50	MDIS 활용	초급	제한없음	2
51	NARA-PC 활용	초급	제한없음	6
52	SGIS 예뉼	초급	제한없음	4
53	SGIS 플러스 활용	초급	제한없음	2
54	국가통계포털(KOSIS) 활용	초급	제한없음	2
55	나라통계시스템 통계제공 사용자교육(신설)	초급	제한없음	2
56	통계데이터 비밀보호(신설)	초급	제한없음	3
57	통계를 활용한 보고서 작성방법	중급	제한없음	4
B-4 데이터 분석(16)				
B-4-1 R(2)				
58	R 기초	초급	제한없음	12
59	R 활용	중급	제한없음	12
B-4-2 SAS(2)				
60	SAS를 활용한 고급 통계분석(신설)	초급	제한없음	6
61	SAS를 활용한 중급 통계분석(신설)	중급	제한없음	6
B-4-3 SPSS(2)				
62	SPSS 초급 통계분석	초급	제한없음	9

행정통계자료와 민간자료를 한곳에!

통계데이터센터 서비스

통계데이터센터가 새로운 서비스로
정보화 사회를 선도합니다.

행정자료를 수집하여 가공한 **행정통계자료(통계등록부)**,
통계청이 제공하는 승인된 **통계기초자료(마이크로데이터)** 등
통계자료뿐만 아니라 **민간자료**까지 한 곳에서 분석이 가능한 통계데이터센터(SDC)

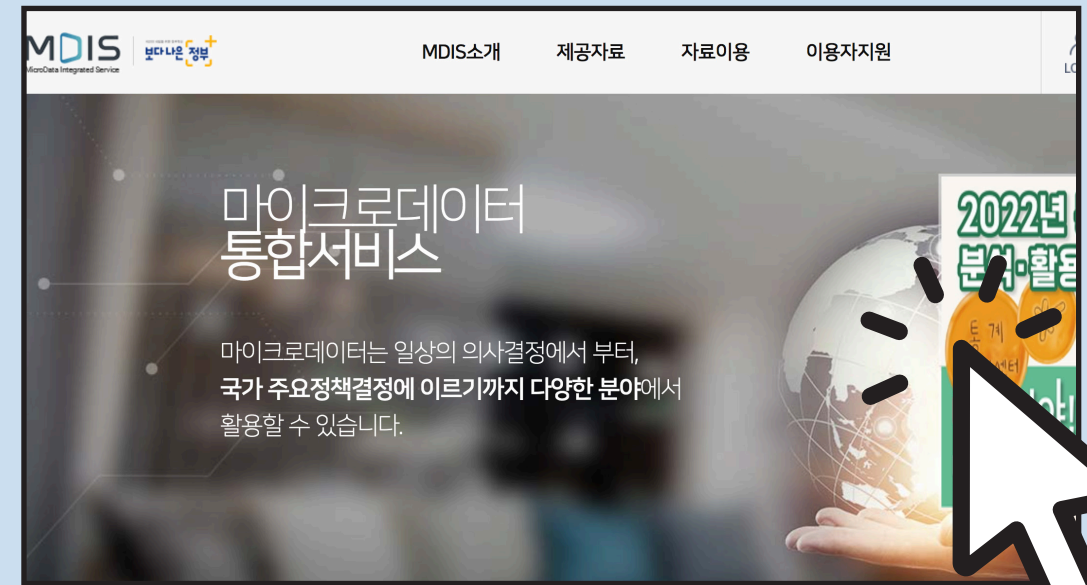


제공서비스

분석센터 이용서비스 <ul style="list-style-type: none"> 데이터분석 플랫폼 제공 (분석시스템, 통계패키지) 통계자료 및 민간자료, 이용자 반입 자료 등 연계·분석 	주문형 분석서비스 <ul style="list-style-type: none"> 시간 및 거리상 센터 방문이 어려운 이용자를 위한 서비스 이용자가 원하는 형태로 결과 제공(비식별화처리된 형태)
교육지원 <ul style="list-style-type: none"> 대학생을 대상으로 통계데이터 센터의 실제 데이터를 통한 다양한 분석 체험 기회 제공 통계학과 등 관련학과와 정규 전공 수업 실습 분석 지원 및 이용 지원 	전문가분석 지원 <ul style="list-style-type: none"> 분석 경험이 없는 이용자를 위한 분석 지원 및 상담
명부서비스 이용 <ul style="list-style-type: none"> 통계작성기관이 승인통계 작성 목적으로 요청할 때 조사표본 명부 제공 	

