

통계의 창

WINDOW OF STATISTICS

2024.
SUMMER
VOL.33



여론조사 결과는 왜 각각 다를까?

초거대 시가 쏟아올린 일상 속 통계활용의 시대 | 통계를 누구나 쉽고 재밌게 즐길 수 있는 KOSIS 통계시각화콘텐츠
공공과 민간 데이터의 '효율적 공유와 통합' 통해 새로운 가치 창출 | 굴뚝 없는 황금산업, MICE산업

CONTENTS

통계의 창
2024. Summer
Vol.33

발행일 2024년 6월 16일
발행인 안형준
발행처 통계교육원
기획 황현식, 정명진, 박현지
주소 대전광역시 서구 한밭대로 713(월평동) 통계센터 통계교육원
전화 042-366-6151, 6152
팩스 042-366-6498
이메일 mjjung@korea.kr, hyunjii@korea.kr
디자인 및 인쇄 (주)피그마리온(02-516-3923)

ISSN 2005-1379
©2024. 통계교육원
※ '통계의창'에 실린 내용은 필자 개인의 의견이므로 필자의 소속기관이나 본지의 공식적인 견해를 대변하는 것은 아닙니다.



2

통계 FOCUS

- 초거대 AI가 쏟아올린 일상 속 통계활용의 시대** 018
신우람 | 통계청 통계정보플랫폼과 사무관
- 통계를 누구나 쉽고 재밌게 즐길 수 있는 KOSIS 통계시각화콘텐츠** 024
강명규 | 통계청 통계서비스기획과 사무관
- 통계로 보는 올해 '뚝뚝해지는' 부자 키워드7** 032
이승주 | Newsis 기자
- 생성 AI 시대 최고의 교수법** 040
박남기 | 광주교대 교수
- 공공과 민간 데이터의 '효율적 공유와 통합' 통해 새로운 가치 창출** 046
김창훈 | KRG 대표
- 굴뚝 없는 황금산업, MICE산업** 052
최서우 | 한국데이터연구소 대표
- Open New Era! 첨단 기술 기반 스포츠 경기분석 시스템** 060
천제민 | 스위트케이 상무이사
- 빅데이터와 시대에 살고 있는 우리들에게 절실히 필요한 직관력 및 상상력, 독서량** 068
김선일 | 빅હે브솔루션(주) 대표이사
- 통계로 바라보는 세상이야기** 076
신동헌 | 도서출판 지일박스 대표

1

통계 ISSUE

- 여론조사 결과는 왜 각각 다를까?** 004
이은수 | 케이스탯컨설팅 CS조사실 실장
- 선거 여론조사의 시각화 자료를 제대로 이해하기** 010
김진호 | 스위스 경영대학 시빅데이터학과 교수

3

통계 EDU

- '데이터 분석 제목 달기' 심층 알아보기** 080
강양석 | Deep Skill 대표
- 스몰 데이터로 새로운 가치 창출하기** 084
구자룡 | 벨류바인 대표

01 여론조사 결과는 왜 각각 다를까?

이은수 | 케이스탯컨설팅 CS조사실 실장

최근 국회의원 총선거가 치러졌고 그 이전에 수없이 많은 여론조사 결과가 매일 쏟아져 나왔다. 그런데 이러한 여론조사 결과가 조사기관마다 각각 다르고 심지어는 동일한 기관에서 발표하는 여론조사 결과도 수시로 다르게 나타나서 후보자로 나서는 정치인들은 물론 국민들에게도 혼란을 주고 여론조사에 대한 불신이 점차 팽배해지고 있다. 이렇게 조사결과가 조사기관마다 다르게 나타나는 이유에 대해서 가장 중요한 몇 가지만 설명하고자 한다.



사회 여론조사의 결과에 영향을 미치는 가장 중요한 요인은 조사방법과 표본추출 Frame, 질문방법, 설문항목의 배치 순서, 설문항목의 wording 등이다. 참고로 필자가 설문조사 관련 강의를 다니다 보면 많이 듣는 질문 중 하나가 “표본수가 적으면 대통령 지지율(또는 고객만족도 점수)이 더 떨어지는 것 아닌가?”라는 질문인데 표본수는 조사결과의 신뢰성에 대한 문제이지 조사결과의 방향에 대해 영향을 주는 요인은 아니다. 즉 표본수가 적으면 표본오차가 늘어나서 그만큼 신뢰성이 줄어드는 것이지 대통령 지지율이나 고객만족도 점수 등이 높아지거나 낮아지는 것과는 영향이 없다.

대부분의 여론조사는 표본조사이다

표본조사란 관심의 대상인 전체 집단(대상 모집단, Target Population)의 일부 또는 전체인 조사 대상자(표집모집단, Sampled Population)를 선정하고 표집모집단으로부터 추출한 표본을 대상으로 조사하여 얻은 표본의 특성인 통계량(Statistic)으로 모집단의 특성인 모수(Parameter)를 추정 또는 검증하는 과정이다.

예를 들어 대통령의 지지율을 조사하고자 한다면 관심의 대상인 전체 국민(대상 모집단)으로 부터 표본을 추출할 표집모집단(국내에 거주하는 19세 이상의 성인 남녀)을 선정하고 이 중 1,000명의 표본을 추출하였다면 1,000명의 표본을 대상으로 조사하여 얻은 대통령 지지율(통계량)로 전국민의 대통령 지지율(모수)을 추정하는 것이다.

그런데 동일한 조사를 여러 번(예를 들면 100번) 반복한다고 가정하자. 그러면 100번의 조사결과가 모



두 동일하게 나타나지 않고 조금씩 다르게 나타날 수 있다. 이는 각각의 표본을 추출하는 과정에서 대통령에게 호감이 있는 응답자들이 조금 덜 추출될 수도 있고, 조금 더 추출될 수도 있으며 이러한 표본으로 추출된 응답자들의 구성에 따라 대통령 지지율이 조금씩 달라질 수 있을 것이며 이렇게 표본을 추출하는 과정에서 발생하는 오차를 표본오차라고 한다.

한편 표본조사란 조사목적에 따라 그에 가장 타당한 조사대상자(표집 모집단)의 특성을 정의하고 조사대상자의 특성에 따라 조사 방법이 결정되며 조사항목, 질문방법 등이 결정되는데, 이러한 표본추출, 조사방법 등에 따라 조사결과가 (대부분은 오차 범위 내에서) 조금씩 달라지게 된다.

■ 먼저 조사방법의 차이

현재 국내 정치여론조사의 대부분은 전화조사와 ARS조사가 사용된다. 경우에 따라서 web-up 설

문지를 이용한 자기기입식 조사(흔히 온라인 조사라 한다)와 개별면접조사(face to face interview)가 사용되기도 하지만 이는 흔히 사용하는 방법이 아닌데 온라인 조사는 정치여론조사의 경우 표본의 모집단 대표성 문제가 있고 개별면접조사는 조사원에 의해 발생하는 오류와 비용의 문제, 표본추출 등의 현실적 문제로 인해 자주 사용되지 않는다.

ARS조사는 기계음에 의해서 질문항목을 듣고 응답자가 응답하는 조사방식이며 전화조사는 조사원이 직접 전화를 통해 읽어주는 질문항목을 듣고 응답하는 방식이다. 각종 언론매체의 진행자들이 '전화면접 조사'라는 용어를 사용하는데 이는 ARS조사와 구별하기 위해서인 것 같은데 그냥 '전화조사'라고 하면 될 것으로 필자는 생각한다. 즉, 전화조사와 ARS 조사로 구분하여 명칭하는 것이 더 타당하다고 생각한다. 아무튼 두 조사방법의 가장 큰 차이는 응답률이며 응답률은 표본의 모집단 대표성 문제와 직결된다. 예를 들어 ARS조사의 경우 (조사 지역과 설문지 길

이에 따라 다르지만) 정치여론조사의 평균 응답률이 2%~4% 정도이며 전화조사의 경우 평균 응답률이 10~20%로 나타난다. 즉 ARS조사의 경우 통화가 연결된 사람 100명 중에 2~4명이 응답을 하는 반면 전화조사의 경우 100명중에 평균 10~20명이 응답을 하며 따라서 ARS 조사보다는 전화조사가 훨씬 모집단의 대표성이 담보될 것이다.

결국 ARS 조사는 표본의 모집단 대표성이 상대적으로 떨어지며 따라서 정치적/사회적 이슈에 관심이 많은 집단이 응답할 가능성과 이에 따라 조사 결과가 모집단의 현상과 다소 다르게 나타날 가능성이 전화조사보다 상대적으로 높다는 문제가 제기되고 있다. 예를 들면 정치적 관심이 높은 사람이 많이 추출되는 ARS조사의 경우 전화조사에 비해 지지정당 질문에서 '무응답/중도층'이 적게 나타날 수밖에 없고 따라서 지지정당이나 정치인에 대한 지지율이 영향을 받을 가능성이 있는 것이다.

즉 두가지 조사방법의 차이 → 응답률 차이 → 응답자의 정치적/사회적 관심 차이 → 조사결과의 차이로 연결될 수도 있다는 의견이다.

한국조사협회에서는 ARS조사에 대해 “ARS를 이용한 조사가 과학적인 조사방법이 아니라는 점에 동의하고 향후 ARS조사를 수행하지 않을 것을 결의한다”고 규정하고 있다.

반면 한국정치조사협회 회원사들을 중심으로 ARS 조사를 진행하고 있는 조사기관들도 상당수 존재하는데 “응답률 차이에 의한 응답의 bias가 생각보다 크지 않으며, 오히려 전화조사에서 조사원의 성향에 의한 bias가 더 클수 있다”는 의견을 제시하기도 한다. 즉 응답자들이 명확하게 응답을 하지 않는 경우 조사원들이 임의로 판단을 하여 응답처리하는 경우가 발생하고 이 때 조사원들의 성향에 의한 bias가 발생할 수 있다는 입장이다. 물론 조사원에 의한 전화조사를 진행하는 기관은 모든 전화조사 내용을 녹음하거나 실시간 감청하는 등의 데이터 신뢰성 확보를 위한 노력을 하지만 100% 정확하게 관리가 된다고 보증할 수 없는 한계가 존재하는 것도 현실이다.

여기서 두 조사방법의 효율성에 대한 필자의 개인적 견해를 밝히는 것은 적절하지 않다고 생각하고 다만 조사목적과 이에 따른 응답자의 특성에 따라 효율적인 적절한 조사방법을 선택하여 사용하는 것이 좋지 않을까 하는 의견을 조심스럽게 표명하고 싶다.

■ 다음은 표본추출 Frame의 차이

현재 전화조사 진행시 표본을 추출하는 Frame은 RDD(Random Digital Dialing)방식과 안심번호를 통한 표본추출 방법이 주로 사용된다. (과거에는 유선의 경우 KT 전화번호부 등록 DB를 사용하기도 했지만 이 또한 모집단의 대표성 문제가 있어 현재는 거의 사용하지 않은 방법이다.)

RDD는 무선번호의 010 - □□□□ - □□□□ 에서 뒤의 8자리 번호를 무작위로 발생시켜서 기계가 전화를 걸고 (결번이 아닌) 유효한 번호이면 조사를 시도하는 방식이다. 그런데 이 RDD 방법은 조사대상자의 성별/연령/거주지역을 알 수가 없는 한계가 있다. 예를 들어 전국의 대통령 지지도를 조사한다고 할 때 통상적으로 지역별/성별/연령별 인구비례에 의한 표본수를 할당하는 것이 일반적이다.

그런데 010 - 1111 - 1111 번호를 가진 홍길동이라는 사람의 연령, 거주지역, 성별 등에 대한 개인정보가 없다. 따라서 무작위로 번호를 추출하는 RDD 조사방법의 경우 전국의 지역별 / 성별 / 연령별 표본구성의 quota를 맞추기가 현실적으로 더욱 어렵다. 따라서 조사가 완료된 후 실제 인구성비에 의한 가중치를 부여하여 분석하는데 문제는 가중치 부여 비율이 너무 크다는 것이다.

즉 실제 인구구성비와 조사가 완료된 표본구성비의 차이가 너무 크고 그 차이를 가중치로 조정하다보니 그 과정에서 발생하는 오류가 발생할 가능성이 크다는 것이다. 특히 전국 1,000표본을 조사한다고 할 때 강원/제주 등과 같이 인구수가 적은 지역의

경우 최소한의 유의한 표본수도 안 되는 표본을 조사하여 그 결과를 기준으로 해당 지역의 가중치를 부여하는 분석하는 것은 위험하다는 생각이다.

이와 같은 단점을 보완하기 위하여 안심번호를 사용한다

안심번호란 지역별/성별/연령별 휴대폰 번호를 통신사에서 제공받아서 조사를 진행하는 방식이다. 단 개인정보 보호를 위하여 실제 휴대폰 번호가 아닌 가상번호를 제공한다. 따라서 실제 지역/성별/연령별 인구구성비에 거의 유사한(동일한) 표본을 조사할 수 있으며 따라서 가중치 비중은 거의 실제 인구구성비와 다르지 않다는 장점이 있다.

반면 통신사별 이용자 구성(비중)이나 특성에 따른 차이가 있을 것이며 이에 의한 오류가 발생할 가능성이 있다는 단점도 제기되고 있다. 그러나 RDD조사의 실제 인구구성비와 크게 다른 표본 구성을 가중치로 부여하는 과정에서 발생할 수 있는 오류는 최소화할 수 있다는 차이가 있다. 그러나 개인정보 보호 차원에서 아직 선거여론조사에만 안심번호를 사용할 수 있다는 제약이 있다.

RDD에 의한 표본추출은 즉각적인 조사가 가능하다는 장점이 있는 반면 원래 표본구성 quota를 맞추

어 조사하기가 어렵다는 단점이 있고 안심번호에 의한 표본추출은 표본구성 quota를 맞추기가 상대적으로 용이한 반면 비용이 비싸고 안심번호를 사전에 미리 신청해야 한다는 (조사 10일 전에 신청해야 한다) 단점이 있는데 결국은 조사비용과 조사의 시급성에 의해 전화조사 표본추출 Frame을 결정하며 이에 따라 조사결과가 달라질 수 있는 것이다.

상기 설명한 두가지(RDD vs 안심번호, 전화조사 vs ARS) 요인 이외에 전화를 받지 않는 응답 대상자에 대한 반복 접촉(contact)도 조사결과에 영향을 미친다. 표본조사에서 중요한 요인 중 하나는 표본 구성의 모집단 대표성이다. 예를 들어 RDD나 안심번호를 통해서 확보된 조사 대상자들에게 전화를 걸었을 때 결번은 아닌데 안 받는 사람도 있을 것이며 “지금 바쁘다”고 거절한 사람에 대한 재접촉 여부와 그 방식에 따라 조사 결과가 달라질 것이다.

만약 전화를 안 받는 사람에게는 다시 접촉을 시도하거나 “지금 바쁘다”고 거절한 사람에게 통화가 가능한 시간을 물어서 그 시간에 다시 전화를 걸어서 접촉을 하는 방식과 전화를 걸어서 바로 전화를 받고 바로 응답을 해 주는 사람들만을 대상으로 조사를 진행하는 방식은 모집단의 대표성 차이가 크며 이 또한 조사 결과의 차이로 나타나는 것이다.

후자에 비해 전자의 방식이 모집단의 대표성이 압도적으로 높은 것이 사실이며 따라서 전자의 방식

대로 조사가 진행되어야 한다. 참고로 한국조사협회에서는 통화중이거나 부재중인 사람에게는 3번 이상 재접촉을 시도하여 최초 조사대상자로부터 응답을 받도록 장려하고 있다.

그밖에 조사결과에 영향을 미칠 수 있는 또 하나의 중요한 변수는 설문 구성 항목의 배열 순서이다. 예들 들어 정당지지도와 대통령 지지율을 조사하는 경우 대통령 지지율을 먼저 질문하고 정당지지도를 질문하는 경우와 정당지지도 → 대통령 지지율의 순서로 질문하는 경우의 조사결과도 달라질 수 있는 것이다.



추가적인 quota를 맞추려면 추가적인 비용과 기간이 소요될 수는 있지만 정확한 조사를 진행하기 위해서 고려할 필요가 있다고 생각한다.

응답자의 정치적 성향 구성비에 대한 quota 할당

마지막으로 필자가 한가지 개인적으로 지적하고 싶은 사항은 국정지지율이나 정당지지도 같은 정치여론조사의 경우 응답자의 정치적 성향에 대한 질문과 분류방법이다.

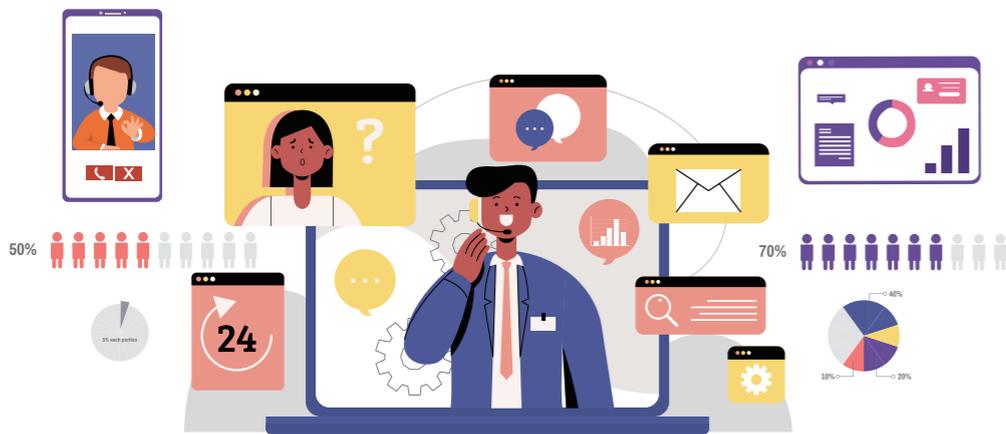
현재 정치여론조사에서 응답자의 정치적 성향이 조사결과에 매우 중요한 영향을 미치고 있다. 그런데 조사기관마다 심지어는 같은 조사기관에서 진행하는 조사에도 응답자들의 정치적 성향에 대한 구성비가 조사마다 조금씩 각각 다르다는 것이다.

이는 앞서 언급한 것처럼 표본을 추출하는 과정에서 정치적 성향이 진보적인(또는 보수적인) 사람이 적거나 많이 추출될 수 있고 이에 따라 조사결과가 달라질 수 있는 것이다. 따라서 한국조사협회, 한국조사연구학회, 또는 중앙선거여론조사심의위원회, 통계청 등의 유관기관에서 주기적으로(예를 들면 1년 단위나 6개월 단위) 한국 유권자들의 정치적 성향 구성비를 정확하게 조사하여 모든 조사기관들이 동일한 기준으로 quota를 할당하여 조사하는 방안이 필요하지 않을까 생각한다.

즉, 현재 지역별/성별/연령별 인구구성비에 따라 Quota를 할당하여 조사를 진행하는데 정치여론조사의 경우 여기에 동일한 정치적 성향 구성비의 기준을 추가하여야 하지 않을까 하는 생각이다. 물론

더불어 응답자 개인의 정치적 성향에 대한 분류를 응답자의 응답에 의존하고 있다는 것도 문제로 지적하고 싶다. 예를 들면 동일한 정치적 성향을 가지고 있는 두사람에게 본인의 정치적 성향을 질문할 때 한 사람은 자신의 정치적 성향이 진보, 또 다른 사람은 자신이 보수적 이라고 응답할 가능성이 적지 않다고 생각한다. 즉 정치적 진보와 보수에 대한 판단기준이 사람마다 제각기 다를 것인데 응답자에게 개인의 정치적 성향을 질문하는 것은 타당하지 않다고 판단된다. 유권자의 정치적 성향을 판단할수 있는 몇 가지 질문을 하여 이러한 질문의 응답결과를 종합하여 응답자의 정치적 성향을 분류하는 방법이 필요하지 않을까 생각한다. 오랜기간 실무적 경험이 있는 업계 전문가들과 한국조사연구학회 학자들의 공동연구에 의한 방법이 충분히 마련될 수 있을 것으로 기대한다.

여론조사회사는 여론을 수집하여 파악하는 기관이며 여론 형성에 영향을 끼쳐서는 안되는 기관이다. 따라서 가장 객관적이고 정확하게 여론을 파악하기 위하여 표본추출 과정과 조사과정에서 발생할 수 있는 모든 표본오차와 비표본오차를 최소화할 수 있도록 조사비용 등의 현실적 한계를 극복하고 업계와 학계, 유관 기관에서 모두 최선의 노력을 다해 관리해야 할 것이다.



02

선거 여론조사의 시각화 자료를 제대로 이해하기

김진호 | 스위스 경영대학 시빅데이터학과 교수



흔히 여론조사의 시대라고 불리는 요즘 우리는 바야흐로 각종 여론조사의 홍수 속에 묻혀 살고 있다. 특히 선거철에는 당이나 후보자에 대한 전화 조사가 수시로 걸려오고 신문과 TV에서는 각종 그래프와 수치로 장식된 조사결과를 실 사이 없이 발표한다. 이제는 매스컴의 어떤 기사도 여론 조사결과를 요약한 그래프 혹은 숫자를 함께 제시하지 않으면 사람들은 그 기사를 무언가 근거가 없는 주장처럼 여기곤 한다. 문제는 사람들이 여론 조사의 결과를 종종 올바르게 이해하지 못하거나 심지어는 조사 결과의 유리 혹은 불리에 따라 여론조사를 불신하는 경우가 많다는 것이다. 예를 들면 미국 부통령이었던 휴버트 험프리는 “정치인들은 자기 후보에게 유리한 결과가 나온 여론조사는 훌륭한 여론 조사라고 하고, 불리한 여론조사는 비과학적으로 행해진 조사라고 우긴다.”고 말했다. 이 말은 요즘에도 선거 여론조사에 대해 정치가들이 종종 하는 말이다.

비판적으로 보는 안목 길러야

선거 여론조사에 대한 불신을 해소하기 위해서는 유권자들, 즉 조사결과를 대하는 사람들이 우선 조사에 대한 안목을 길러야 한다. 조사에 대한 안목을 높인다는 말은 조사 과정에서 일어날 수 있는 잘못을 잘 인식하고, 그런 잘못이 일어났을 때는 그것을 피해가면서 결과를 올바르게 해석하는 것을 말한다. 그렇다면 선거 여론조사에서 가장 많은 왜곡이

일어나는 부분은 어디일까? 조사 결과를 시각화 하는 과정에서 가장 많은 왜곡이 일어난다.

많은 숫자나 분석 결과들을 요약해서 잘 설명할 수 있는 가장 효과적인 방법이 그래프 등을 사용하여 시각적으로 나타내는 것이다. 그래프는 가로축, 세로축, 점, 선, 숫자, 글자, 심볼 등을 복합적으로 사용하여 양적인 숫자들을 시각적으로 요약한다. 따라서 그래프를 대하는 사람의 입장에서는 익숙하지도 않은 많은 숫자들을 머리를 써서 생각할 필요

도 없이 단지 보는 것만으로도 숫자들 속에 포함된 사실을 파악할 수 있다. 하지만 문제는 그래프를 이용하여 정보를 전달할 때 가장 많은 왜곡이 일어난다는 사실이다.

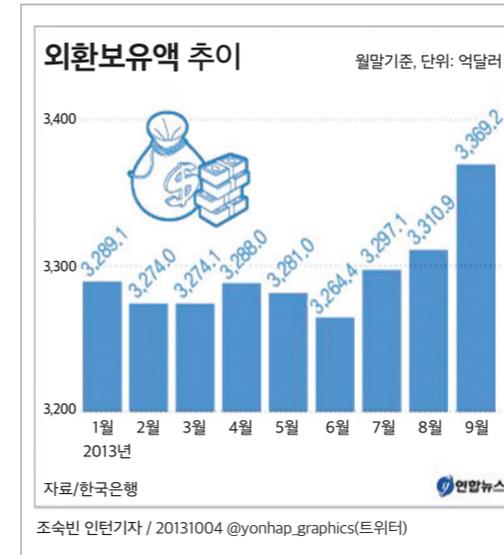
왜 그럴까? 많은 숫자(데이터)를 그래프로 그릴 때에는 가능한 한 간단하게 데이터가 갖고 있는 정보를 생기는 그림으로 전달해야 한다. 이때 중요한 것은 데이터를 단순화(simplicity)하면서도 데이터가 갖고 있는 사실(fact)을 충실하게(loyalty) 전달해야 하는 것이다. 하지만 이 두 가지 원칙(simplicity & loyalty)은 자칫하면 서로 상충하기 쉽다. 그래서 그래프를 그리는 것이 언뜻 보기에 매우 쉬운 듯하지만 상당한 기술(art)이 필요한 것이다.

실제로 그래프에서 속임수가 가장 많다는 사실은 데이터의 단순화 과정에서 사실에 대한 정확성을 유지하기가 어렵다는 것을 말해 준다. 다시 말하면 그래프의 미숙한 사용으로 데이터를 너무 단순화하면 의도하지 않았더라도 실제와는 전혀 다른 인상을 주는 그래프가 될 수 있다. 더욱이 그래프를 그리

는 사람이 사실을 의도적으로 왜곡하여 그리게 되면 그래프는 사실에서 크게 동떨어지게 된다. 그러므로 그래프가 제시될 때 그것을 비판적으로 볼 수 있는 안목을 길러서 혹시 있을지 모르는, 의도적인 혹은 비의도적인 '왜곡'에 빠지지 않는 능력을 키우는 것이 매우 중요하다.

그리는 사람 마음대로 왜곡

그래프 중에서 가장 흔한 것은 선을 이용한 선 혹은 막대 그래프다. 이 그래프들은 그리기가 쉬울 뿐만 아니라 많은 숫자 속에 숨어있는 경향을 잘 나타내어 주기 때문에 데이터의 분석이나 예측에 가장 많이 쓰인다. 그러나 데이터의 왜곡이 가장 많이 일어나는 그래프이기도 하다. 가장 흔한 사례는 그래프의 밑 부분을 잘라내는 것이다. 신문에 제시되는 많은 그래프가 지면절약 등의 이유 때문에 이런 식으로 흔히 그려진다. 이런 그래프는 원래의 차이를 부풀리기는 하지만 속임수가 아니면서도 독자에게 주는 인상은 크게 다르다.

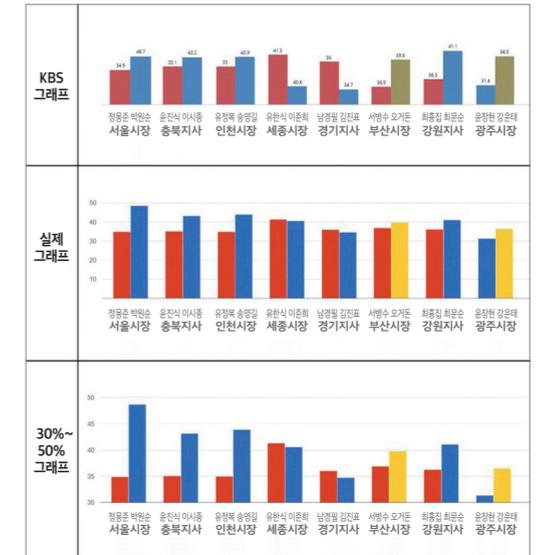


외환보유액 추이

위의 그래프는 월별로 외환보유액을 나타낸 것이다 (2013년).¹⁾ 9월의 외환보유액은 6월에 비해 거의 3배나 되는 것처럼 보이지만 실제로는 증가율이 2.9%에 불과하다. 그래프의 밑부분이 잘리면서 차이가 과장되게 나타나는 것이다.

위의 그림은 KBS가 2014년 6.4 지방선거를 앞두고 공개한 여론조사 결과의 일부인데 후보별 지지도를 나타내는 막대그래프가 특정 정당에 유리하도록 왜곡되어 그려져서 논란에 휩싸였었다. 막대그래프마다 눈금의 기준이 달라, 새누리당 후보가 실제 지지도보다 높은 지지도를 얻은 것처럼 비춰져 논란이 됐던 것이다. 위쪽 그래프는 KBS에서 방영된 그래프, 가운데 그래프는 제대로 그린 그래프, 그리고 맨 아래 그래프는 0에서 30%의 밑부분을 잘라낸 그래프이다.²⁾

서울시장과 충북도지사의 경우에는 위와 가운데의



MBC, SBS 공동조사

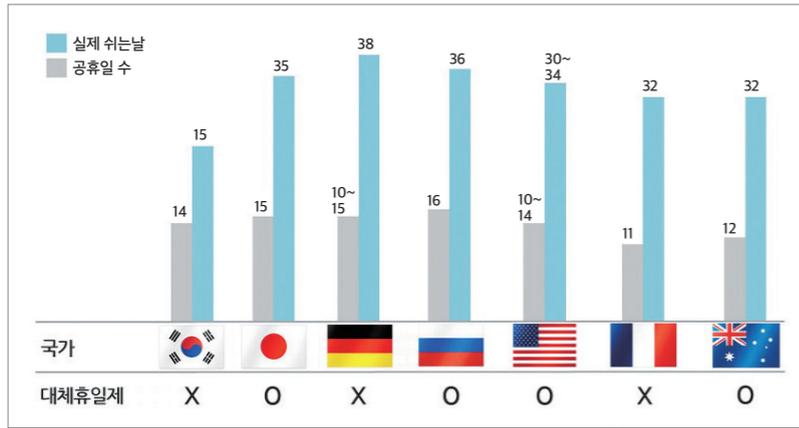
그래프가 큰 차이가 없다. 하지만 세종시장과 경기도지사의 경우에는 불과 1%p 정도의 차이가 나는 접전 상황인데도 불구하고 (가운데의 그래프의 중앙), KBS 뉴스에서 방영된 그래프에서는(맨 위 그래프의 중앙) 새누리당 후보가 크게 앞서는 것처럼 그린 것이었다. KBS가 선거를 앞두고 특정 정당에 유리하도록 편향된 보도를 한다는 비난을 받아, KBS는 뒤늦게 특정 정당에 유리한 모습을 보일 의도는 전혀 없었다고 해명하기 바빴다. 맨 아래 그래프는 0에서 30%의 밑부분을 생략한 그래프로서 후보간의 경쟁에 대한 전혀 다른 인상을 주고 있다.

그래프를 그리는 사람이 우선 지켜야 할 사항은 그래프의 수직 눈금의 크기를 일관성 있게 유지하는 것이다. 그러나 조사결과를 요약하는 그래프에서 이런 기본적인 원칙이 지켜지지 않는 경우가 종종 있다. 아래의 그림은 주요 국가별 공휴일 현황을 막대그래프로 그린 것이다.³⁾

역시 수직 축의 눈금이 그래프마다 제각각이다. 예

1) <https://www.yna.co.kr/view/GYH20131004000500044>

2) <http://www.vop.co.kr/A00000759397.html> 3) <http://www.valken.net/718>



주요 국가별 공휴일 현황 (2009년 기준, 자료: 한국문화관광연구원)

를 들어 빨간 사각형으로 나타낸 것과 같이 같은 15 일이지만 한국과 일본의 높이가 다르다.

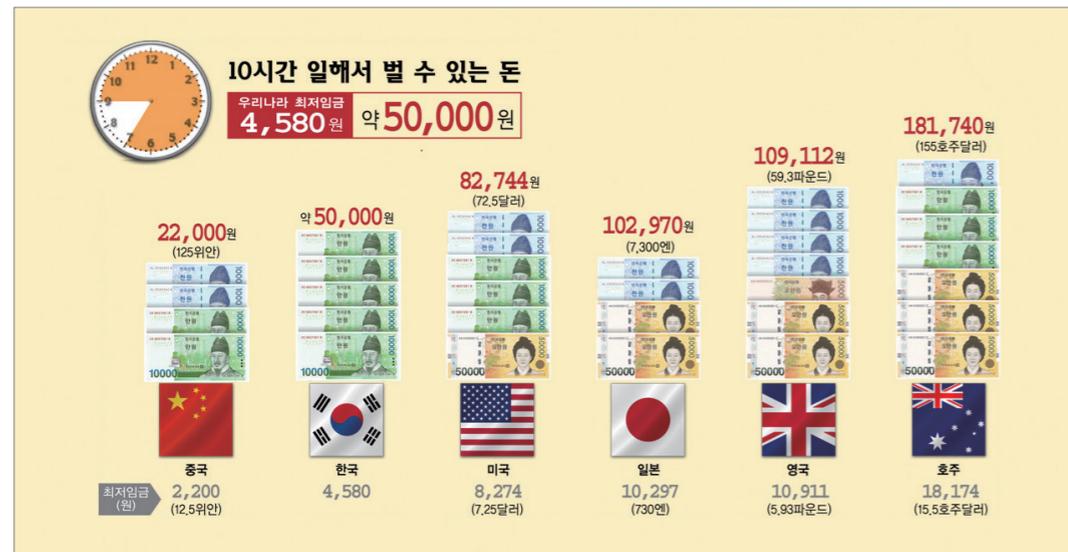
아래의 그림은 최저임금으로 10시간 일해서 벌 수 있는 돈을 그린 것인데 역시 수직축의 눈금이 제각각이다.⁴⁾

중국에서는 22,000원인 붉은 선 높이가 일본에서는 102,970원과 같게 그려졌다.

물론 이런 왜곡 주장에 대해 정작 그래프를 작성한

당사자는 각 막대그래프의 위에 구체적인 숫자를 적어 놓았으니까 그 숫자를 참조해서 보면 되지 않느냐고 반박할 수도 있다. 하지만 그래프의 원래 목적이 양적인 숫자들을 시각적으로 요약하여 보는 사람이 쉽도록 하는 것이다. 사람들이 그래프를 보면서 거기에 적힌 숫자들의 크기를 다시 비교해야 한다면 그래프를 그리는 취지가 무색해진다.

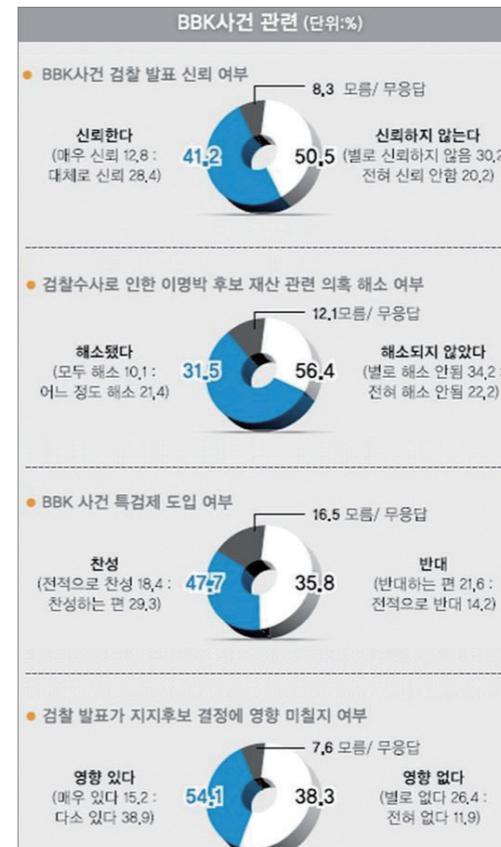
아래의 원그래프는 한 방송에서 소위 ‘김영란법’의



최저임금으로 10시간 일해서 벌 수 있는 돈



김영란법의 국회통과에 대한 여론조사 결과



BBK 사건과 관련한 여론조사

국회통과에 대한 여론조사 결과를 파이 형태로 그린 것이다.⁵⁾

64.0%인 ‘잘했다’ 응답의 크기가 7.3%인 ‘잘못했다’ 응답의 크기와 비슷하게 그려졌다. 심지어 ‘잘못했다’ 7.3%가 ‘모르겠다’ 28.7%보다 더 크게 그려졌다. 퍼센트(%)의 크기에 맞도록 제대로 그린 두 번째 원그래프와 비교하면 처음의 원그래프는 이 법의 국회통과가 마음에 들지 않는다는 왜곡 의도가 쉽게 드러난다.

위 사례와 마찬가지로 %의 크기를 상호비교가 가능하도록 그리지 않은 사례는 적지 않다. 아래의 그림은 BBK 사건과 관련한 여론조사 결과를 그린 것이다.⁶⁾

역시 ‘BBK 사건 검찰 발표 신뢰 여부’에 대한 응답에서 41.2%의 ‘신뢰한다’는 응답이 50.5%의 ‘신뢰하지 않는다’는 응답보다 더 크게 그려졌다. 심지어 그 아래 그래프에서는 31.5%의 크기가 56.4%의 크기보다 거의 두 배나 될 정도다. 역시 어떤 의도에서 이렇게 왜곡했는지를 쉽게 알 수 있는 그래프다.

시각적인 인상과 사실은 크게 다를 수도

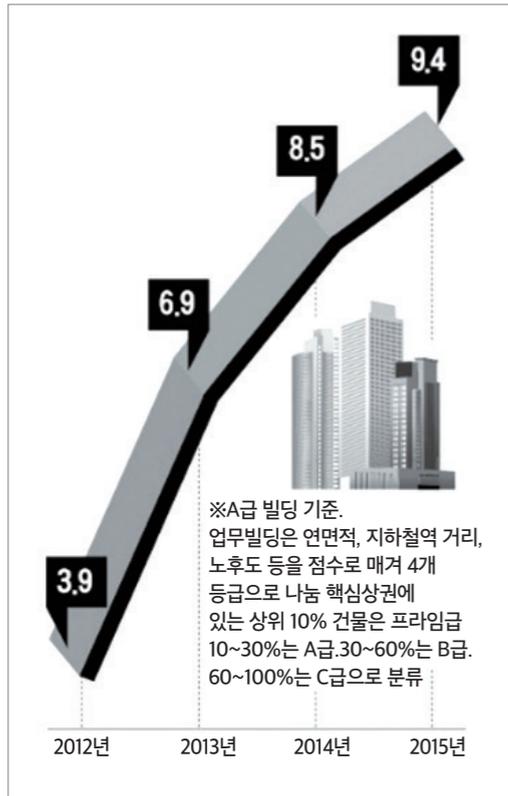
그래프의 시작을 화살표의 밑 부분으로 그리기 시작하면 아무리 소폭증가라도 연장수 마음대로 화살표의 끝을 위치시킴으로써 원하는 만큼 과장해서 표현할 수도 있다. 위의 그림은 강남의 업무빌딩 공실률을 그린 것이다.⁷⁾

2012년에는 공실률이 3.9%였는데 3년 뒤인 2015년에는 9.4%로 2.4배 올랐다. 하지만 선 그래프가 주는 느낌은 그보다는 훨씬 더 공실률이 높아진 인상을 주고 있다.

4) <https://vop.co.kr/A00000496675.html> 5) <https://twitter.com/cozyonu/status/572991122244435968>

6) <http://ppomppu.co.kr/zboard/view.php?id=humor&no=54141>

7) <http://media.daum.net/economic/finance/newsview?newsid=20160121183004721>



강남의 업무빌딩 공실률(단위:%)

아래의 그림은 한 방송국 뉴스에서 정규직 임금체계를 바꾸어야 한다는 논리에 대한 근거로 한국과 유럽 세 나라의 30년 근속 임금표를 그린 것이다.⁸⁾ 한국 정규직의 임금 상승폭이 가장 두드러지게 그려졌다. 예를 들어 영국과 한국의 2배의 차이가 높



한국과 유럽 세 나라의 30년 근속 임금표

이에 있어서는 훨씬 더 큰 차이라는 인상을 주고 있다.

통계학자인 스테픈 캠벨은 “그래프는 진실을 자세히 검사하지 않고는 결코 있는 그대로 받아들여서는 안 된다. 사물은 겉으로 보이는 것과는 전혀 다른 경우가 많다.”고 말했다.

이런 맥락에서 그래프에 대해서는 이렇게 표현할 수 있다. “그래프를 자세히 따져보지 않고는 결코 있는 그대로 받아들여서는 안 된다. 그래프는 겉으로 보이는 것과는 전혀 다른 경우가 많다.” 사실 그래프(그림도표를 포함해서)를 그리는 목적은 열심히 읽으려 하지 않는 독자들과 숫자를 다루는 데 익숙하지 않은 사람들에게 데이터가 갖고 있는 특징이나 본질을 간단명료하게 보여주는 데 있다. 그러나 그래프를 그리는 과정에 있어 수직축과 수평축의 위치나 어떤 눈금을 선택할 것인가는 그래프로 무엇을 나타내려고 하는가 하는 의도에 달려 있다. 더욱이 이러한 선택이 전적으로 그래프를 그리는 사람에게 주어졌으므로 항상 왜곡의 가능성이 존재하는 것이다.

물론 그래프를 그리는 사람의 기술이 부족하거나 그래프에 좀 더 생기를 불어 넣는 과정에서 과장이나 축소가 있을 수 있다. 하지만 자신만의 논조를 부각하기 위해서 의도적으로 과장하거나 왜곡하는 경우가 더 많다. 그렇다면 이러한 왜곡을 막기 위해서는 어떻게 해야 할까? 그래프를 보는 사람과 그리는 사람의 입장으로 구분해 보자.

제대로 이해하기 위한 두 가지 질문

좋은 그래프는 간단하고 정확하게 데이터에 대한 진실을 말해야 한다. 그래프를 보는 사람의 입장에



서 그래프의 공정성을 평가하기 위해서는 두 가지 질문을 던져야 한다.

첫째는 그래프가 전체 그림을 보여주고 있는가 하는 것이다. 그래프의 밑 부분이 잘라져 있는지, 축을 변화시킨다면 전혀 다른 인상을 주는 그래프가 될 수 있는지, 만일 그렇다면 데이터를 왜곡하지 않고 바른 모양을 나타내는 그래프는 어떤 것인가를 반문해야 한다.

둘째는 눈금이 (특히 수직축의) 과장되어 있지 않은가 하는 물음이다. 과장된 눈금은 잘못된 인상을 독자들에게 강요하는 경우가 많다. 그러면 과장되지 않은 적절한 눈금이란 어떤 눈금일까? 그 답은 데이터에 포함된 중요한 차이나 흐름의 변화를 꼭 보여 주어야 하는 동시에 별로 중요하지 않은 것들이 과장되지 않도록 눈금을 정하는 것이다. 더욱이 수직축이나 수평축의 눈금이 무엇을 나타내는지 표시조차 안 된 그래프는 의도적으로 과장하거나 속이기 위한 것이 대부분이다.

그래프를 만드는 사람이 유의할 점은 좋은 그래프를 만들기 위해서는 데이터의 본질에 대한 이해와 미적인(artistic) 기술이 필요하다는 것이다. 지금까지는 주로 미적인 기술에 중점을 두어 재미있고 눈에 잘 띄는 그래프를 그리려 하였지만 더욱 중요한 것은 데이터의 본질을 파악하고 이를 그대로 전달할 수 있는 그래프를 그리는 것이다. 특히 눈금의 크기를 적절하게 정하기 위해서는 그래프를 만드는 사람의 데이터의 본질에 대한 이해와 경험이 필수적으로 요구된다.

구체적으로는 먼저 데이터의 변화를 보여주어야지 눈금이나 축의 변화에 따른 축소/과장을 강조해서는 안 된다. 또한 숫자를 나타낼 때에는 그 숫자의 크기와 직접적으로 비례가 되게 하고 특히 그림도표를 사용하는 경우에는 그 숫자의 차이가 면적이거나 부피의 차이와 같도록 유의해야 한다. 또한 필요한 경우에는 정확하게 말을 덧붙이는(labeling) 것도 좋은 방법이다.