빅데이터를 활용한 서비스 (홈페이지 내 제공)

통계청, 정부부처, 지방자치단체, 연구기관 등 모든 기관의 마이크로데이터를 한 곳으로



보다 심도 있고 다양한 분석을 원한다면
지금 바로 MDIS를 클릭해보세요.

■ 서비스 소개 (2021년 11월 기준)

가. 서비스명 : 마이크로데이터통합서비스(MDIS, mdis.kostat.go.kr)

나. 제공통계수 : 21개 주제별 총 327종 통계 제공(통계청 49종 및 통계작성기관 278종)

다. 제공형태 : 마이크로데이터(통계에 따라 사람, 사업체, 가구 기반 자료)

기준	주요 통계
인구·가구	경제활동인구조사, 가계동향조사, 국내인구이동통계, 사망원인통계, 가계금융복지조사, 지역별고용조사, 인구주택총조사, 인구동향조사, 생활시간조사, 사회조사 외 8종
통계청 사업체·농어가	전국사업체조사, 광업·제조업조사, 농가경제조사, 기업활동조사, 농림어업총조사, 농산물생산비조사, 경제총조사, 어가경제조사, 운수업조사 외 14종
행정통계 및 기타	귀농귀촌인통계, 영리법인기업체행정통계, 신혼부부통계, 주택소유통계, 중장년층행정통계, 퇴직연금통계, 일자리행정통계, 기업생멸행정통계
통계작성기관	전국다문화가족실태조사, 가족실태조사, 자동차주행거리통계, 직종별사업체노동력조사, 보육실태조사, 기상관측통계, 국민여가활동조사, 외래관광객실태조사, 한부모가족실태조사, 청소년종합실태조사 외 207종

■ 서비스 내용

가. 구분 : 자료의 민감성 정도에 따라 공공용, 인가용으로 구분 운영

나. 수수료
- 무료 : 공공용 자료
- 인가용 : 선택제 수수료 부과

다. 서비스 방법
- 추출·다운로드 : MDIS 포털에서 직접 무료 다운로드
- 원격접근서비스 : 승인 후 이용자가 집·사무실 등에서 통계청 서버 접속 후 활용
- 이용센터 : 승인 후 지정된 장소를 방문·활용

■ 문의

- 연락처 : 재단법인 한국통계진흥원
- 전화 : (02) 512-0167 FAX : (02) 515-0240
- 주소 : (우) 06097 서울특별시 강남구 선릉로 612, 6층
- E-mail : MDIS@stat.or.kr

통계청에서 국가통계를 활용하세요!

통계청은 통계개발·활용·교육에 필요한 모든 정보와 도움을 제공합니다.

다양한 국가통계정보 제공사이트를 활용하세요.



통계교육원



sti.kostat.go.kr

국내 유일의 국가통계교육 전문기관

통계작성 및 활용 전문통계과정,
기관맞춤형과정, e-러닝 과정

통계데이터센터



data.kostat.go.kr

행정통계자료와 민간자료를 한곳에

행정통계자료(통계등록부), 민간자료의
연계·융합이 가능한 데이터 플랫폼

MDIS



mdis.kostat.go.kr

원하는 자료를 직접 분석 및 요청

온라인으로 추출/다운로드 선택 시
공공용 마이크로데이터를 무료로 분석 활용 가능

KOSIS



kosis.kr

국가통계 쉽게 찾기

국내, 국제, 북한의 주요 통계를
한 곳에 모아 알기 쉽게 분류해 제공

SGIS



sgis.kostat.go.kr

지도 위 통계정보 살펴보기

인구, 가구, 주택, 사업체 통계 등 각종 통계를
지도(GIS) 위에서 한눈에 파악



통계청
통계교육원