8) https://www.google.co.kr/search?q=%EA%B7%B8%EB%9E%98%ED%94%84+%EC%99%9C%EA%B3%A1&sa=X&biw=1067&bih=521&tbm=isch&tbo=u&source=univ&ved=0ahUKewjEitb7yq_KAhUBxJQKHTnmCyAQsAQIGg

초거대 AI가 쏟아 올린 일상 속 통계활용의 시대

신우람 | 통계청 통계정보플랫폼과 사무관



국제학술지 네이처는 매년 우주과학, 생명공학, 환경, 물리학 등 다양한 과학기술분야에서 뛰어난 업적을 보인 과학자 10명을 선정하여 ‘네이처10’을 발표하고 있다. 2023년에는 인류 최초로 달 남극에 찬드라얀 3호를 착륙시킨 인도 우주연구기구(ISRO)의 칼파나 칼라하스티 박사와 멸종위기종의 번식을 돕는 유전공학기술을 개발한 하야시 카츠키 박사 등이 포함된 10명의 과학자가 선정되었다. 그런데 지난해 ‘네이처10’에는 사람이 아닌 사물이 과학자들과 함께 이름을 올려서 세간의 관심을 받았다. 바로 챗GPT가 포함된 것이다. 이에 대해 네이

처는 챗GPT가 비록 사람은 아니더라도 2023년 과학계에 지대한 영향을 끼쳤고, 학습된 데이터로 대화를 이어가는 것이지만 이를 통해 과학자들이 일하는 방식을 변화시키고 있기 때문에 ‘네이처10’에 선정되었다고 밝힌 바 있다.

전문지식을 학습한 AI와의 대화를 통해 과학자들의 일하는 방식이 변화하고 있다는 네이처의 발표는 시사하는 바가 크다. 새로운 정보를 탐색하고 이와 연관된 종합적인 통찰력을 얻기 위해서 우리는 일상적으로 검색 기능을 사용하고 있다. 웹사이트의

모든 정보를 검색해주는 포털의 검색엔진이나, 국가통계포털에서 통계를 찾아주는 검색 기능도 기본적으로 궤를 같이하는 부분이 바로 키워드 검색이다. 특히 국가통계포털에서는 많은 사용자가 통계에 대해 궁금한 점을 키워드로 정리해서 검색창에 입력하고 검색 버튼을 클릭함으로써 수많은 통계정보의 바다에 풍덩 뛰어들게 된다. 그러나 대부분의 검색 결과에는 내가 원했던 정보뿐만 아니라 굳이 알고싶지 않은 불필요한 내용들까지 검색의 그물망에 함께 끌어올려지는 바람에 내가 필요했던 정보가 맞는지 하나하나 선별하고 검증하는 지난한 과정도 어쩔 수 없이 거쳐야 한다.

국가통계의 성장과 통계이용자의 불편

많은 독자들이 이미 경험했듯이, 국가통계포털에는 1300여종의 통계와 23만개 가까운 통계표가 수록되어 있어 국민들이 시간과 공간의 제약없이 자유롭게 통계를 이용할 수 있다. 그리고 통계의 주제는 매우 다채롭다. 인구부터 보건, 범죄, 환경, 국토 이용, 물가와 금융, 환경에 이르기까지 어떠한 인류의 모든 지식체계와 생활상이 국가통계의 포괄범위에 있다고 해도 과언이 아닐 듯하다. 통계의 주제범

위가 넓다고 해서 각각의 통계의 깊이가 낮은 것도 아니다. 국가통계가 작성되기 위해 통계분류체계를 반드시 사용해야 하고, 용어와 변수도 공식적이고 전문적인 단어와 개념을 사용하고 있다. 또한 통계에 기반한 정책 개발과 평가가 자리 잡는 것과 함께 증거 기반 의사결정 시대가 열리면서 새로운 통계도 지속적으로 개발되고 있다. 그 결과 최근 5년 동안의 국가통계종수는 1,142종에서 1,332종으로 증가하였고, 통계표 수도 145,988개에서 227,158개로 증가하는 등 국가통계의 성장세가 가파르다.

그러나 통계의 전문성이 강화되고 주제범위도 다양해지는 등 통계정보 제공의 질적, 양적 성장이 강조되는 장면의 이면에는 통계를 이용하고자 하는 일반 국민이 통계 이용에 어려움을 호소하는 목소리도 함께 커지고 있는 것도 현실이다.

2023년 일반 통계이용자 5,488명을 대상으로 통계정보서비스 이용 시 문제점을 묻는 설문조사 결과를 살펴보면, 통계정보를 이해하는데 어려움을 겪는다고 답한 비율이 36%로 가장 높았고, 그 다음으로 통계검색이 어렵다고 답한 비율이 35%를 차지했다. 쓸모있는 통계를 많이 제공한다고 해서 국민이 통계를 편하게 이용하는 것은 아니라는 의미



다. 결국 쉬운 용어로 통계를 설명하고 통계에 접근하는 방식도 더 편리하게 바뀌어야 한다는 결론에 다르게 된다.

규칙 기반 챗봇의 등장과 한계

통계이용자가 보다 쉽게 원하는 통계정보까지 접근할 수 있도록 이미 국가통계포털에서는 2021년부터 챗봇 서비스를 실시하고 있다. 민간기업과 공공기관을 포함해 이 시기에 만들어진 대부분의 챗봇은 규칙 기반 챗봇의 모습을 띄고 있다. 규칙 기반 챗봇이란 사전에 미리 정의된 다양한 규칙과 응답 집합에 따라 작동되는 챗봇을 의미한다.

즉 예상되는 질문을 미리 만들고, 이에 대응하는 답변도 서로 쌍을 이루도록 잘 정비된 시나리오에 따라 움직이면 최상의 성능을 보여준다. 그런데 사전에 서로 합의된 질문과 응답의 범위에서 조금만 벗어나면 제대로 작동되지 않을 가능성이 높다. 특히 정확한 용어를 사용해야 답을 준다는 점이 큰 한

계로 지적된다. ‘가계가처분소득’이나 ‘비임금근로자’, ‘합계출산율’ 같은 전문적인 통계용어를 정확히 언급하면서 질문해야 답을 들을 가능성이 높다. 실제로 현재 국가통계포털에서 서비스 중인 코봇에 2023년의 출생아 수가 몇 명인지 질문하면 기특하게도 인구동향조사의 통계표로부터 출생아 수가 23만명이라는 정확한 답을 찾아서 제시해 준다. 출생아 수를 물어보는 질문에는 인구동향조사에서 답을 찾도록 하는 ‘규칙’이 제대로 작동하고 있기 때문이다.

그런데 이 규칙에 조금만 변조를 주면 어떻게 될까? 출생아 수 대신 우리가 일상생활에서 사용하는 아주 편안한 대화의 방식을 가정해보자. ‘2023년에 사람들이 아이를 얼마나 많이 낳았어?’라고 질문하면 코봇은 인구동향조사 대신 엉뚱하게도 한부모가족실태조사나 지자체의 사회조사 결과를 제시해 주며 슬그머니 물러난다. 무슨 질문을 하는지 제대로 이해하지 못했기 때문이다. 만약 일상언어를 더 잘 이해하는 시라면 어땠을까?



통계정보 접근에 있어 초거대 AI의 역할

초거대 AI를 구성하는 핵심요소는 대규모언어모델로, 수천억개의 매개변수를 포함하고 있기 때문에 자연어의 이해능력이 뛰어나고 동시에 학습데이터로부터 생성한 정보를 쉬운 단어로 풀어서 설명하는 능력 또한 탁월하다.

즉 질문자가 통계에 대한 사전지식이 풍부하지 않아도 사회·경제 현상 등 우리 주변의 여러 현상에 대해 일상언어로 질의하면 초거대 AI를 품은 통계 챗봇은 그 의도와 맥락을 파악하여 통계수치나 용어를 해설해준다. 유사한 용어가 있다면 그 차이점도 알려줌으로써 국민 눈높이를 맞추어 정보를 제공할 수 있어 어려운 통계에 대한 문턱이 대폭 낮아지게 될 예정이다. 예를 들어 ‘가정에서 마음대로 쓸 수 있는 돈’이라거나 ‘자영업자나 돈을 받지 않고 일하는 가족’, ‘여성 한 명이 평생 낳을 것으로 예상되는 아이 수’라는 식으로 표준적인 통계용어가 아닌 일종의 개념을 들어서 질문하면 그 의미가 무엇인지 이해하고 답변을 찾아 준다.

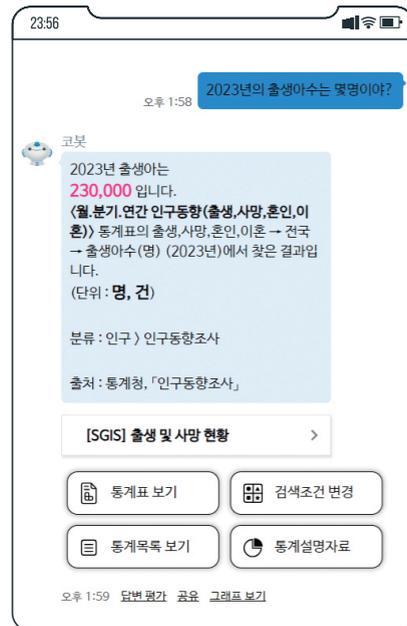
초거대 AI의 우려점과 대안

다만 초거대 AI는 장점만 있는 것이 아니다. 정확한

답을 찾지 못할 경우, 방대한 학습 내용 중에 비슷한 부분만 찝기기해서 잘못된 정보를 주는 이른바 환각 현상이 발생할 수 있다. 초거대 AI의 환각 현상을 꼬집는 대표적인 사례로 세종대왕이 한글 창제 과정에서 집현전 학자들을 꾸짖던 도중 그만 맥북을 집어 던졌다는 일화나 허쉬 초콜릿은 한국전쟁 이후 미국으로 건너간 허쉬 부자에 의해 탄생되었다는 등의 황당한 이야기가 회자된다. 초거대 AI에 통계에 대해 질문해 보면 어떤 답변을 하게 될까? 현재 서비스되고 있는 초거대 AI에 의사들이 평균적으로 아이를 얼마나 많이 낳는지 볼 수 있는 통계가 있는지 물었더니 합계출산율 감소에 대한 일반적인 현황과 평균 초산연령에 대한 정보를 알려주고 있다.

다른 초거대 AI 서비스는 인구주택총조사 결과를 언급하면서 남/여 의사의 자녀수를 각각 1.76명, 1.45명으로 그럴듯하게 제시하고 있다. 사실 질문을 교묘하게 했을 뿐으로 의사의 평균 자녀수가 조사된 공식 통계는 실제로 존재하지 않는다. 그럼에도 불구하고 그러한 통계를 찾지 못했다는 정직한 답변을 하는 대신, 두루뭉술하게 말을 돌리거나 없는 정보를 억지로 만들어 내는 것은 생성형 AI를 활용함에 있어 걱정스러운 요소이다.

새로운 통계 챗봇은 이러한 환각 현상을 최소화



‘출생아수’ 질의에 대한 코봇의 답변



‘출생아수’ 대신 자연어를 사용했을 경우의 답변

하기 위해 검색 증강 생성서비스(RAG: Retrieval Augmented Generation)를 적용한다. 언어를 이해하고 표현하는 능력은 인간의 검증된 초거대 AI 언어모델을 활용하되, 실제로 정보를 검색하고 답변을 생성하는 부분은 전문적인 문서들이 고차원적인 벡터로 저장된 통계지식데이터베이스에서만 찾아오는 방식이다.

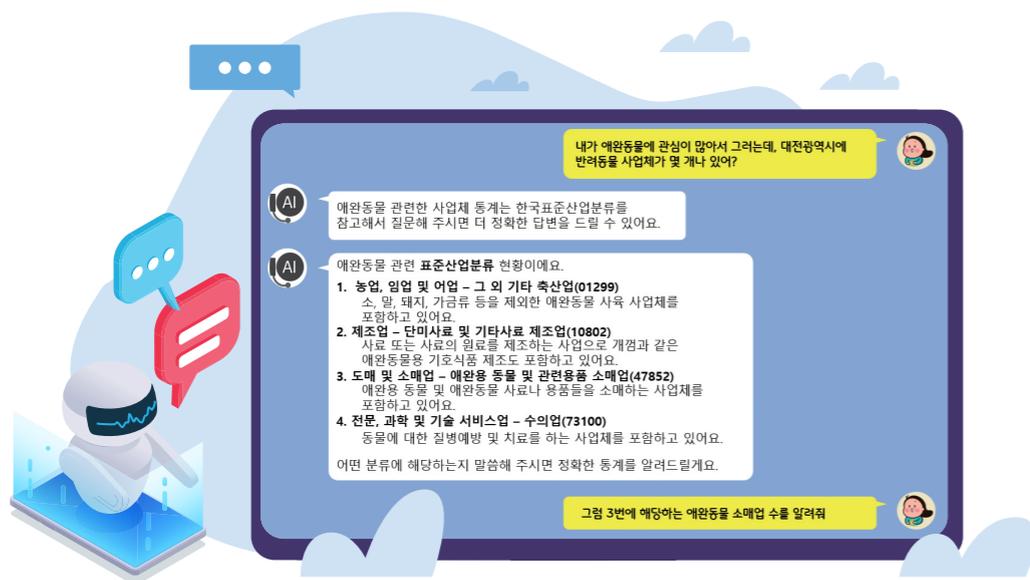
이 데이터베이스에는 1300여종의 통계에 대한 메타데이터와 설명자료, 정보보고서 등이 저장된다. 주제별 통계를 가장 상세하게 설명하고 특징들까지 해석한 보도자료나 간행물도 함께 포함한다. 이때 사용되는 학습데이터들은 이미 통계청 누리집이나 국가통계포털 등을 통해 완전히 공개된 자료라는 공통점이 있다. 공개된 데이터로부터 정보를 찾는 방식이므로 혹시 있을지 모를 데이터 유출사고나 개인정보가 노출되는 문제상황으로부터 완전히 자유로울 수 있다는 의미이다.

결과적으로 통계이용자들은 종합적인 통계정보를 접할 수 있고, 무엇보다 우리가 일상에서 흔히 사용하는 단어와 문장으로 설명을 들을 수 있게 된다. 필요할 경우 정보를 요약하여 핵심만 간추릴 수 있다

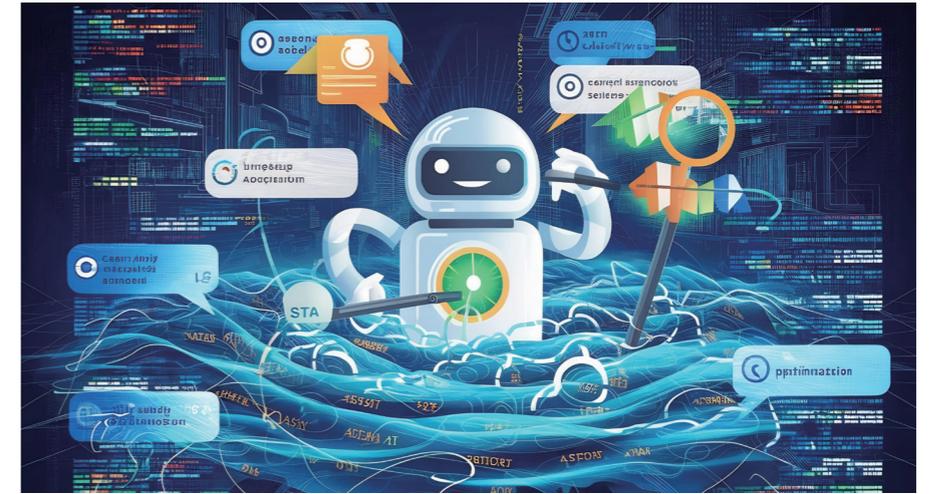
는 점도 장점이다.

초거대 AI 통계챗봇은 통계분류에 대해서도 학습해 나갈 예정이다. 앞서 언급한 질문의 예시처럼, 의사의 평균적인 자녀수를 알 수 있는 통계는 없다. 그러나 만약 초거대 AI가 통계분류를 이해하고 있다면, 의사의 평균 자녀수와 관련된 통계는 없지만 표준 직업분류 상 의사가 포함된 ‘전문가 및 관련 종사자’의 출생자녀수 통계표를 대안으로 제시할 수 있게 된다. 물론 의사가 포함된 상위분류이므로 활용에 주의하라는 언급도 함께 제공한다.

이용자와 대화도 가능하도록 설계된다는 점도 새로운 통계챗봇만의 차별점이다. 만약 통계이용자의 질문 내용이 모호하거나, 지나치게 광범위한 내용을 질문한다면 통계챗봇은 역으로 질문의 의도를 재확인함으로써 질문자가 궁금한 내용이 무엇인지 정확히 파악한다. 예를 들어서 반려동물 사업체수를 질문한다면 통계챗봇은 표준산업분류를 제시하면서 농업, 임업 및 어업에 해당하는 애완동물 사육 사업체 수를 묻는 의도인지, 아니면 애완동물을 소매하는 도매 및 소매업에 대해 묻는 질문인지를 역



이용자와의 대화를 통한 질의내용 구체화 과정



으로 물어봄으로써 의도와 맥락을 정확히 파악하여 답변을 제공하는 식이다.

초거대 AI 통계챗봇이 가져올 미래모습

이처럼 혁신적인 통계접근방식은 2024년 5월부터 11월까지 서비스 구축이 추진된다. 올해 연말에 시범운영을 통해 성능을 확인하고 개선사항을 보완한 후 2025년 1월부터 본격적인 대국민 서비스가 시작된다. 또한 통계청은 초거대 AI의 기술발전 추이를 지켜보면서 앞으로는 더 다양한 영역에 초거대 AI를 접목할 예정이다.

2026년부터 시작될 예정인 ‘통계정보플랫폼 및 원포털 구축사업’이 완성되는 2028년부터는 초거대 AI가 통계표의 수치 해석을 돕고 이용자의 통계분석까지 지원하게 된다. 공간정보와도 접목해서 분석 결과를 통계지도로 변환해주고, 데이터 특성을 가장 잘 표현하는 형태로 그래프를 그려주는 등 시각화 영역까지 확장될 예정이다. 이 단계까지 도달하면 국민들은 초거대 AI 통계챗봇과의 대화를 통해 일상생활에 가까이 있는 의사결정까지 활용할 수 있게 된다. 예를 들면 ‘대전시 내에서 이사를 가려는데 최근 5년간 집값이 덜 올랐으면서 주변에 학교가

많은 동네를 추천해줘’라고 챗봇창에 입력하면 초거대 AI가 인구주택총조사, 주택매매동향조사, 전국사업체조사 등의 통계자료를 공간정보와 연계하여 분석하고 가장 적합한 후보지를 추천하는 것도 가능해진다.

초거대 AI는 이제 막 걸음마를 시작했다. 문장을 이해하고 정보를 찾아서 제시해주는 수준을 넘어 앞으로는 소리와 영상까지 이해하여 추론할 수 있게 된다. 더 시간이 지나면 통계를 비교하여 분석하고, 자료 간 유사성과 차이를 통찰력 있게 제시해주는 것은 물론, 나아가 인류의 일하는 방식까지 바뀌는 것으로 전망된다. 사람이 할 일들을 AI가 대체하는 시대에 대한 두려움과 우려도 일각에서는 가지고 있을 것이다. 그렇지만 다가올 가까운 미래에서는 상황판단과 의사결정에 꼭 필요한 통계를 배경 지식과 경험이 풍부한 전문가들만 쓰는 시대는 아닐 것이다. 누구나 쉽게 필요한 통계를 찾아서 이해하고, 분석을 통한 시사점까지 도출해내는 이 통계 활용과정이 국민의 일상생활에서 움틀 수 있기를 바란다. 통계에 대한 문턱을 아주 낮추기 위한 긴 레이스의 출발선에 서 있는 초거대 AI 통계챗봇의 활약에 기대를 걸어봐도 좋겠다.

통계를 누구나 쉽고 재미있게 즐길 수 있는 KOSIS 통계시각화콘텐츠

강명규 | 통계청 통계서비스기획과 사무관



AI, IoT, 로봇 등으로 대표되는 ICT 기술의 융합은 우리 사회를 눈부시게 발전시키고 동시에 세상을 너무도 빠르게 변화시키며, 미래를 예측하기 어렵게 하는 불확실성을 가져왔다. 이렇게 빠르게 변화하며 불확실한 상황 속에서, 데이터에 기반한 과학적 의사결정도구로서 “통계”는 그 필요성과 중요성이 더 높아지게 되었다. 통계청은 이렇게나 중요한

통계를 생산하는 일과 함께, 국민 누구나 언제든지 통계를 사용할 수 있도록 통계를 제공하는 역할 또한 맡고 있다. 24년 3월 현재 기준 1300여종, 22만여개의 통계표를 국가통계포털(이하 KOSIS)에 수록하여 관리한다. 통계의 중요성과 높은 접근성에 힘입어, KOSIS를 통해 수많은 국민이 통계를 이용하며, 그 이용건수 또한 꾸준히 증가하고 있다.

방대하고 이해하기 어려운 통계? 지표 중심의 통계시각화콘텐츠!

수많은 통계표를 담고 있는 KOSIS에 대해 ‘원하는 데이터를 찾기 쉽지 않다’, ‘문자 위주의 통계표는 이해하기 어렵다’와 같은 의견들이 지속적으로 제기되어왔다. 특히 통계 이용 경험이 적은 일반 이용자 입장에서 방대한 통계데이터 중 원하는 정보를 찾고 그 변화나 추이를 파악하고 이해하긴 더욱 어려웠다. 통계 이용률의 증가와 함께 국민이 통계를



[그림1] KOSIS 통계표 이용건수 및 방문자수

더욱 잘 활용하고 이해하도록 도울 방법의 필요성이 대두되었다.

그래서 통계청은 국민 누구나 원하는 통계를 쉽게 찾고 잘 활용할 수 있도록 다양한 통계의 핵심내용(지표)을 시각적으로 표현한 통계시각화콘텐츠를 다양하게 제공하고 있다. 각각의 통계시각화콘텐츠는 인구·경제·국제 등의 다양한 주제에 대해 이용자 참여요소, 인포그래픽 등의 재미요소를 더해 통계에 대한 거리감을 줄이고, 더 나아가 통계 이용 저변을 확대하는데 기여하고 있다.

통계시각화콘텐츠와 함께 KOSIS 200% 활용하기

통계표 중심의 KOSIS를 이용하는데 약간의 어려움이 있다면 통계시각화콘텐츠를 함께 활용함으로써 통계에 대한 접근성과 이해도를 높일 수 있다. 가령 이용자가 관심있는 테마의 통계시각화콘텐츠를 선택해서 손쉽게 지표를 찾고 해당 지표의 원통계표를 통해 추가적인 정보탐색을 할 수 있다. 또한 각 지표는 시각적 요소로 표현되어 그 추이변화를 누구나 손쉽게 확인할 수 있다. 여기에서는 대표적인 통계시각화콘텐츠 3종을 소개해보려 한다.



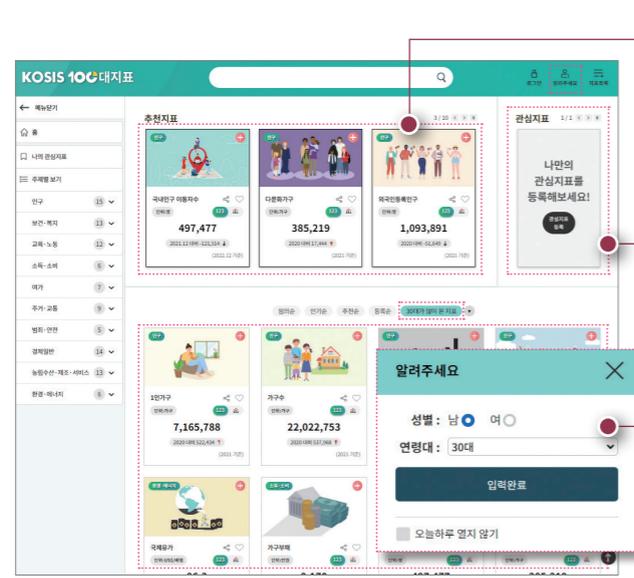
현재 서비스 중인 통계시각화콘텐츠

- 1 ●● **KOSIS 100대 지표**
국민적 관심을 반영한 100개의 지표 제공
 - e-지방지표(시각화)
각 지역별 세부지표 제공
 - 통계로 시간여행
과거의 통계를 통하여 과거의 모습을 살펴보는 콘텐츠
 - 해·석·남·녀
남녀 간 차이를 통계를 통해 탐구
- 2 ●● **인구로 보는 대한민국**
인구 관련 지표를 통해 인구정보를 쉽고 재미있게 체험
 - 나의 물가 체험하기
개인별 소비패턴에 따른 나의 물가를 산출하고 이를 공식물가와 비교해 보는 콘텐츠
 - 통계로 보는 자화상
통계, 앙케이트 등을 활용하여 본인에 대한 인포그래픽 생성
 - 지역경제상황판
지역별 주요 경제 지표 비교
 - 경기순환시계
주요 경제지표들의 경기순환 흐름 제공
 - 3D 인구 체험
가상현실로 구현한 인구 쿼츠
 - 버블차트로 보는 통계
경제와 사회의 연관성을 시각적으로 제공
- 3 ●● **세계속의 한국**
전 세계의 취합된 통계정보제공
 - 세상을 그리는 통계
다이나믹차트(통계동영상) 제공
 - 통계웹툰
통계를 이용한 웹툰 제작/매월 1회
 - 카드뉴스
최근 이슈를 통계를 통해 해석/매월 1회
 - 내가 말하는 통계
디지털콘텐츠 공모전 입상작 게시

01 KOSIS 100대 지표



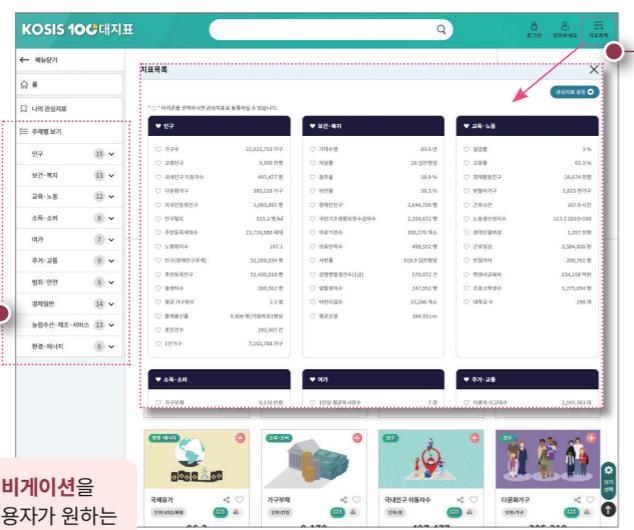
우리나라의 전반적인 상황을 다양한 주제별로 폭넓게 살펴보고 싶다면 ‘KOSIS 100대 지표’를 이용할 수 있다. ‘KOSIS 100대 지표’는 국민적 관심과 영역별 대표성을 고려하여 선정된 100개의 지표를 그래프 등의 시각적 요소로 표현하여 누구나 이해하기 쉽다. 뿐만 아니라, KOSIS 이용실적과 지표 선정 연구를 통해 사회변화가 반영된 지표들은 이 사용자에게 시의성있는 정보를 제공할 수 있다.



텍스트를 덜고 '지표카드' 형식을 이용하여 그래픽 직관성 강화

관심지표 등록을 통해 자주 이용하는 지표들만 모아 저장

'알려주세요' 창을 통해 이용자의 성별, 나이를 입력받아서 동일 집단의 이용자가 많이 본 지표를 추천



우측 상단 '지표목록'을 통해 모든 지표에 대해 한 눈에 파악

좌측 내비게이션을 통해 사용자가 원하는 지표에 쉽게 탐색

01



각 지표의 값을 시각적으로 도표로 제공하고, 시점조정, 인쇄, 공유, 원통계표 링크 등 다양한 사용자 편의 기능 제공

해당 지표와 관련성이 높은 지표를 '연관지표'로 제공하여 이용자의 통계 탐색의 기회를 확대

02 인구로 보는 대한민국



고령화·저출생 등 우리나라 인구 현상을 통계로 확인해보고 싶다면 ‘인구로 보는 대한민국’을 이용하는 것도 좋은 방법이다. 방대한 인구 통계자료를 142종의 지표로 압축적으로 제시하며, 우리나라 인구의 과거·현재·미래를 한 눈에 살펴볼 수 있는 인구피라미드, 인구 지표와 사회·경제 지표간 연관 관계를 살펴볼 수 있는 관계도맵 등 다양한 시각적 요소를 적용하였다. 인구 관련 사회이슈에 대한 관심이 높아진 요즘, 인구 변화와 그로 인한 영향들을 보여준다.



인구에 대한 다양한 질문 (클릭시 해당 질문에 대한 지표로 이동)과 애니메이션 효과로 시작하여 이용자 흥미유발

PC·스마트폰·태블릿에서도 편하게 활용할 수 있도록 반응형웹을 적용하여 모바일 접근성 강화

① 인구상황판



우리나라 인구관련 정보를 시점별로 대시보드를 통해 한 화면에서 한 눈에 조망

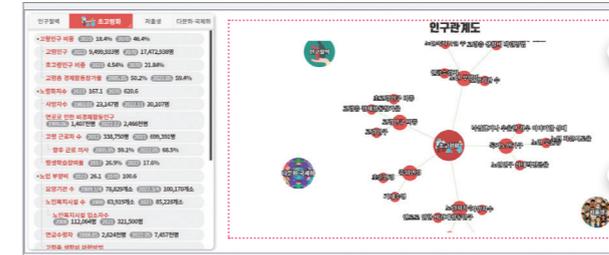
기존 13종의 인구지표를 총 142종으로 확대하여 더 다양한 정보 제공



출산율, 기대수명 등과 같은 인구변동요인설정을 통해 각 설정값과 시간에 따른 인구피라미드의 변화 관찰

02

② 인구이야기



지표간의 관계에 주목하여 관련 지표들을 관계도맵으로 제공

움직임 기능 추가



4가지 대주제에 관한 모든 지표에 대해 설명과 차트 제공

③ 인구더보기



블록형

우리나라 인구와 관련된 지표들을 다양한 방법으로 제공



양자비교



러닝차트 (시계열의 흐름에 따라 차트 변화)

03 세계속의 한국



‘세계속의 한국’은 전 세계의 국가별 통계정보를 누구나 손쉽게 검색하고 비교하여 국제 사회에서 우리나라의 위상을 확인할 수 있는 콘텐츠이다. 복지·소득·사회·인구 등 다양한 분야의 국제통계를 한 곳에 모음으로써 각 국가의 전반적인 통계정보를 확인하고 더 나아가 국가 간의 비교를 통해 한 단계 더 깊은 통찰을 얻을 수 있다.

‘지표별보기’, ‘국가별보기’, ‘비교해보기’ 세 가지 테마로 통계정보 제공

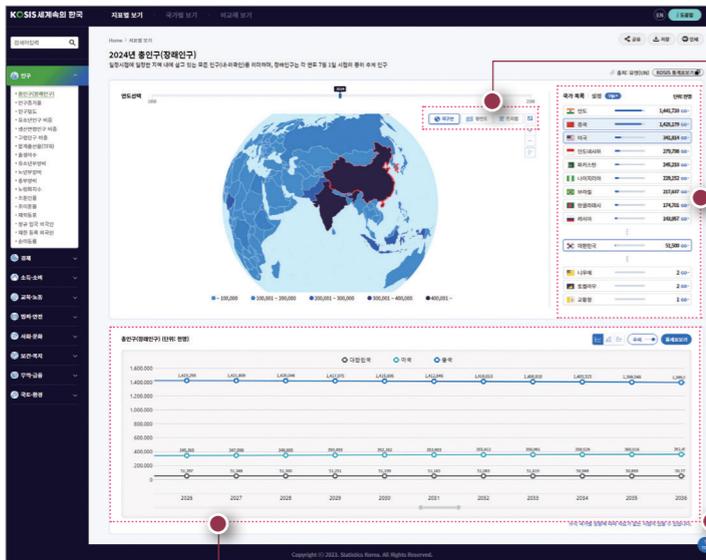


영문버전 제공

조회수를 바탕으로 주요 국제통계 지표에 대해 우리나라 수치 제시

9개 영역 87개 지표

① 지표별보기



지구본·평면도·트리맵 등 다양한 형태로 각 지표 탐구

수집된 국제통계를 바탕으로 각 지표에 대해 전 세계 현황을 한 눈에 조망

다양한 그래프 모형 선택 가능

설정버튼을 통해 원하는 국가목록 구성 후 클릭을 통해 하단에 간단한 국가비교 기능 제공

03

② 국가별보기



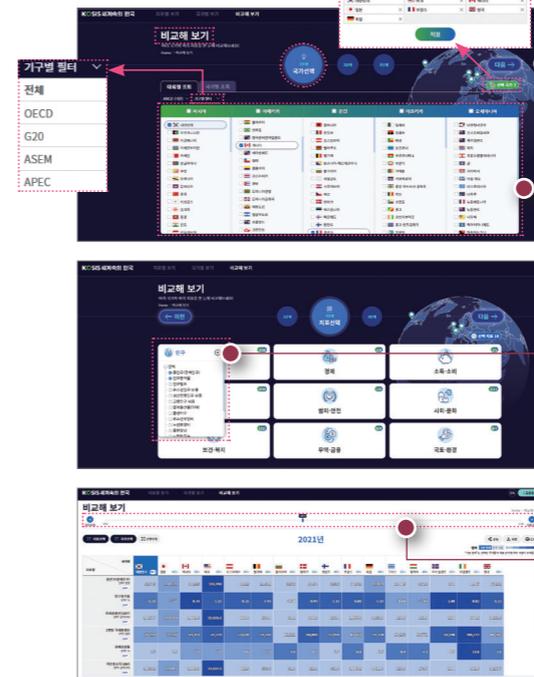
각 통계표별로 분산된 통계정보를 국가별로 한데 모아 제공함으로써 각 국가의 전반적인 현황 파악 용이

카드형: 각 지표별 추세선과 상세보기를 통해 연도별 변화·상세값 제공

수집된 통계가 있는 모든 국가들에 대해 선택 가능

목록형: 각 영역별로 국제통계지표 제공

③ 비교해보기



다종의 국가와 지표에 대해 연도별로 비교할 수 있는 코너

1 관심있는 국가 선택 대륙별로 기본 분류, 필터를 통해 OECD 등 기구별 구분 가능

2 비교하고 싶은 지표를 영역별로 여러개 선택

3 결과화면 시점변경 등의 기능과 함께 통계청이 수집하여 보유한 값에 대해 결과 제시

국민의 소중한 답변이 모여 만들어진 통계는 그 잠재적인 가치가 무궁무진하다. 하지만 아무리 유용한 통계라도 제대로 전달될 수 없다면 그 가치는 제한적일 수 밖에 없을 것이다. 이때 전문적이고 명확한 시각화와 직관적인 설명을 담은 통계시각화콘텐츠는 통계를 전달하는 훌륭한 수단이 될 수 있다. 통계 분석 능력을 갖춘 전문가가 뿐만 아니라, 누구나 통계를 쉽게 접하고 활용할 수 있는 통계시각화콘텐츠를 통해 더 많은 국민이 통계의 가치를 누릴 있도록 지속적으로 노력하겠다.

통계로 보는 올해 ‘똑똑해지는’ 부자 키워드 7

이승주 | Newsis 기자



통계를 보면 현재 경제상태가 보이고, 아는 만큼 똑똑하게 준비할 수 있습니다. 올해 1분기 통계를 살펴보니 가슴이 답답합니다. 내 월급통장 빼고 다 오른다는 말처럼 물가는 오르는데 차곡차곡 모은 저축액은 제자리 걸음이죠. 올해 똑똑하게 가게 살림을 운영하려면 어떻게 해야 할까요. 정답은 바로 통계에 있습니다. 올해 1분기 경제 통계를 관통한 키워드 7개를 살펴보겠습니다. 이를 기반으로 올해 똑똑하게 살림살이를 꾸려볼까요?

1
환율

**환율 1400원 돌파...
역대급 고회환율에 물가 더 오르나?**

올해 환율이 올랐습니다. 환율이란 우리나라에서 쓰는 원화가 세계에서 어느정도 가치를 지니는지

보여주는 척도인데요. 특히 기축통화로 여겨지는 미 달러를 기준으로 정해집니다. 환율을 보면 1달러가 우리 돈으로 얼마나 가치를 지니는지 알 수 있어요. 환율이 올랐다는 것은 1달러를 가지려면 더 많은 원화가 필요하다는 뜻이니, 반대로 우리 원화의 가치는 떨어졌다는 뜻이죠.

그런 환율이 올해 꿈틀대더니 4월부터 오르기 시작해 14일에는 급기야 1400원도 넘어섰습니다. 물론 그날 하루 중 잠깐 1400원을 넘어선 것에 불과하지만 이를 간과해선 안 됩니다. 환율이 1400원을 넘긴 것은 역대 세 번에 불과하기 때문이죠. 심지어 1997년 국제통화기금(IMF)사태라 불리는 외환위기, 2008년 서브프라임 모기지 사태가 촉발한 글로벌 금융위기, 2002년 레고사태 등 우리 경제가 어려워졌을 때거든요.

그렇다보니 이번 1400원을 넘어선 고회환율 기조를



하나은행이 발표한 원달러 환율의 3개월 간 변화

두고 우려가 나옵니다. 이번에 환율이 오른 이유는 ‘중동사태’ 때문인데요. 지난해 10월 이스라엘과 팔레스타인의 분쟁이 난 뒤 그 양상이 지난 4월부터 확대되기 시작했거든요. 위기 상황이라고 느낀 전세계 사람들이 달러를 비축해놓으려고 사재기를 하기 시작했고, 고회환율이 됐다 설명이예요. 고회환율이 얼마나 갈 것이냐, 이게 중요해졌죠. 잠시 1400원을 넘어선 것이라면 크게 문제가 없을텐데요. 금방 잦아들 줄 알았던 러시아-우크라이나 전쟁도 계속되잖아요. 중동 사태 역시 장기간 이어진다면 고회환율 기조가 계속될 겁니다. 고회환율이 꽤 오랜시간 이어진다면 우리는 어떻게 될

까요. 가장 우려해야 할 점은 ‘물가’예요. 지금도 물가가 올라 가슴을 치는 분들 많을거예요. 생선이란 과일 값이 왜이리 올랐는지 시장보기 두렵다는 분들도 많죠. 그런데 이보다 물가가 더 오를 수 있거든요.

우리나라는 꽤 많은 것들을 해외에서 수입해오거든요. 그런데 환율이 높아 우리 원화 가치가 떨어졌다는 것은 즉 해외에서 무언가 사올 때 더 많은 돈을 내야 한다는 뜻입니다. 기름 한 방울 안나오는 우리나라에서 석유를 수입할 때 돈이 더 많이 들테니, 이를 이용해 물건을 파는 우리 기업들도 물건값을 올릴 수 있거든요. 이를 포함 광물 등 다른 자원들의 수입 비용도 오르면서 국내 물가가 치솟을 수 있습니다.

실제로 이미 1분기부터 우리 물가는 꿈틀거리고 있어요. 한국은행이 발표한 ‘2024년 3월 수출입물가지수’에 따르면 올해 3월 수출입물가지수는 137.85(2015=100)로 전월(137.24)대비 0.4% 올랐거든요. 1월 이후 수입물가는 계속 증가했고요. 그러니 우리는 혹여 있을 고물가 위기에 대비해야 합니다. 이런 상태가 계속된다면 내 월급 빼고 다 올랐다는 말이 재현될 수 있어요. 내 월급으로 야무지

게 지출하는, 생활비를 절약할 수 있는 방법을 찾는 게 필요하죠. 이 때를 대비해 미리미리 여윌돈을 마련해두는 것도 도움이 될 거예요~

2
수출
**수출 역대치 목표 7000억달러...
수출 활력에 경제 살아날까?**

물가가 오르면서 삶이 팍팍해질 수 있다니 앞이 컴컴한가요. 하지만 예단하긴 이릅니다. 우리 경제가 물가도 비싸고 전반적으로 침체됐는지 몰라도, 이를 회복할 카드가 있거든요. 바로 '수출'입니다. 우리나라에서 수출은 참 중요합니다. 유럽이나 동남아처럼 관광자원이 풍부하거나, 중동처럼 원유가 나는 것도 아니라 부를 창출할 방법은 많지 않지만요. 우리 기업들이 열심히 해외로 수출하면서 벌어들이는 돈이 꽤 되거든요.

다만 문제는 승승장구하던 우리 수출 실적이 지난 2년 코로나19를 겪은 뒤 고전을 면치 못했다는 건데요. 심지어 무역수지도 역대급 적자를 기록하면서 그들이 드리워졌죠. 하지만 우리 기업들과 정부 등이 합심해 으쌰으쌰 힘을 모은 결과 지난해 말부터 수출 실적은 회복되기 시작합니다. 특히 올해 1분기에는 꽤 많이 회복됐네요. 산업통상자원부가 발표한 수출입 통계에 따르면 올

해 1분기에 1637억 달러(220조291억원)를 수출했습니다. 이게 어느 수준인지 감이 안오시나요? 지난해 1분기 보다 8.3% 증가했어요. 물론 작년에 워낙 저조했다 보니 그에 비교해 좋아진게 뭐 대수냐고 반문하실 수 있어요. 중요한 건 암울했던 시간을 지나 점차 회복되고 있다는 거예요.

그 비결이 무엇이나. 바로 우리 수출의 효자인 '반도체' 덕분인데요. 코로나19로 어려웠던 반도체 업계에 생기가 돌면서 수출 실적도 덩달아 오른 겁니다. 올해 1분기 수출 실적을 자세히 살펴보니 반도체(50.7%)·자동차(2.7%)·일반기계(0.9%)·석유제품(3.8%)·선박(64.3%) 등 10개 품목이 고르게 증가했네요.

4월 수출 실적도 증가할 것으로 예상되는데요. 이를 시작으로 2분기를 지나 하반기까지 우리 수출 회복세는 계속될 것으로 보고 있어요. 어느 정도 좋아질 것 같냐고요? 미래를 예측하기란 쉽지 않지만 이를 주관하는 정부부처인 산업부의 발표를 보면 어느정도 가능성은 할 수 있습니다. 산업부가 올해 수출 목표로 역대 최고치를 넘어서는 7000억 달러를 내세웠거든요. 물론 목표치이긴 하지만 그만큼 이전보다 수출이 크게 회복될 여지가 있다고도 볼 수 있겠죠. 그럼 수출 실적이 좋다면 무엇이 좋을까요. 우선 우리가 다니고 있는 기업들이 적자 등을 겪지 않을 테니 여러분은 좀 더 안심하고 회사를 다니셔도 되고

요. 우리나라에서 고물가 등으로 경제가 어려울 수 있는데 반대로 해외에서 돈을 많이 벌어드는 격이니 경제 침체 위험이 줄어든다고도 볼 수 있겠죠. 그리고 개인적으로 투자하시는 분들이라면 구체적으로 수출 실적이 좋아질 업종을 살펴보는 것도 도움이 될 수 있겠죠? 올해 수출에 주력하는 기업과 업계 중에 어디가 크게 회복할지 지켜본다면 여러분의 투자 방향도 잡힐테니까요.

3
유가
**중동사태에 유가 또 100달러 전망...
기름값 치솟을까?**

수출이 날개를 단다고 하니 갑자기 마음이 편해지셨나요? 정부에서 목표치를 높이며 기대감을 키우고 있다고 안심해선 안 됩니다. 그 와중에 기름값이 꿈틀대고 있거든요. 우리나라는 기름 한 방울 안나오는 나라라 보니 석유를 비롯해 천연가스 등 에너지를 수입합니다. 그러다 보니 바로 이 유가 등이 오르면 정말 치명적이죠.

앞서 말씀드렸듯 중동에서 분쟁이 점차 심화하면서 우려가 커지는데요. 그 중에 하나가 유가가 치솟으면 어쩌냐는 겁니다. 왜냐하면 중동은 대표적인 산유국이거든요. 자칫 중동 국가들이 기름을 제때 넉넉히 팔지 않으면 유가가 오르게 될테니까요. 실제



로 이런 상황을 우려한 세계은행(WB)에서 올해 중동사태가 심화한다면 국제 유가가 치솟을 수 있다고 전망했습니다. 배럴 당 100달러를 넘어설 수 있다고 했는데, 현재 80달러대라는 점을 고려하면 꽤 부담이 될 수준이죠.

다만 과도한 우려는 금물입니다. 국내 기관인 에너지경제연구원에서는 현재 수준인 85달러 대를 유지할 것 같다고 관측했거든요. 그렇다면 뭐가 맞는 말일까요. 바로 이 중동사태가 얼마나 확대되는지가

'24년 1Q 수출입 실적 (통관기준 잠정치) (단위 : 억 달러, %, 관세청)

구분	'23년				'24년			
	1Q	2Q	3Q	4Q	1Q	1월	2월	3월
수출	1,512 (Δ12.8)	1,558 (Δ12.0)	1,571 (Δ9.7)	1,681 (+5.7)	1,637 (+8.3)	548 (+18.2)	524 (+4.8)	566 (+3.1)
수입	1,740 (Δ2.2)	1,592 (Δ13.2)	1,507 (Δ21.6)	1,586 (Δ10.7)	1,548 (Δ11.1)	544 (Δ7.9)	481 (Δ13.1)	523 (Δ12.3)
무역수지	Δ228	Δ35	+64	+96	+90	+4	+43	+43



두바이유



2월부터 4월까지 3개월 간 두바이유 기준 국제유가 추이. 뉴욕상업거래소(NYMEX)에서 거래되는 두바이 선물가격으로 산출(계약단 위: 1,000배럴)



관건이 될 겁니다. 이들 국가가 다투던 끝에 비정상적으로 원유를 크게 줄여 팔거나 이를 중단시키는 일이 벌어진다면 과거 유가가 너무 올랐던 '오일쇼크'가 재현될 수 있다는 겁니다.

그렇다면 어떤 상태에서 유가가 쇼크 상태까지 치솟게 될까요? 에너지경제연구원은 이란이 어떻게 나오는지 주목하고 있습니다. 원래 다투던 이스라엘과 팔레스타인은 산유국이 아니었거든요. 이 분쟁이 산유국인 이란까지 확대된 만큼 이 이란의 행동에 따라 유가가 오를 수 있다는 설명입니다.

이란이 수탈했다면서 갑자기 석유 수출을 줄이거나 차단할지 살펴야 합니다. 이란의 석유가 90% 넘게 중국으로 가고 있는데 이를 줄여버리는 건 아닌지, 다툼이 호르무즈 해협을 봉쇄하는 수준까지 되는 건 아닐지 주시하세요. 그 때 '최악의 시나리오'로 여겨졌던 100달러 넘어서는 고유가가 현실화될 테니까요.

물론 이란이 이처럼 도발할 확률은 크지 않습니다만, 러시아-우크라이나 전쟁도 예상보다 장기간에 접어들었잖아요. 늘 최악의 순간까지 고려해 준비하는 것이 필요하겠죠. 과도하게 우려하진 않더라

도 위기에 대비하려는 자세 말이죠. 만일에 사태를 우려한 정부가 컨틴전시 플랜을 세우고, 국내 유가가 오르면서 부담이 커질 것을 대비해 유류세를 만지작거리고 있지만요. 여러분 또한 중동 뉴스에 귀 기울여 보는 건 어떨까요.

4

금 투자

금 한돈에 40만원?...
금 투자 해야 할까?

이쯤되니 돈은 어떻게 벌어야 할지 막막하시죠? 물가와 유가는 오르고, 수출은 아직 회복세라고 하니 말이죠. 없는 살림에 조금이라도 알뜰살뜰 투자해서 볼러보고 싶은 마음은 매한가지 아니겠어요. 이때 여러분의 귀를 솔깃하게 하는 것이 있으니 바로 '금'이 아닐까요.

금값이 그야말로 금값입니다. 금값은 올초 g당 8만 5449원 수준이었는데요. 신한은행에 따르면 4월에 접어들면서 10만원을 넘어섰습니다. 12일에는 10만 7580원이 됐는데요. 한돈(3.75g)으로 환산하면 40만 3425원이네요.



신한은행에서 발표한 국내 금 시세

아니 금 한돈이 40만원을 넘어섰다니, 이제야 얼마나 금값이 금값인지 와닿으시나요. 금값이 오른 이유는 이를 사고파는 사람들이 늘어났기 때문입니다. 한국거래소에 따르면 지난달 일 평균 거래대금은 68억 6000만원으로 평소의 2.4배 수준이거든요. 즉 사람들이 이전보다 금을 더 많이 사들이면서 가격이 치솟았다는 건데요.

이 역시 중동 사태 때문입니다. 중동 지역에 긴장감이 커지자 투자자들이 갑자기 어떻게 될지 모른다는 불안함을 느끼게 된 거죠. 다른 투자처보다 이런 리스크에도 변함없는 금에 모여들게 된 거예요. 이처럼 세계경제가 불안하더라도 크게 가치가 바뀌지 않아 안정적인 것들을 '안전자산'이라고 해요. 금이 바로 대표적인 안전자산이거든요. 그래서 세계적으로 위험요인이 커지면서 불안감이 감돌게 되면 금과 같은 안전자산의 가격이 오르는 경향이 있습니다.

여기에서 생각할 게 있습니다. 그렇다면 금값이 여기에서 더 오를지, 혹은 지금이 고점일지 말이죠. 중동사태가 계속된다면 금을 찾는 사람들이 많아져 더 오를 것이고요. 그렇지 않다면 더 치솟긴 힘들겠죠. 이미 현재 금값이 역대급 수준이거든요. 그럼에도 안전자산이니 다른 투자수단보다 든든하다는 분들이라면 지금이라도 사도 좋겠지만요. 이미 너무 올랐고 중동 사태가 더 심화되긴 어렵다고 본다면 고민해보는 게 필요한 시점입니다. 자신의 투자성향에 따라 현명하게 결정하세요.



서울 지역 연립·다세대 원룸(33㎡이하) 평균 월세 추이
(출처: 앱 '다방'을 운영하는 스테이션3가
국토부 자료 기반으로 도출한 통계(보도자료 배포분))

5

월세

원룸 월세 100만원에 전세품귀...
무주택자는 어디로?

하늘 아래 집이 이렇게 많은데 두발 뺀고 잘 수 있는 내 보금자리 하나 찾기가 왜이리 힘들걸까요. 얼마 전까지 강통전세에 전세사기로 세입자를 울리더니 이제 월세까지 괴롭습니다. 손바닥 만한 원룸의 월세가 100만원을 넘어섰거든요. 부동산 정보 플랫폼 다방에 따르면 지난 2월 기준 서울의 준공 5년 이하, 보증금 1000만원 기준 연립·다세대 주택(전용면적 33㎡ 이하)의 평균 월세가격은 101만 5000원이나 합니다. 이는 다방에서 국토교통부 실거래가를 기준으로 서울에 준공 5년 이하, 보증금 1000만원 기준인 매물을 대상으로 산출한 통계예요. 지난해 2월과 비교하면 9%나 올랐죠.

왜 이렇게 월세가 오른 걸까요. 월세를 원하는 분들이 늘어났기 때문이죠. 지난 1월 국토교통부 실거래 공개시스템을 살펴보니 전국 빌라 전월세 거래량은 총 2만 1146건이었는데요. 이중 전세가 9268건, 월세는 1만 1878건이었어요. 임대차 계약 중에서 월세가 56.2%나 차지하네요. 이는 국토부가 관련 통계를 작성하기 시작한 2011년 이후 1월 통계로는 가장 높은 수준이거든요. 2021년 34.4%에서 2022년 42.8%, 지난해에는 53.2%였어요.

물가도 유가도 오를 수 있는데 월세까지 오른다니

허리가 휘어질 지경입니다. 왜 그런걸까요. 최근 집값이 떨어지면서 깡투기꾼이나 사기꾼들이 났던 전세들이 깡통이 됐습니다. 깡통전세를 비롯해 전세사기가 줄줄이 터졌죠. 이를 본 사람들이 전세를 피하기 시작한겁니다. 집을 사자니 너무 비싸고 그렇다고 대출받기도 쉽지 않으니 말이죠. 전월세 중에 골라야 하는데, 전세가 아닌 월세를 찾을 수밖에 없게 된겁니다. 수요가 몰리니 가격이 오를 수밖에요.

내집 없는 것도 서러운데 전세가 두려워 선택한 월세가 이렇게나 많이 올랐다니 걱정되시죠. 월세가 더 오를 수 있다는 것을 유념하고 안전한 전세매물을 찾아보거나 이를 내집 마련의 기회로 삼아보시는 건 어떨까요.

6 주담대에 예금금리까지 '뚝'... 저축 말고 집 살까?

월세는 너무 비싸고 전세는 위험하고, 그렇다면 이참에 나도 집을 마련해볼까 싶은 분 있으시죠. 내집을 마련하고픈 마음은 누구나 굴뚝같겠지만 그게 쉽나요. 집 한 채가 수억에서 비싸면 수십억원이나 하니 엄두가 나지 않잖아요. 집값은 워낙 비싸다 보

니 매입을 고민한다면 가장 먼저 살펴야 할 게 대출입니다. 대출없이 집을 사는 건 쉽지 않다보니 대출이 쉽게 나오는지, 저리로 빌릴 수 있는 상황인지 두루 고려해야 하죠.

지난해부터 내집 마련이 어려워졌습니다. 그 이유가 바로 대출에 있는데요. 정부에서 가계부채가 너무 늘어나니 대출을 받기 어렵게 규제를 강화한 것도 있는데요. 대출 금리가 꽤 올랐기 때문이죠. 코로나19로 한때 역대 최저수준의 기준금리를 기록했지만 금리가 다시 올랐거든요. 이에 따라 주택담보대출 금리도 높아졌습니다. 실제로 대출금리가 점점 올라 부담이 크다는 민원이 속출했을 정도예요. 금융감독원에 따르면 지난해 금융민원과 상담 등으로 72만6061건이 접수됐어요. 이중 금융민원이 9만 3842건인데 7.7% 늘었습니다.

무주택자를 주저하게 만들었던 이 주택담보대출 금리가 앞으로 안정될 수 있다는 전망도 나옵니다. 그 이유는 바로 '코픽스(COFIX·자금조달비용지수)'에 있어요. 코픽스가 떨어진겁니다. 은행연합회에 따르면 지난 3월 신규 취급액 기준 코픽스는 3.59%로 전월(3.62%)보다 0.03%포인트 하락했습니다. 지난해 11월 4.00%까지 올랐던 코픽스가 4개월 연속 하락하고 있는데, 이게 과연 무슨 의미



은행연합회가 발표한 최근 3개월 코픽스(COFIX) 신규취급액 추이 (매월 신규로 취급한 수신상품 금액의 가중평균금리로, 시장금리의 변동이 신속하게 반영되는 특징이 있음)

를 지닐까요.

코픽스는 금리를 산정하는 기준입니다. 은행권에서 이를 기준으로 주택담보대출 금리든 예금금리든 각종 금리들을 정하거든요. 무엇보다 중요한 건 주요 은행들이 이를 주담대 금리에 곧바로 반영하기에, 이를 보면 앞으로 주담대 금리가 어떻게 될지 가능할 수 있어요. 앞으로 새롭게 주택담보대출이든 전세자금대출을 받을 때 더 낮은 금리를 적용받을 수 있거든요.

반대로 코픽스가 내린다고 좋지만 한건 아닙니다. 주담대 금리는 물론 예금금리도 덩달아 내려가거든요. 은행에 저축할 때 적용되는 금리도 떨어진다는 거예요. 저축하는 메리트가 이전보다 사라지겠죠. 자 코픽스가 앞으로도 더 떨어질지 살펴보는 건 어떨까요. 대출받는게 내게 유리할지, 있는 돈이라도 예금하는게 유리할지 금리 변화를 살펴면서 계산해보세요.

7 N잡러 꿈틀대는 N잡러 2% 육박... 쥐꼬리 월급인데 나도 부업할까?

중동 사태가 심화하면서 유가와 물가가 오를 수 있는데다 월세는 원룸 마저 서울에선 100만원을 넘어섰습니다. 금이라도 사모아볼까 했는데 이미 발빠른 사람들로 금값도 금값이 됐고요. 수출 실적이 좋아진대지만 내게 바로 체감되기까지 한참 걸릴테니 가슴이 먹먹해집니다. 부족한 생활비는 어떻게 벌어야 할까 우려하는게 나만은 아닌 것 같아요.

기존에 다니던 직장은 물론 부업까지 한 취업자가 소폭 늘어났어요. 통계청 경제활동인구조사 마이크로데이터를 보니 올해 1분기에 부업을 한 적 있는 취업자가 55만2000명이더라고요. 지난해 1분기에



45만1000명이었으니 22%나 늘어난거예요. 이 같은 N잡러 인원이 엄청난건 아닙니다. 전체 취업자 중에서는 1%대에 불과하거든요. 물론 2019년 1분기 1.34%에서 1.97%로 오르면서 2%대에 육박해졌고요. 증가세를 보면 청년층과 40대에서 뚜렷하다는 점에 주목됩니다. 연령별로 증가세를 살펴보면 청년층 부업자는 전년 대비 가장 큰폭인 30.9% 늘었습니다. 40대 부업자도 27.7% 증가했구요.

다만 늘어난 부업 대부분이 '배달'이나 '택배' 등이다 보니 부업으로 노동시간이 늘어난 것 대비 소득 증가는 미미한 상황이에요. 이처럼 경제 상황에 리스크가 커지다 보니 N잡을 고려할 분들도 늘어날텐데, 그 소득 증가는 미미한 상황도 고려해주세요. 너무 암울하다고요? 현 상황에 과도하게 걱정하기 보다 내게 맞는 방안을 찾아보는 건 어떨까요. 생활비 지출을 현명하게 줄이는 팁을 찾아보고 안전하면서도 수익률이 조금이나마 높은 투자처를 찾아보면서 안정적인 주거를 만드는 방법요. 쉽지 않다고요? 다시 한번 통계를 들여다보는 건 어떨까요?

생성 AI 시대 최고의 교수법

박남기 | 광주교대 교수



1. 들어가며

4월 21일은 과학의 날이다. 초등학교 2학년 선생님이 페이스북에 다음과 같은 협조요청 글을 올렸다.

▶ 영상추천 요청 | 2학년 과학의날 행사로 “과학상상화 그리기”를 합니다. 주제는 “과학 기술이 발전된 미래 사회의 모습을 상상하여 그림으로 표현한다.”인데요. 2학년 수준에 적합한 영상(40~80분량) 있으면 추천 부탁드립니다. 영화도 되고, 애니메이션도 됩니다. 가장 맘에 드는 영상 추천해 주시는 분께 제가 스벅 아메쿠폰 보내드립니다. ^^

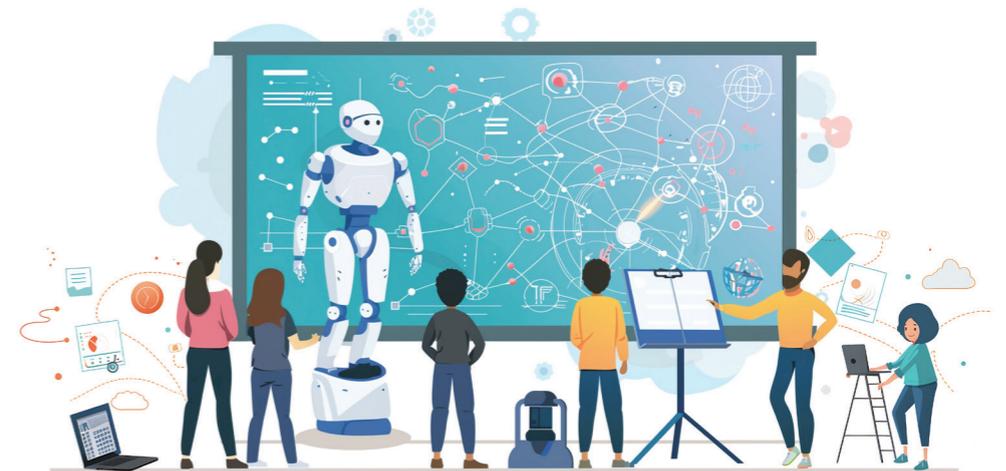
몇 분 선생님들께서 영화나 동영상을 추천한 것을 보고, 생성 AI가 가지고 있는 수업용 영상자료 추천 기능을 활용해 보았다. 선생님이 올린 글을 복사해서 ChatGPT에게 입력하니 몇 개의 영상을 추천하면서 추천이유, 주요 내용, 시청 방법, 영상 길이 등의 정보를 함께 제공했다. 제시된 답변을 복사해서 보냈더니 내가 커피 쿠폰에 당첨되었다며 보내

왔다. 그래서 기계가 답한 것을 그대로 보낸 것이 아니라 대신 열심히 답한 다른 선생님께 드리라며 사양했다.

생성 AI는 수업 자료와 관련하여 위에서 이야기한 온라인 자료 추천 외에 텍스트 자료 개발, 지침 개발, 시각 자료 제작, 그림 인식 기능 활용한 수업 자료 제작, 토론 주제 제시, 역할극 시나리오와 대본 제작 등등 교수자가 필요로 하는 수업 자료를 거의 대부분 만들어준다. 교수자들이 이러한 사실에 대해 잘 알고 있을 것으로 생각되는데 실제로 사용하는 예가 아직까지는 많지 않은 것 같다.

‘4차 산업혁명의 증기기관’이라는 생성 AI는 교육계의 마약일까 신약일까? 생성 AI 시대, 최고의 교수법은 무엇일까? 호리병을 빠져나온 지니를 다시 집어넣기는 어렵다. 인공지능이 이미 나와 있고, 교수자와 학생들이 널리 사용하고 있는 상황에서 위험성만 강조한다고 해서 사용을 막을 수 있을 것 같지는 않다. 스마트폰을 다양하게 활용할 역량을 갖춘 사람과 그렇지 못한 사람 간에는 삶의 질에서 큰 차이가 난다. 생성 AI를 잘 활용할 수 있는 사람과 그렇지 못한 사람의 차이는 스마트폰 경우보다 더 커질 가능성이 있다.

이 글에서는 필자가 출판한 책 「생성 AI 시대 최고



의 교수법(2024. 천재교육)』을 바탕으로 생성 AI 사용시 유의점을 밝히고, 수업 준비·수업 진행·학생 평가 등의 수업 활동, 생활지도를 포함한 학급경영 활동 관련해서 어디까지 활용 가능한지에 대해 간단히 소개하고자 한다.

2. 생성 AI 활용 시 유의점

AI를 교수 학습 활동에 활용할 때 유념할 것이 있는데, AI 의존도 심화가 자칫 교육자와 학습자의 교육력 및 역량 약화로 이어질 가능성이 있다는 점이다. 시든 가상(증강)현실이든 모두 교사의 교육력이 튼튼해야 교육의 효과를 높여주는 역할을 한다. AI 시대의 교사는 교육 내용만이 아니라 교육철학, 생활 지도를 포함한 학급경영 역량, 학습동기 유발을 비롯한 교수법 등에 더욱 관심을 가져야 한다. 그렇지 못하면 AI로 인해 교사의 교육력이 저하되면서 교사

무용론이 대두될 수도 있다.

학습자 또한 AI 의존형이 아니라, AI 활용형 인간이 될 수 있도록 자신의 기본 역량을 길러가야 한다. 그렇지 않을 경우 역량을 갖춘 뛰어난 인간이 되는 것이 아니라, AI에 의존하는 무능한 인간으로 전락하게 될 수도 있다. 예술 활동을 비롯한 창의적 활동에 대한 AI 의존도가 높아지면 인간의 고급역량마저 저하되게 될 것이다. 우리가 바라는 것은 AI 도움을 받아 뛰어난 역량을 갖춘 인간이 되는 것이지, AI에 의존하는 무능한 인간이 되는 것이 아님을 기억해야 한다. 이는 교육자와 학습자 모두에 해당한다.

이미 잘 알려진 것처럼 AI의 윤리적 사용과 사생활 보호, 그리고 거짓말을 사실처럼 얘기하는 환각 현상(hallucination)에도 유의해야 한다. 특히 교사들은 학생들이 생성 AI에 의존하거나 중독되지 않도록 유의하면서 협업하는 역량을 기르도록 지도할 필요가 있다.



3. 생성 AI 시대 최고의 교수법

최고의 교수법이란 어떤 특정의 교수기법이 아니라 교수가 자신에 맞는 교수법을 찾아 끝없이 노력하는 과정 자체를 의미한다. 생성 AI 시대 최고의 교수법은 생성 AI를 활용하여 자신에 맞는 교수법을 만들어가는 지속적인 노력 자체를 의미한다. 수업 준비, 수업 진행, 학생 평가, 수업혁신 등의 수업 관련 활동과 성공적인 수업의 토대가 되는 학생 학부모와의 소통 및 학생 행동경영을 포함하는 제반 학급경영에도 생성 AI를 활용할 수 있다.

가 수업 준비에의 활용

교수자들은 수업 준비와 관련하여 어디까지 생성 AI를 활용할 수 있을까? 수업 준비활동의 순서에 맞춰 지속적으로 실험을 해본 결과 거의 모든 준비 활동에 활용할 수 있음을 알 수 있었다. 강의(수업) 계획

서 만들기, 차시별 강의록(수업지도안) 만들기, 수업용 PPT 슬라이드 개요 만들기(PPT 전용 AI를 활용하면 PPT 제작 의뢰도 가능), 수업 주제별 동기유발(재미있는 수업 진행) 아이디어 얻기 및 프로그램 제작, 수업 주제별(차시별) 활동계획 작성 등 대부분의 것을 생성 AI에게 요청할 수 있다. 수업 자료 제작과 관련해서는 텍스트 자료, 지침 개발, 가상 사례 만들기 혹은 실제 사례 찾기, 온라인 자료 추천 및 시각 자료 제작, 토론 주제, 요약 자료 만들기, 역할극 시나리오와 대본 만들기 등도 요청이 가능하다.

아울러 첫시간 마음열기(아이스 브레이킹) 프로그램 만들기, 학습공동체 형성 및 강화 활동 프로그램 만들기까지도 가능하다. 잘 사용하면 수업 준비에 들어가는 시간을 크게 줄이면서 재미있고 질 높은 수업이 되도록 준비할 수 있게 되었다. 풍부하고 질 높은 자료, 학생 맞춤형 수업 자료 제작은 수업의 효과를 더욱 높여줄 것이다.

나 수업 진행

지금까지 사용해본 내 경험에 비춰볼 때 생성 AI가 할 수 있는 것은 그에게 맡겨 쉽게 처리하고자 하는 유혹을 이겨내기가 상당히 어렵다. 고급 역량이 제대로 길러지지 못한 어린 학생들이 생성 AI에 노출될 때 발생할 수 있는 중독성과 의존성 문제를 해결할 수 있는 프로그램이 함께 운영되어야 할 것이다. 그 프로그램이 성공적임이 입증될 때까지는 제한된 범위에서 소극적으로 활용하는 것이 바람직할 것이





다. ChatGPT 회사가 13세 이하 아동의 가입을 금하는 이유도 그러한 문제 때문이다. 초등학교에서도 굳이 사용하고 싶다면 5학년 이상에서 ‘모의실험 기반 학습’ 등 극히 제한적 범위에서만 사용하길 권한다. 물론 교사의 지도 및 감독 역량, 과목의 특성에 따라 사용 학년에는 융통성이 있을 수 있다.

중고등학교에서도 수업 중에 사용하고자 할 때에는 문제점을 충분히 알리고, 그러한 문제가 발생하지 않도록 지속적인 훈련을 시키면서 제한적으로 사용하는 것이 바람직하다. 다만, 대학에서는 관련 문제를 적시하고, 학생들이 그러한 문제에 대해 책임져야 함을 알리면서, 적극적으로 사용해도 좋을 것 같다. 이제는 생성 AI와의 협업 역량을 키우는 것이 필요한 시점이 되었다. 활용 효과와 발생하는 문제점에 대한 관련 연구가 세계적으로 다양하게 이뤄지고는 있지만 아직 확실한 것은 없다는 점을 교수자들이 명심하길 바란다.

수업진행과 관련하여 생성 AI를 활용할 수 있는 범위는 개인 맞춤형 학습 지원, 학생들의 글쓰기 능력 향상 지원, 생성 AI 도움을 받을 수 있는 학습 활동,

생성 AI 비판적 활용, 생성 AI 활용 수업 활동 등으로 나눠볼 수 있다. 학습 지원과 관련한 ChatGPT의 가장 큰 강점은 학생 개인 맞춤형 학습 지원이 가능하다는 것이다. 아직은 많은 한계가 있지만 교수자나 학생이 요청하면 개인화된 맞춤형 학습 자료를 일부 생성할 수 있다. 그런 다음 학생의 특정 요구와 학습 스타일에 맞게 결과를 조정할 수 있다.

글 쓰는 역량을 거의 갖추고 있지 않고, 글 쓰는 것도 두려워하는 학생이 많은 경우 교사 혼자서 이 많은 학생들을 개별 지도하기가 어렵다. 하나의 방안은 생성 AI와 협력하여 학생들 글쓰기 지도를 하는 것이다. 이를 위해 학생들에게 ChatGPT와 협력적 글쓰기 방식을 소개하고 활용하도록 안내할 수 있다. 이러한 과정을 거치면서 개인의 글쓰기 역량과 태도 등이 강화되면 그다음에는 기계의 도움 없이 혼자서 쓰도록 유도해야 한다. 도움을 받는 것은 메마른 펄프에서 물이 나오도록 돕기 위한 마중물에 해당함을 학생들이 깨닫게 해야 한다. 그리고 수업 중에 기계 도움 없이 글을 쓰게 하여 학생들이 그러한 역량을 기르고 있는지를 평가할 필요가 있다. 이러한 제반 과정에서 교사가 해야 할 가장 중요한 역할은 학

생들이 글을 잘 쓰고 싶다는 강한 동기를 갖도록 유도하는 것이다.

수업 중에 생성 AI를 활용하면서 생성 AI가 제공한 답에 대한 비판 역량, 오류 찾는 역량을 길러주는 활동을 통해 고급역량을 길러줄 수도 있다. 생성 AI를 활용한 수업 활동 예로는 모의실험 학습, 1대1 토론, 조별 활동(브레인스토밍)에서의 사고의 폭 넓히기, AI와의 협업 역량 제고 등을 들 수 있다.

다 학생 평가

인공지능을 활용한 새로운 평가 시대가 열리고 있다. 교육활동 중에서 가장 힘든 것 중의 하나가 평가인데, ChatGPT를 비롯한 생성 AI가 크게 도움을 줄 수 있는 분야의 하나가 바로 평가이다. 기계가 너무나 쉽게 평가문항을 만드는 것을 보더니 명퇴했던 중등교사가 다시 복직하고 싶다고 하던 이야기가 생각난다.

평가 활동 관련해서 생성 AI를 활용할 수 있는 영역은 각종 평가문항 만들기, 채점 기준표(rubrics) 만들기, 텍스트형 과제 채점 및 코멘트, 보고서 피드백, 생활기록부 작성 등 다양하다. 구체적인 내용은



(참고문헌) 박남기(2024.05). 생성 AI 시대 최고의 교수법. 서울:천재교육

생략한다.

라 학급경영

학급경영은 교수자가 하는 수업경영, 생활지도, 학부모 관계, 교실환경 구성 등등의 제반 교육 활동을 일컫는 용어이다. 수업 역량만 뛰어나면 수업이 성공할 것이라고 생각하는 교수자가 많다. 수업의 성패를 좌우하는 토대는 수업 경영을 포함한 학급경영 역량이다. 학급경영에 필요한 제반 아이디어 구하기, 자료 제작 등에서 생성 AI를 활용할 경우 교사의 업무 효율성과 효과성이 크게 높아질 것으로 기대된다.

4. 나오며

생성 AI 시대라고 해서 가르침과 배움의 본질이 크게 바뀌지는 않을 것이다. 교수자들은 가르침과 배움은 무엇인지, 배움은 어떤 과정을 통해 어떻게 이뤄지는지, 그리고 교수자의 역할은 무엇이 되어야 할지에 대해 지속적으로 연구하며 역량을 강화해야 한다. 생성 AI 시대에도 교수자가 갖춘 지식과 역량은 아주 중요하다. 생성 AI를 잘 활용하려면 이용하고자 하는 분야에 대한 충분한 지식과 경험을 갖추고 있어야 한다. 잘 알지 못하면 생성 AI가 제공한 답변의 오류 식별이 어려워 활용도가 떨어진다. 또한 생성 AI 활용으로 인한 ‘학습의 빈익빈부익부 현상’이 가속화되면서 교수자 및 학생들의 지식과 역량 양극화는 더욱 심해질 수 있다.

생성 AI 활용은 피할 수 없는 상황이다. 교수자와 학생의 생성 AI 활용을 막을 수 없다면 국가 차원에서는 위험성을 최소화하기 위한 제도적 보완 장치를 마련하고, 교육기관 차원에서는 교사와 학생이 윤리적으로 활용할 수 있는 역량을 길러주기 위해 노력해야 할 것이다. 생성 AI 시대 교육의 미래는 활용에 따른 위험성을 어떻게 최소화할 것인가에 달려 있음을 다시 한번 강조하며 글을 맺는다.

공공과 민간 데이터의 ‘효율적 공유와 통합’ 통해 새로운 가치 창출

김창훈 | KRG 대표



‘범람하는 데이터 홍수’ 시대, 가치있는 데이터로 전환해야

바야흐로 우리는 ‘데이터 홍수’ 시대에 살고 있다. 전 세계에서 매일 쏟아지는 데이터양은 2.5엑사바이트(EB: 25억GB)에 달하는데, 이는 1GB짜리 영화 250억편에 맞먹는 양이다. 하지만 이처럼 엄청난 양의 데이터 중 90%는 원시 데이터 수준에 그친다. 이런 원시 데이터를 ‘가치 있는(Valuable) 데이터’

로 만들기 위해선 유사 데이터 간의 적절한 통합과 효율적인 가공 과정이 반드시 필요하다. 그렇지 않다면, 대부분 원시 데이터는 일정 시일이 경과하면 90%는 ‘무용한 데이터’로 전락할 가능성이 크다. 여기서 우리가 주의깊게 바라봐야 할 것은 신뢰할 수 있는 공적 기관에서 발표한 데이터를 어떻게 적절하게 접목하고 가공, 분류해 우리 삶에 보탬이 되게 하는 소스(Source)로 재탄생하게 만들 것인가이다. 이와 관련해 최근 여러 나라가 공공 데이터와 민



간 데이터를 공유하고 결합해 의료나 복지 등 공적 영역에서 의미 있는 데이터를 생성해 정책 또는 제도적으로 반영하는 작업에 나서고 있다. 하나의 단적인 예를 보자. 최근 통계청에서 MZ세대 속성에 관한 통계자료를 발표했는데, 그들 중 절반이 넘는 숫자가 서울, 경기, 인천 등 수도권에 거주하는 것으로 조사됐다. 이 데이터가 아니더라도 지방에 젊은 층이 없다는 것은 누구나 상식적으로 알고 있다. 여기서 정책적으로 고민해야 하는 것은 MZ세대들이 자신이 태어난 지역에 거주할 수 있는 대안을 만들어야 한다는 점이다. 이를 위해선 MZ세대의 현 주거 환경에 관한 기본적인 공공 데이터와 민간에서 보유하고 있는 특정 지역 채용수요나 생활 만족도, 문화 인프라, 취업 환경 등의 데이터를 결합한다면, 수도권 집중화 해소를 위한 정책적 실마리도 제공할 수 있다는 점이다.

데이터 공유 통한 공적 영역에서의 활용 효과

개인이나 기업이 보유한 민간 데이터는 공적 영역에서 정책 의사 결정을 내리는 데 유용하게 사용된

다. 가령, 코로나19 같은 전염병이 만연할 때 표적 맞춤형 대응이 가능하다. 또한 더 발전된 도시 모델을 연구하거나, 환경보호와 관련된 다양한 정책을 수립하는데 도움을 줄 수도 있다. 또한 시장 경제 체제에서 시장이 제대로 작동되는지 모니터링할 수 있고 이는 소비자 보호로 이어지는 선순환 효과로 이어진다.

공식 통계를 작성할 때도 때로는 민간 데이터를 활용하고, 분석하는 게 비용효율적인 경우도 많다. 인구 이동이나 가격, 인플레이션, 교통 분야 등에서 더 빠르고 현실에 부합되는 결과를 얻을 수 있다. 특히 민간 데이터를 적절하게 활용하면 공공 데이터를 취합하는 데 소요되는 많은 자원과 비용도 절감할 수 있다.

컨설팅 기업 맥킨지(Mckinsey)에 따르면 데이터양은 매년 50% 이상 증가한다. 다만, 공익 목적으로 작성된 공적 데이터와 별개로 개인이 소유한 데이터들은 각기 사용 용도별로 갖고 있기 때문에 굳이 공유의 가치를 느끼지 못한다. 하지만 중요한 것은 이러한 민간 데이터를 현명하게 공유한다면 이 데이터는 풍부한 사회적 자산으로서 그 가치는 더 커

질 수 있다는 점이다. 가령 기업들이 보유한 데이터를 상호 공유하게 되면 효과적인 정책 결정에 중요한 모티브가 된다. 특히 ‘공공선’을 위한 민간 데이터 활용은 공공 데이터가 커버하지 못하는 사각지대까지 아우른다.

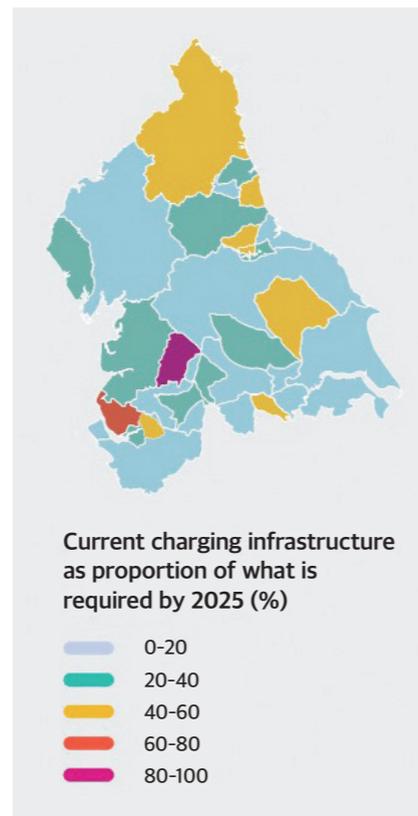
때문에 전 세계적으로 공공 데이터와 민간 데이터 간의 공유와 결합을 통해 새로운 가치를 창출하는 사례가 점차 많아지고 있다. 공유 사례 가운데, 대표적으로 사회 인프라(SoC) 분야, 교육 분야, 의료 분야, 복지 분야 등의 글로벌 사례를 몇 가지 소개한다.

분야별 공공-민간 데이터 공유 사례

사회 인프라(SoC)

타 산업에 비해 농업의존도가 높은 미국의 대표적인 주인 캘리포니아는 해마다 ‘물 부족’ 사태에 시달린다. 이를 해소하기 위해 캘리포니아주는 지난 2006년부터 주정부 주도로 수자원 배분 정책을 펼치고 있다. 여기에는 ‘캘리포니아대학교 산하 지구 연구소’, ‘세계식량센터’, ‘반도체 기업 인텔’ 등이 참여하고 이들 기관에서 제공받은 지식 데이터를 신기술 기반의 분석 도구를 활용해 가공작업을 수행하고 있다. 이같은 작업의 결과물을 통해 캘리포니아주정부는 효과적인 수자원 배분 정책을 추진, 가시적인 성과를 거두고 있다.

미국 테네시주 채터누가(Chattanooga)시는 주변에 3개의 주(States)와 16개의 카운티(County)가 있다. 채터누가는 민간이 보유 중인 운송 데이터베이스에 접근할 수 있는 툴(Tool)을 공유하고 있다. 이 지역에 소재한 20개 이상의 화물 중개인과 운송 회사는 운송 데이터베이스를 공유해 상품의 이동 경로를 추적할 수 있다. 여기에서 해당 지역의 도로 이용자를 정량화하거나 화물 병목 현상에 대한 원인을 파악해 주 정부 차원에서 운송 및 물류 관련한 이슈들에 대해 적절한 대응을 가능하게 한다.



영국지리공간위원회는 민간 데이터 공유를 통해 전기차 충전 인프라 시설 계획을 수립하고 있다

이 데이터베이스는 정부 기관 및 기업이나 교통전문가를 대상으로 무료로 제공된다.

영국의 지리공간위원회(Geospatial Commission)는 공공과 민간 데이터를 공유해 전기자동차가 어떻게 교통 여건을 개선할 수 있는지에 대한 연구 결과를 발표했다. 공유 작업을 통해 지역별 전기자동차 채택률을 결정하는 도구로 활용하는 한편 충전 인프라 시설이 필요한 지역이 어디인지를 조사한다. 더 나아가 특정 지역의 전기자동차 수요를 충족하기 위한 인프라 개발 계획을 사전에 결정하게 된다.

실업이 사회적 문제로 대두되는 유럽에서도 여러 텔레콤 기업에서 제공되는 익명화된 데이터를 통해 통화기록 및 통근 패턴을 분석, 지역별 실업 현황과 해당 지역민들의 심리상태를 분석해 대국민 서비스를 제공하기 위한 통찰력을 얻고 있다.



세계경제포럼(WEF)은 미래 노동환경 변화와 일자리 전망 정책 발걸을 위해 채용 전문 기업인 링크드인이 보유한 데이터를 활용하고 있다

교육

교육 분야는 민간 데이터가 가장 효과적으로 적용할 수 있는 분야다. 빅테크 기업 구글은 코딩 등 수요가 많을 것으로 예상되는 분야를 포함해 교사와 학생들을 위한 다양한 리소스와 프로그램을 제공한다. 교사들은 다양한 앱을 통해 수집된 정보를 가지고 수업 계획부터 채점까지 이전과 다르게 효율적으로 작업할 수 있다.

대표적인 인력 채용 기업인 링크드인(Linked in)은 IT 관련 일자리 수요 데이터를 미국 노동부에 제공하고 있는데, 이 데이터는 노동부의 공공 데이터와 결합해 인재 양성 프로그램 개발에 활용되고 있다. 또한 세계경제포럼(WEF)은 미래 노동시장 변화와 이에 대응한 일자리 발굴과 교육 영역에서 링크드인이 보유한 미래 기술과 노동시장 변화 등에 관한 데이터를 공유해 미래 기술 발전에 부응하는 인력 양성 정책에 나서고 있다.

의료

의료데이터는 공공 및 민간에서 다양한 데이터가 축적돼 있다. 이처럼 산재된 의료데이터를 민간과 공공이 상호 공유하고 교환하면 공공 의

료 영역에서 좋은 성과를 기대할 수 있다. 안전하지 않은 불량 식품이나 코로나 같은 바이러스 발생 등의 건강상 이슈를 국민에게 미리 경고할 수 있다. 민간-공공 데이터 간 상호 공유를 통해 코로나19 팬데믹 상황에서 특정 지역의 코로나 발생 상황을 실시간으로 알려줘 경각심을 제공하는 한편 봉쇄 기간 동안 사람들의 동선을 추적하기 위해 위치 기반 데이터를 활용하기도 했다.

민간과 공공 데이터의 공유는 많은 연구자에게 정보 접근성을 제고시키는 한편 이미 수행한 작업과 결과물에 대한 데이터와 정보를 기반으로 새로운 선진 연구를 진행할 수 있다. 세계보건기구(WHO)는 2022년도에 WHO 주도로 수행된 모든 프로젝트 결과물에 대한 데이터 공유 및 관리 계획을 수립해야 한다는 정책을 발표한 바 있다.

영국의 로체스터대학교는 파킨슨병으로 고통받는 사람들의 손동작, 균형감각이나 보행 등에 관한 데이터를 수집하기 위해 자체 앱을 개발했는데, 6시간 만에 7천여 명 이상의 사람들이 앱을 다운받아 자신의 정보를 기꺼이 제공했다. 만일 일반적인 루트로 이 데이터를 얻으려 했다면 아마도 최소한 수개월이 걸리는 까다로운 작업이었을 것이다. 로체스터

대학교는 여기서 얻은 데이터를 다른 연구자와 환자에게도 공유함으로써 파킨슨병을 치유하는 연구에 많은 도움을 주고 있다.



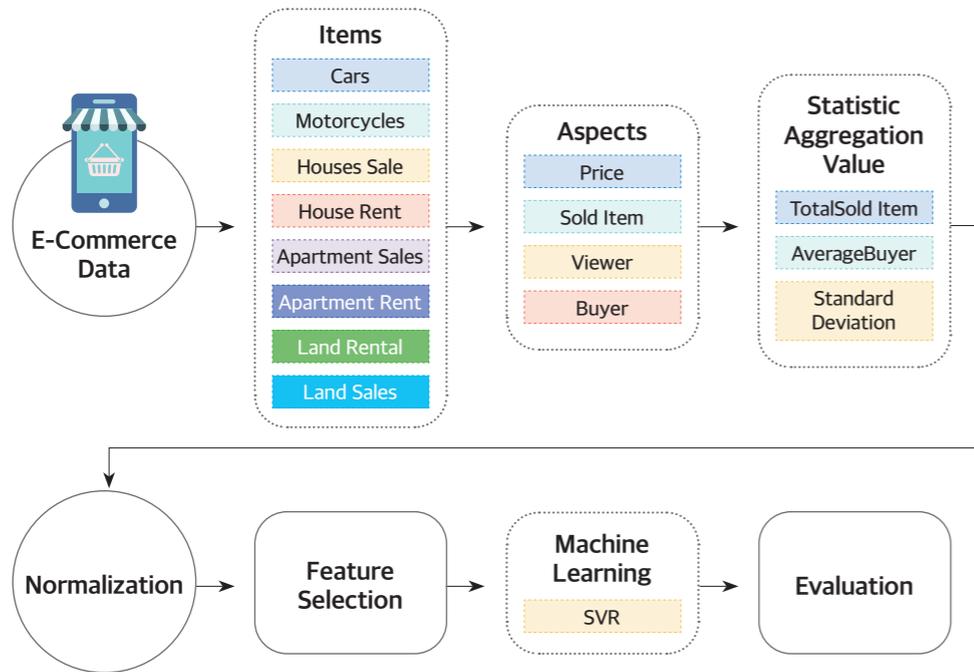
복지

복지 분야에서도 민간과 공공 데이터 공유는 훌륭한 성과를 만들 수 있다. 주거 환경 개선을 위한, 또는 사회 취약층에 관한 공적 데이터도 민간이나 특정 기관에서 보유 중인 데이터와의 상호 연계를 통해 효과적인 대책 수립을 위한 훌륭한 소스가 된다.

인도네시아 텔콤대학의 데디 라만 위자야(Dedy Rahman Wijaya) 교수팀은 개발도상국들이 직면한 과제인 빈곤 타파를 위한 정책개발에 민간 데이터를 활용한 새로운 모델을 제시해 세계적인 과학 저널지 네이처에 소개되는 등 많은 주목을 받았다. 인도네시아는 전체 국민의 10% 가량이 빈곤층에 속한다. 문제는 이같은 빈곤층을 조사하기 위해서

는 대량의 설문조사와 대면조사가 필요하고, 설령 결과가 발표된다고 하더라도 시기적으로 정책을 개발하기까지 많은 시일이 걸린다는 점이다. 빈곤 정책은 타이밍이 중요하다. 과거 데이터를 가지고 현실을 반영하는 데는 한계가 있다는 점을 인식한 데디 라만 위자야 교수팀은 빈곤층 및 그들의 거주 지역, 소비행태를 조사하기 위해 5년마다 실시하는 대량의 설문조사 대신 기존 데이터를 활용하는 방식을 제안했다. 그것은 인도네시아 최대 통신사가 갖고 있는 e커머스 데이터를 활용하는 것이다. 기존에 빈곤 상태를 추정하는 방식으로 많이 활용된 것은 야간 조명 밝기였다. 일반적으로 빈곤 지역은 중-상류층 지역보다 야간 조명 밝기가 약할 것이라 가정에 착안한 것이다. 하지만 문제는 이같은 가정이 빈곤층을 추정하기에는 한계가 있다는 점.

하지만 e커머스 데이터는 해당 지역의 실질적인 소비 지출 정도를 파악할 수 있기 때문에 빈곤 수준을 추정하는 데 더 효과적인 데이터라고 할 수 있다. 물



인도네시아 텔콤대학의 Dedy Rahman Wijaya 교수팀은 빈곤 추정을 위한 설문조사 대안으로 e커머스 데이터와 머신러닝 알고리즘을 활용할 것을 제안한다

론 여기에는 머신러닝 같은 신기술 영역이 뒷받침되어야 한다. 데디 라만 위자야 교수팀은 빈곤층을 조사하는 데 필수적인 설문조사와 인구 조사 대신에 더욱 저렴하고 비용효율적인 방법으로 특정 기업이 보유 중인 e커머스 데이터와 머신러닝 알고리즘을 활용하는 방법을 채택했다. 이같은 방식으로 업로드된 정보는 실시간으로 업데이트돼 정책 당국자들에게 시의적절한 빈곤 대책을 수립하고, 정책의 우선순위를 결정짓는 데 중요한 기여를 할 수 있다는 것이다.

공유를 위한 선결 과제

현재 국내에서 2023년 기준으로 약 9만여 건의 공공 데이터가 개방돼 있다. 미국 역시 대기질부터 농산물 시장 정보 등을 포함해 100만 개 이상의 공공 데이터 세트가 온라인상에서 무료로 공개돼 있다. 하지만 환경이나 생체 인식, 역학 및 물리적 상태에 관한 데이터는 대부분 민간이 보유하고 있다. 따라서 민간 데이터를 공적 데이터와 공유할 수 있는 여건 조성이 필요하다.

우선 민간 데이터 활용도를 높이려면 강력한 인센티브 제도가 필요하다. 공익을 위한 데이터 공유는 해당 데이터를 보유하고 있는 민간의 도움이 절대적으로 필요하다. 이를 위해선 세제 혜택 등의 구체적인 세밀한 인센티브 정책을 펼칠 필요가 있다.

두 번째는 AI와 빅데이터 기법 등 신기술을 통해 공공과 민간 데이터의 공유와 융합을 지원할 수 있어야 한다. 앞서 인도네시아 빈곤 추정 사례처럼, 공공 데이터의 질을 제고시키고 민간 데이터와의 시너지 효과를 극대화하기 위해서는 인공지능이나 머신러닝, 딥러닝 등의 최신 디지털 기술 적용 수준을 제고해야 한다. 물론 관련 전문인력 양성에도 적극 나서야 한다.

셋째는 공공 데이터와 민간 데이터가 그 본질에 부합되게 구축돼야 한다. 국내에서 실시 중인 대표적인 조사 가운데 산업별 실태조사가 있다. 공공기관에서 매

년 조사하고 있는 주요 산업 실태조사는 해당 산업에 종사하는 기업 매출이나 인력 보유, 수출현황, 정책적 유의점 등을 조사한다. 반면, 같은 산업에 연관된 민간 보유 데이터는 비즈니스 친화적인 목적으로 작성된 데이터다. 즉, 데이터의 속성이 다른 것이다. 이것을 무리하게 하나로 혼용해 조사하게 되면, 당초 목적과 어긋난 데이터가 생산될 수 있다. 따라서 데이터 생산 주체별로 지향하는 목적을 정확하게 이해해야 한다.

넷째는 개인 정보 보호 문제이다. 공공과 민간 데이터를 공유하는 문제는 흔히 정부 기관과 민간이 정보를 교환할 때 주로 발생한다. 문제는 이런 결합을 통해 새로운 가공 데이터가 만들어질 때 관련 개인들은 자신의 개인 정보가 제대로 보호받는지에 대한 의구심이 든다는 점이다. 공공에 활용된다는 목적하에 개인 데이터가 아무런 안전장치 없이 이곳저곳에 제공된다면 개인들은 불안할 수밖에 없다. 2021년 한 연구에 따르면 미국내 학교 앱의 60%가 제삼자에게 데이터를 전송한 것으로 조사됐다. 공립학교에서 사용하는 앱 중 18%는 수천 개의 네트워크로 연결돼 있어, 정보를 악용할 가능성이 있는 위험한 채널에도 전송된 것으로 보고됐다. 따라서 해당 데이터를 교환하는 경우 상호 간 책임 소재를 확실히 하는 한편 투명성을 높이는 노력이 필요하다.



굴뚝 없는 황금산업, MICE 산업

최서우 | 한국데이터연구소 대표



MICE산업 정의

MICE는 Meeting(회의/기업회의), Incentive Travel/Trip(포상관광), Conference/Convention(국제회의/컨벤션), Exhibition&Event(전시, 이벤트)의 영문 앞글자를 합친 용어이다. 초기의 MICE 용어와 개념은 주로 대규모 국제회의와 전시회 등을 통칭하기 위한 목적으로 사용되었으나, 현재는 포상관광, 이벤트 등을 포함하여 개별 분야를 통합한 개념으로 사용되고 있다(김철원·허준, 2011). MICE라는 용어는 1990년대 중반 싱가포르에서 유래되었다. 이 용어는 정보 교류 및 토론을 목적으로 국제기구와 협회 등이 개최하는 회의 및 비즈니스 이벤트 행사를 의미하며, 국가나 단체에 따라 다양한 형태로 사용되고 있다.



[표1] UN Tourism의 MICE 분류체계 및 포괄범위

구분	내용
기업회의 (Meeting)	<ul style="list-style-type: none"> 기업이 주최하는 비즈니스 목적의 회의로, 참가자는 동일 기업 내 회의, 기업의 그룹사 내 회의, 고객사·사업제공사 등의 관계 간에 개최되는 회의를 의미함
총회 및 협의회 (Convention, Congresses and Conferences)	<ul style="list-style-type: none"> 비즈니스 외 목적으로 개최되는 회의를 총칭하며, 참가자를 소집하기 위하여 별도의 장소 섭외 및 시간 선정 등의 요인이 결정되어야 함. 이 같은 행사에는 다음과 같은 유형이 포함됨 ① 공공컨퍼런스 및 강의(Public conferences or lecture): 기본적으로 큰방향이 없는 회의로, 청중에게 요구되는 과거경험이나 특별지식이 없는 회의 ② 정부회의(Government conferences): 전문적·정치적 회의로 정부기관 또는 정부기관 대 민간담당자간에 개최되는 회의 ③ 일반소집회의(General assembly): 조직내구성원 또는 내부관계자에게만 한정하여 소집되는 회의 ④ 컨벤션(Convention) : 입법기관, 사회적 또는 경제적 그룹간에 개최되는 일반적, 공식적인 회의로 참가자들에게는 정보제공, 의견·정책의 합의 도출을 목적으로 하는 회의 ⑤ 과학적 총회(Scientific congress): 여러 가지 사안에 대한 심도 있는 연구를 수행하기 위하여 전문적, 학술적 목적으로 개최되는 회의로, 참가자들간에 상호적인 담론이 요구되는 회의 * 예: 포럼(forum), 콜로키움(colloquium), 세미나(seminar), 심포지엄(symposium) 등
전시회 (Exhibition)	<ul style="list-style-type: none"> 무역전시회(trade show) 관련 활동만을 포괄함
인센티브 (Incentive)	<ul style="list-style-type: none"> 인센티브활동(Incentive activities)과 관련한 행사를 의미함

(자료: UN Tourism(2008). Global Meeting Initiative. 1. 2-3 재인용)

세계연합(United Nations)의 관광전문기구인 UN Tourism에서는 MICE 용어를 산업별로 구분하여 사용할 것을 권고하고 있으나, 국가 및 기관에 따라 상이하게 사용되고 있다(황희곤, 2010). ‘MICE산업’ 용어는 한국 외 싱가포르, 일본 및 태국 등 아시아 지역에서만 사용되고 국제적으로 통용되고 있지는 않다. 그 이유는 UN Tourism(세계관광기구)에서 ‘MICE산업’보다는 ‘회의산업(Meetings Industry)’ 또는 ‘국제회의산업(International Meetings Industry)’으로 분류하여 사용하고 있기 때문이다([표 1] 참조).

기 언급한 바와 같이 국내에서는 ‘MICE산업’이라는 용어를 사용하고 있고, 산업의 분류를 Meeting(회의/기업회의), Incentive Travel/Trip(포상관광), Conference(국제회의), Exhibition(전시)으로 규정하였다. 산업마다 소관부처가 다른데 문화체육관광부는 Meeting(회의/기업회의), Incentive Travel/Trip(포상관광), Conference(국제회의), 산업통상자원부는 Exhibition(전시)을 소관하고 있다. 최근 국제회의와 전시회가 동시에 열리는 등 융복

합 국제회의의 개최가 증가 양상을 보이고 있기 때문에 각 부처 간 협업이 절실히 필요한 실정이다.



한편, 문화체육관광부에서는 「국제회의산업 육성에 관한 법률」을 제정하는 등 국제회의산업을 중심으로 MICE산업을 육성하고 있다. 관련 법령을 살펴보면, 코로나19 이후 엔데믹을 맞이하여 국제회의의 소규모화 경향을 반영하여 국제회의의 기준을 개선(국제회의산업법 시행령 일부개정/22.12.27.)하였고, 관광산업 중 국제회의업에 기업회의를 포함(관광진흥법 및 국제회의산업법 일부개정/22.9.27.)하였다. 개정된 국제회의의 정의는 「국제회의산업 육성에 관한 법률 시행령」 제2조(정의)에 근거하고, ‘국제회의’란 상당수의 외국인이 참가하는 회의(세미나·토론회·전시회·기업회의 등을 포함한다)로서 대통령령으로 정하는 종류와 규모에 해당하는 것을 의미한다.

「국제회의산업 육성에 관한 법률 시행령」 제2조(국제회의의 종류·규모)에 따라 국제회의의 정책적 범주는 다음 [표 2] 중 하나에 해당하는 회의를 의미한다. 현재 우리나라가 활용하는 법정 용어인 ‘국제회의’라는 용어는 해외 참가자 유치에 집중하여 외화 획득에 주력하고, 단기간의 경제 활성화에 기여할 수 있는 관광 산업 육성의 측면에서는 의미 있는 용어이다.

[표2] 「국제회의산업 육성에 관한 법률 시행령」 제2조의 국제회의 유형

구분		내용
제1호	국제기구 및 국제기구 가입 기관·법인·단체 주최 회의	가. 해당 회의에 3개국 이상의 외국인이 참가할 것 나. 회의 참가자가 100명 이상이고 그 중 외국인이 50명 이상일 것 다. 2일 이상 진행되는 회의일 것
제3호	국제기구, 기관·법인·단체 주최 회의	가. 「감염병의 예방 및 관리에 관한 법률」 제2조제2호에 따른 제1급감염병 확산으로 외국인이 회의장에 직접 참석하기 곤란한 회의로서 개최일이 문화체육관광부장관이 정하여 고시하는 기간 내일 것 나. 회의 참가자 수, 외국인 참가자 수 및 회의일수가 문화체육관광부장관이 정하여 고시하는 기준에 해당할 것

(자료: 「국제회의산업 육성에 관한 법률 시행령」, 대통령령(제33127호), 2022년 12월 28일 시행.)



MICE산업 중요성

1) 고부가가치성 및 지역 경제 파급성

MICE산업은 대규모 외국인 관광객을 유치하여 호텔, 쇼핑 등 연관 산업의 경제 파급효과를 높이는 고부가가치 산업으로, 일각에서는 ‘굴뚝 없는 관광산업’으로 불린다. MICE 산업은 우리나라 관광산업 중 9.2%(3조 5,528억 원) 비중을 차지하는 등(문화체육관광부 2020), 높은 경제 파급효과 때문에 국가 전략산업으로 정부 정책차원에서 육성·진흥·촉진하고 있다.

국제회의산업은 MICE 산업의 각 분야 중 핵심적 산업이다. MICE 산업의 직접국내총생산(MDGDP) 3조 5,528억 원 중 국제회의산업이 68.5%(2조 4,335억 원) 비중을 차지하고 있고, 나아가 이외 생산유발효과 약 5조 9,268억 원, 소득유발효과 약 1조 3,678억 원, 수입유발효과 약 4,898억 원, 부가가치유발효과 약 2조 7,012억 원, 간접세유발효과 약 314억 원, 취업유발효과 53,007명, 고용유발효과 34,355명 등 타 분야 대비 경제적 효과가 높기 때문이다(한국관광공사, 2020).

한편, 국제회의에 참가하는 외국인 참가자의 경우 일반 방한 외국인 관광객과 비교하여 상대적으로 소비지출 규모가 크게 나타났다. 2019 MICE 참가자 조사(한국관광공사, 2020)에 따르면 Meeting과 Convention 부문의 외국인 참가자 1인 평균 소비 지출 금액은 약 2,797,084원으로 나타났으며, 이는 2019년 외래관광객조사(문화체육관광부, 2020) 기준 외래관광객 1인 평균 소비 지출 금액(약 1,445,799원) 대비 약 1.9배 높은 수준이다. 이처럼 MICE 행사 참가를 위해 방한한 외국인 참가자의 지출 비용이 일반 외국인 관광객 대비 높기 때문에 관광산업 중 MICE 산업이 지역 경제에 미치는 파급성이 높음을 알 수 있다.



2) 타 산업과의 연계 다양성 및 융복합성

UN Tourism(2008)은 국제회의산업의 구조를 크게 회의 공급 중심의 핵심산업(Core Industry)과 회의 수요 중심인 확장산업(Extended Industry)으로 구분하고 있으며, 각 핵심산업과 확장산업 내 개별 산업은 매우 다양하다고 주장하였다. 특히 국제회의산업이 발달한 국가나 도시 지역에서는 PCO(Professional Conference Organizer) 및 PEO(Professional Exhibition Organizer)와 같은 회의/전시 전문 기획업체, 회의 전문 시설업 등 핵심 산업뿐만 아니라, 관광산업에 포함된 교통, 숙박, 엔터테인먼트, 레저, 쇼핑 및 식음료 등의 분야가 집약적으로 발전한다.

또한, 통신, IT 및 인쇄출판 광고 등 다양한 확장 산업들이 유기적으로 연계되어 융복합 산업의 특성을 지니게 된다. 국제회의산업은 교류와 혁신을 추구하는 산업과 함께 성장하고 상생하는 것을 통해 고부가가치와 파급효과를 극대화할 수 있고 이를 위해선 '전문성'과 '협치성'이 필수적으로 요구된다.

국제회의를 개최하기 위한 전, 중, 후 과정에서 다양하고 복잡한 요소들을 통합적으로 기획, 준비, 운영 및 관리하려면 하드웨어, 소프트웨어, 휴먼웨어의 유치와 개발, 집적이 필수적이다. 또한, 국제회의산업의 내부 및 외부 이해관계자 간의 협력적 거버넌스 구축은 지속 가능하고 선순환적인 산업 발전에 매우 중요한 역할을 한다.

3) 산업 동반 성장 및 혁신성

국제회의산업은 경제, 사회, 정치, 문화, 기술 및 인프라 등의 다양한 외부 환경에 의해 직접적이거나 간접적으로 영향을 받는다. 이 산업은 국제회의의 개최 주제와 관련된 최신 지식과 정보를 집적하고, 새로운 지식의 생산과 유통을 촉진하며 융합하는 지식 기반형 산업이다. 각 관련 산업이 개별적이고 독립된 산업이기 때문에 산업 자체로

는 높은 부가가치를 가지고 있으며, 상호 의존성이 강하기 때문에 관련 연관 산업과의 통합을 통해 육성할 경우 시너지 효과를 쉽게 창출할 수 있는 종합 서비스 산업이다.

국제회의산업은 교류와 만남을 위한 플랫폼을 제공하고, 이를 통해 국제회의의 주제와 관련된 최신 정보와 지식을 집적한다. 새로운 지식의 생산 및 혁신을 촉진하고, 혁신적인 지식을 상품화하여 거래하고 유통하는 것을 돕는다. 이러한 과정에서 혁신성과 교류성이라는 특성을 지니게 된다.

국제회의 산업은 혁신적인 지식 교류와 기술 발전에 따라 회의 개최와 관련된 기획 및 관리 기술의 혁신과 도입이 활발히 이루어지고 있으며, 이를 통칭하여 마이스 기술(MICE Technology)이라 한다. 대표적인 예시로 코로나19 팬데믹으로 인해 직접 대면 만남이 어려워졌음에도 불구하고, 비대면 온라인 회의를 통한 만남과 교류는 계속되었다. 이 과정에서 온라인 회의 관련 MICE 기술의 혁신과 발전이 촉진되었다.

MICE 산업 관련 통계

현재 중앙정부에서는 국제회의의 관련 법률인 「국제회의산업 육성에 관한 법률」에 의거하여 국제회의의 관련 정책을 MICE 범위로 확대하고, MICE 유형에 따라서 관련 정책을 추진하고 있다. 또한, 이에 발맞춰 정책 및 마케팅 전략 수립의 기초자료를 확보하고자 한국관광공사와 지역관광공사에서는 MICE 산업과 관련된 통계를 생산하고 있다.

한국관광공사에서는 관련 통계로 공급자 측면에서 「MICE 산업통계 조사연구」, 수요자 측면에서 「MICE 참가자 조사」를 매년 수행하고 있다. 「MICE 산업통계 조사연구」는 첫째, 우리나라에서 개최되는 미팅(M), 인센티브 투어(I), 컨벤션(C)에 대한 행사 규모(개최 건수, 전체/외국인 참가자 수)를 파악하는 규모 조사와 둘째, 관련 산업을 영위하는 사업체를 대상으로 매출액과 근로자 수 등을 파악하는 사업체 조사로 구분된다. 모든 정책과 사업에 목표와 방향성이 있듯 동 조사는 국가승인통계 및 특수산업분류체계 지정 절차를 밟고 있는 등 장기적 로드맵을 세우고 통계 고도화를 하고 있다. 또한, 중앙 및 지방정부의 정책/사업 계획 혹은 민간기업의 유치/마케팅 전략 수립의 기초자료로 활용되고 있다.



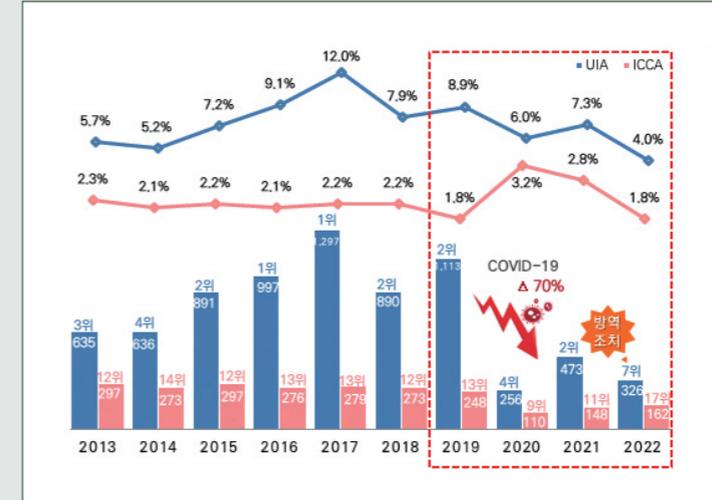
「MICE 참가자 조사」는 국내 미팅(M), 인센티브 투어(I), 컨벤션(C), 전시(E) 행사에 참가한 내/외국인을 대상으로 참가실태를 파악하는 조사이다. 참가자들이 왜 행사에 참가했는지? 출발지는 어디인지? 행사에서 무엇을 경

협하였는지? 행사에 지출한 비용은 얼마인지? 등 전반적 참가실태에 대한 분석을 통해 행사 고도화 혹은 참가자 증대방안 마련을 위한 마케팅 전략 수립의 기초자료로 활용되고 있다. 한편, 「MICE 산업통계 조사연구」의 공급자 비용과 「MICE 참가자 조사」의 소비자 지출비용을 토대로 MICE 산업에 대한 경제적 파급효과를 매년 산출하고 있고, 이에 따른 MICE 산업 육성의 당위성을 확보하고 있다.

한편, 한국관광공사에서는 매년 「국제회의 개최실적 조사」를 수행하고 있다. 동 조사는 우리나라에서 개최된 국제회의의 데이터를 수집하여 국제기구인 국제협회연합(UIA : Union of International Associations)과 국제컨벤션협회(ICCA : International Congress and Convention Association)에 제출하는 조사이다.

UIA와 ICCA에서는 한국을 비롯하여 전 세계에서 제출받은 국제회의의 데이터를 토대로 국가별 그리고 도시별 국제회의의 개최지의 순위를 발표하고 있다. 이러한 순위 결과는 국제회의의 개최지로서의 인지도 및 이미지를 제고하고 나아가 행사 유치마케팅 시 경쟁우위를 점할 수 있다는 장점이 있다. 한국의 지난 10년간의 순위를 살펴보면, UIA 기준¹⁾으로는 2016-2017년 세계 1위를 달성하고 이후 세계 5위권 이내에 위치하고 있으며, ICCA 기준²⁾으로는 세계 15위권 이내에 위치하는 등 한국은 세계적인 국제회의의 개최지이다([그림 1] 참조).

한국관광공사 외에 지역관광공사(서울관광재단, 인천관광공사, 부산관광공사, 광주광역시 관광공사 등)에서도 각 지역의 특성을 고려하여 MICE 산업과 관련된 통계를 생산하고 있다. 각 기관마다 수행하고 있는 통계는 유사한데, 전반적으로 관내 개최 MICE 행사 전수조사, 관내 MICE 산업 영위 사업체 정량 및 정성조사, 관내 MICE 행사 참가자 조사, 경제적 파급효과 등으로 구분된다. 각각의 조사들을 격년 주기로 옴니버스 방식의 조사를 진행



[그림 1] 한국의 UIA 및 ICCA 순위 추이(2013~2022)

하고 있다. 각 지역관광공사가 별도의 통계를 생산하는 주된 이유는 국가 단위에서 진행되는 MICE 통계는 표본 수가 적어 지역 단위로 분석하기에는 한계가 있어 각 지역의 특성에 부합되는 데이터를 수집하기 위함이다.

MICE 산업에 있어 통계는 연구개발(R&D) 투자이다. MICE 산업 육성과 시장 활성화를 위해 전시컨벤션센터, 특급호텔 등 전문 시설이 반드시 필요하다고 여기는 것처럼 산업 통계에서도 없�서 안 될 필수 정보 인프라로 봐야 한다. 그러나 MICE 산업에서 통계는 필요할 때만 하면 되는 일회성 과제, 예산 사정에 따라 늘리고 줄일 수 있는 후순위 사업으로 보는 인식과 관행이 있다. 통계는 시간, 데이터가 쌓이면 쌓일수록 더 유의미한 결과와 시사점 도출이 가능해진다. 당장 원하는 결과를 얻지 못하더라도 과거와 현재의 문제점을 파악하고 미래의 방향성을 설정하기 위한 도구로써 긴 호흡을 갖고 연속성과 지속성을 유지하는 것이 중요하다. 따라서 MICE 산업의 통계를 바라보는 시각과 관행 개선이 절실히 요구된다.



1) 국제기구에서 개최되는 회의 2) 3개국 이상을 정기적으로 순회하는 참가자 50명 이상의 학/협회 주관 회의

(참고문헌)

- 김성태·윤은주·정광민·김미성(2024), 비즈니스이벤트. 백산출판사.
- 김철원·허준(2011), 한국 컨벤션 산업의 역사적 고찰. 『관광학연구』35(10), 517-533.
- 문화체육관광부(2020), 2019 외래관광객 조사. • 문화체육관광부(2020), MICE산업 경제규모 추정 및 성장 지표 개발 연구.
- 문화체육관광부(2023), 포스트코로나 시대의 국제회의산업 진흥 방안 연구.
- 정광민(2020), 국제회의산업 정책추진 실태와 과제. 한국문화관광연구원. • 한국관광공사(2020), 2019 MICE 참가자 조사.
- 한국관광공사(2020), 2019 MICE산업의 경제적 파급효과 분석. • 황희곤(2010), MICE산업 기본법 제정 연구, 한국문화관광연구원.
- UN Tourism(2006), Measuring the Economic Importance of the Meetings Ind.
- UN Tourism(2006), Measuring the Economic Importance of the Meetings Industry.
- UN Tourism(2008), Global Meeting Initiative.

Open New Era!

첨단 기술 기반 스포츠 경기분석 시스템

천제민 | 스위트케이 상무이사



기술의 발전 속도가 빨라도 너무 빠르다

지금으로부터 약 8년 전인 2016년 1월 제47회 다보스 세계경제포럼에서 처음 거론된 4차 산업혁명이라는 용어는 단순한 사회경제적 현상을 설명하는 개념의 단계를 넘어 이제는 우리들 일상 깊은 곳까지 영향을 미치고, 삶의 다양한 측면에서 변화를 이끌어내고 있다.

스포츠 분야에서도 스포츠의 과학화, 스포츠 산업의 디지털 전환(Digital Transformation, DX)이라는 미명 아래 인공지능, 사물인터넷, 클라우드, 데이터 분석 등 디지털 기술을 기반으로 새로운 가치를 창출하기 위한 노력들이 경쟁적으로 시도되고 있다. 이러한 변화의 큰 흐름 속에서 스포츠 경기분석 분야에도 다양한 연구들이 시도되었고 많은 상용 시스템들이 출시되었으며, 최신 기술로 점차 고도화되고 진화하고 있다.

스포츠 경기분석이란, 스포츠 경기 중에 일어나는 여러 가지 현상들, 즉 현상이나 이벤트들을 주관적인 관점이 아닌 객관적인 관점에서 탐색, 관찰하여 이를 기록(Record)하고, 목적에 부합하는 방법론(Technology)을 통해 분석(Analysis)함으로써 최종적인 통찰력(Insight)을 얻는 일련의 프로세스라고 정의할 수 있다.

이러한 스포츠 경기분석을 하는 목적은 과학적이고 체계적인 기록관리를 통해 크게 선수들에 대한 훈련 계획을 세우고, 경기의 전술과 전략을 세우는 데 필요하며, 대중들에게 스포츠에 대한 관심과 재미를 더하도록 하는 데 있다고 할 수 있다. 최근 스포츠 경기분석의 기록(Record), 방법(Technology), 분석(Analysis), 통찰력(Insight)의 일련의 프로세스 과정은 4차 산업혁명의 핵심기술인 DNA(Data, Network, AI)의 기술에 힘입어 더욱 발전하고 있다. 이는 스포츠 경기를 분석하기 위한 기술이 단순한 데이터와 이벤트를 수집하는 환경에서 영상

과 데이터, 영상과 이벤트를 결합한 멀티미디어(Multi-media) 분석 시스템으로 발전해 가는 것을 의미한다. 이러한 멀티미디어 분석 시스템 구현이 가능하게 된 배경에는 컴퓨터(Hardware)의 발전은 물론 영상 추적 및 데이터 분석 기술이 더해져 가능하게 됐다.

첫째, 스포츠 통계 기법 활용(Sports Statistics Techniques)

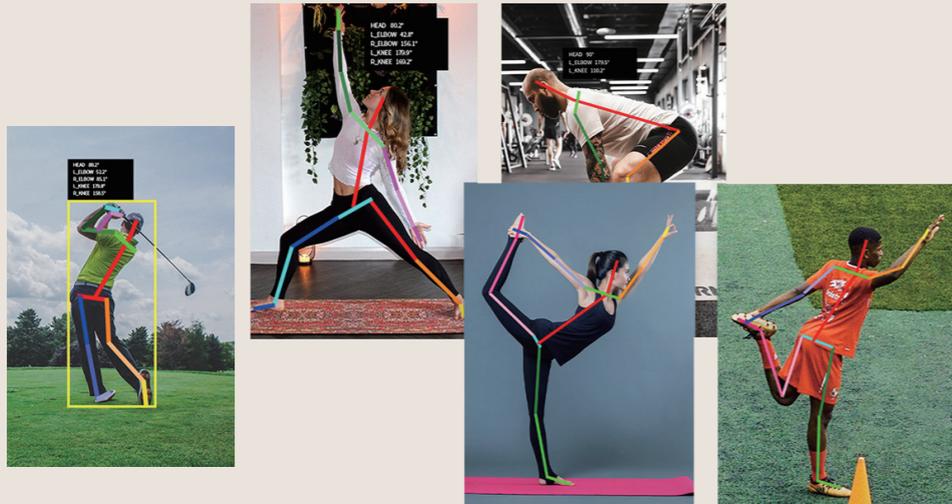
스포츠 경기분석에서 통계 기법을 활용하는 것은 그동안 “스포츠 경기의 주요 요인을 체계적으로 기록하고 통계적으로 분석하는 객관적인 방법” 중 가장 전통적인 방법으로 알려져 왔다. 댄스나 음악 분야에서 사용되고 있는 개성적이고 체계적인 기록 방법과 유사하게 스포츠 통계 분석은 스포츠의 주요 요인들을 체계적으로 기록하는 방법과 기록된 자료를 통계적으로 분석한다는 두 가지의 개념을 함께 내포하고 있다.

체계적으로 기록된 자료는 스포츠 현장에서 주로 팀 스포츠의 전술평가, 기술평가, 움직임 패턴이나 경로 등을 평가하는 자료로 활용될 수 있는데, 코치는 스포츠 통계분석을 통해 선수들에게 정확하고 객관적인 피드백을 즉각적으로(실시간) 제시할 수 있으며 선수들의 오류나 결함을 교정하기 위한 효과적인 자료로 활용할 수 있다. 실제 스포츠 과학의 중요성이 부각되면서 축구, 배구, 하키, 농구, 테니스 등 다양한 스포츠 현장에서 경기분석을 활용한 전술 및 기술평가 그리고 훈련 방법의 적절성에 대한 평가가 이루어지고 있으며, 스포츠 현장에서 직관적인 경험에 의해서만 지도를 하는 것이 아니라 과학적이고 객관적인 자료에 근거해 선수들을 지도하는 기반이 마련되었다.



(Source : 머니투데이)

우리에게 잘 알려진 스포츠 통계 분석의 대표적인 기법으로는 ‘세이버메트릭스(Sabermetrics)’라는 접근법이 있다. 세이버메트릭스란 시즌 동안 누적된 경기 기록을 바탕으로 선수의 기량을 평가하는 통계기법을 의미한다. 이 방법은 실화를 바탕으로 만들어진 소설과 영화 ‘머니볼’로도 유명한 미국의 빌리 빈(Billy Beane)이란 사람이 처음 도입하여, 만년 하위팀이었던 메이저리그의 오클랜드 애슬레틱스(Oakland Athletics)를 포스트시즌에 매년 진출하는 강팀으로 탈바꿈시키면서 선풍적인 인기를 끌었던 통계 기법으로 알려져 있다.



(Source : DSLAB GLOBAL)

둘째, 영상을 이용한 객체 추적 기술(Object Tracking Technology)

영상을 이용한 객체 추적 기술이란 ‘영상으로부터 일련의 데이터 처리 과정을 통해 원하고자 하는 객체를 추적하는 방법’을 의미하며, 접촉식 센서나 하드웨어를 부착하는 형태가 아닌 영상으로부터 데이터를 획득하기 때문에 선수 또는 경기 자체에 어떠한 부정적인 영향도 주지 않는다는 점이 특징이라고 할 수 있다.

시각적인 정보를 전달해 줄 수 있는 영상은 기존의 다른 방식들에 비해 효과적인 정보 전달이 가능한 수단으로 스포츠 현장에서 주로 사용되고 있으며, 영상처리 기술의 발전과 더불어 스포츠 분석은 영상을 이용한 객체추적 기술을 적용하여 분석이 이루어진다.

영상으로부터 얻어진 선수 추적 데이터를 활용하면 선수의 뛰 거리, 이동 위치, 움직인 위치와 시간 등의 다양한 정보를 얻을 수 있으며, 스포츠 영상에서의 객체 추적 기술은 대표적으로 축구 종목 이외에 최근에는 야구, 농구, 배구, 탁구, 테니스, 골프 등 다양한 종목에 적용되고 있다.

객체 추적(Objects Tracking)은 움직이는 물체를 검지하고, 자동 추적 기능이 있는 시스템으로 군사용, 공공용으로 활용되다가 최근에 스포츠 분야의 선수(Player), 공(ball) 등을 추적하는 데 응용되고 있다.

이러한 영상분석 기법은 최근에 인공지능(AI) 기술이 접목되면서 공간 인지, 속도 인지, 선수 관절의 움직임을 파악하고 선수의 움직임을 통해 어떠한 기술이 적용된 동작인지를 자동으로 분석해 내는 기술로 발전하고 있다.

선수의 관절 정보 x, y 좌표값을 추출하는 스켈레톤 모션 트래킹(Skeleton Motion Tracking) 기술은 다양한 영상 수집을 통해 선수의 기술, 공간의 좌표, 볼 궤적, 스피드, 몸의 회전 등의 관절의 움직임에 따른 공통적인 정보를 반복적인 기계학습으로 영상클립을 편집하고 동작, 공간, 상황별로 자동 DB화 할 수 있는 기술을 의미한다.¹⁾

스포츠 분야 중 축구 경기의 객체 추적은 Particle filter라는 알고리즘을 주로 사용하는 것으로 알려져 있다.

셋째, 인공지능 알고리즘 기술을 활용한 경기력 탐색(AI Algorithm Technology)

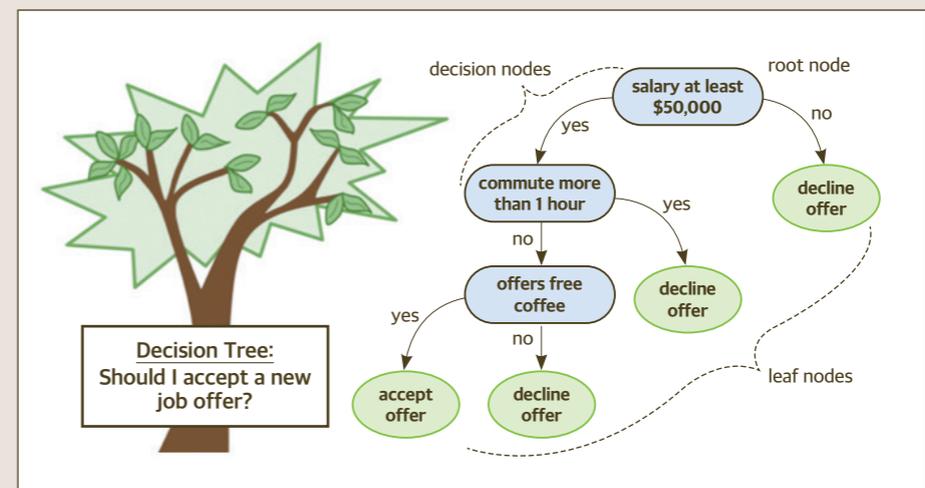
스포츠에서 인공지능(AI) 알고리즘을 적용해서 얻게 되는 가장 큰 장점은 전문가에게 의존하지 않고 경기를 분석할 때 객관성을 얻을 수 있다는 점을 들 수 있다.

인공지능(AI) 알고리즘 기술을 활용한 경기력 탐색 기술은 크게 의사결정트리를 활용한 경기력 탐색, 텍스트마이닝을 통한 이벤트 추출, 연관규칙 마이닝, 선수 간 유사도 분석으로 구분할 수 있다.

① 의사결정트리(Decision Tree)를 활용한 경기력 탐색

단순한 계산 공식이 아닌 좀 더 복잡한 문제에 있어서 널리 쓰이는 방법 중 결정트리가 있는데, 이는 학습된 개념을 가장 잘 나타내 줄 수 있는 속성으로 만들어지며, 훈련 집합에서 일부 데이터 집합을 선택하여 결정 트리를 만들게 된다.

결정트리(Decision Tree)는 의사 결정 규칙과 그 결과들을 트리 구조로 도식화한 의사 결정 지원 도구의 일종이다. 결정트리는 운용 과학, 그 중에서도 의사 결정 분석에서 목표에 가장 가까운 결과를 낼 수 있는 전략을 찾기 위해 주로 사용된다.²⁾ 데이터들이 모두 정확하게 분류된다면 분류 과정이 끝나고 분류 과정이 끝나지 않았다면 새로운 데이터를 추가하여 새 결정트리를 만들게 된다.



(Source : www.packtpub.com)

② 텍스트 마이닝(Text Mining)을 통한 이벤트 추출

대규모 문서정보로부터 암시적이고 이전에 알려지지 않았지만 잠재적으로 유용한 패턴을 발견하는 과정을 의미한다. 인터넷에서 쉽게 볼 수 있는 스포츠 문자중계를 통해 특정 선수들의 패턴이나 관계를 분석할 수 있으며, 이

1) ㈜스다, 2024 [http://home.sportsdiary.co.kr] 2) 위키백과, [https://ko.wikipedia.org/wiki]

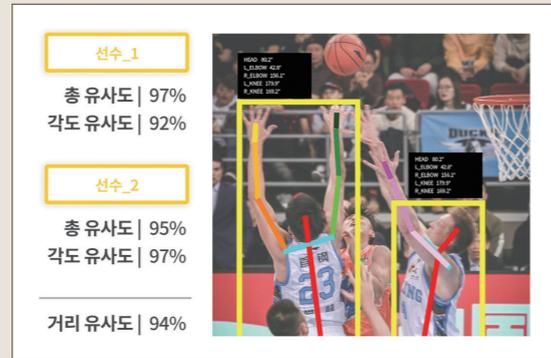
러한 텍스트 정보를 기반으로 키워드 매칭을 통해 경기의 다양한 속성을 추출하는 기술이다.

최근에는 스포츠 복권과 미디어 및 정보통신 기기가 발달하면서 스포츠 빅데이터 기반의 텍스트 마이닝을 이용한 스포츠 경기 예측 분석 서비스에 대한 수요가 커지고 있다.

이는 스포츠 경기에서 변수라고 할 수 있는 선수, 감독, 선수나 경기의 여건, 날씨, 외부 환경 등을 분석하여 경기 결과를 예측하는 것으로서 보다 정확하고 객관적인 판별을 제공하기 위한 기법들에 대해 대중적인 관심과 수요가 커지고 있음을 의미한다.

③ 연관규칙 마이닝(Association Rule Mining)

연관규칙(Association Rule)이란, 데이터 항목들 간에 자주 발생하는 객체들을 찾아내어 그들 사이의 연관관계를 규칙으로 표시하는 기법이다. 어떤 선수가 특정 행위를 할 때 발생하는 또 다른 행위를 찾음으로써 선수의 행동을 미리 예측하여 효과적인 전략을 세우는 데 도움을 받을 수 있다.



(Source : DSLAB GLOBAL)

④ 선수 간 유사도 분석(Similarity Evaluation)

팀 내 선수 간의 유사도를 알아보기 위해서는 선수 별 어떤 이벤트를 유형별로 분석하여 수행할 수 있다. 여기서 이벤트라 함은 경기 중 선수들에게 발생하는 하나의 사건을 의미하여 농구 경기의 경우 득점, 3점 슈트, 리바운드, 어시스트, 인터셉트, 파울 등을 말한다.

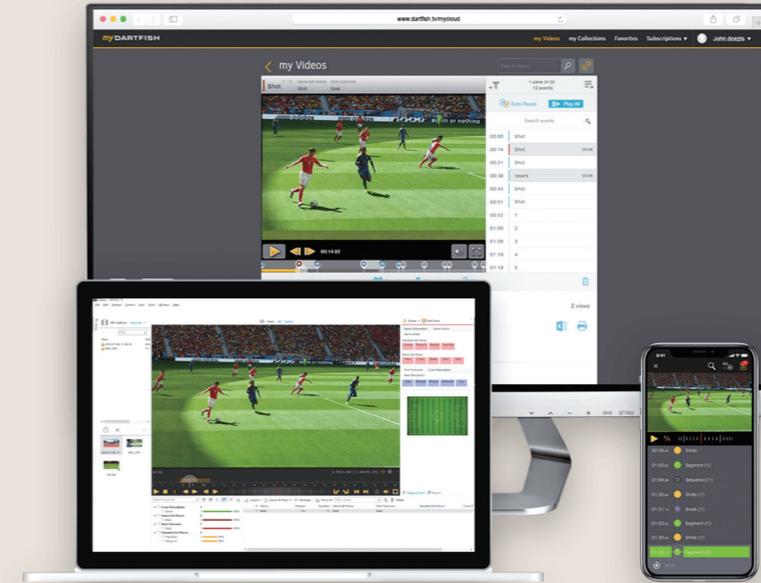
넷째, 3D 입체영상 분석 기술(3D Stereoscopic Video Analysis Technology)

인간이 두 개의 눈으로 입체감을 지각하는 원리와 유사하게 두 개의 카메라 렌즈를 통해 좌우 영상을 동시에 촬영하고, 여러 가지 종류의 방식으로 이를 재생함으로써 이루어지며, 입체영상의 재생 방식은 크게 안경 방식과 무안경 방식으로 구분한다.

첨단 기술 기반 스포츠 경기분석 시스템 사례

(1) 다탈피쉬(Dartfish), 스위스

선수의 동작 분석을 위한 대표적인 소프트웨어 도구로서 슬로우모션을 이용한 영상분석 기술을 기반으로 한다. 10가지의 모듈을 제공하며, 라이브러리, 태깅, 플레이어, DV영상입력, 애널라이저, 인디액션, 사이멀 캠, 사이멀 캠 픽스드, 스트로모션, 스트로모션 픽스드가 있다.



(Source : Dartfish.com)

하나의 PC로 최대 4개의 카메라를 이용해 경기분석을 진행할 수 있으며, 자체 3D 도면 기능은 보다 정확한 측정과 정교한 데이터 산출을 지원한다. 또한, 라이브 컬레버레이션 기능을 통해 동시에 여러 사람이 분석에 참여하여 실시간으로 관련 내용을 공유할 수 있고, 즉각적인 전략 조정을 가능하도록 한다.

그 밖에 사이멀캠 기능과 스트로모션 기능, 데이터 스트림 통합, 통계 작성, 보고서 작성 등으로 다양한 영상 분석을 진행할 수 있다.³⁾

(2) IBM의 슬램트래커(SlamTracker), 미국

IBM의 분석 기술을 이용해 실시간 경기 상황을 보여주는 플랫폼으로서 8년간 테니스 그랜드 슬램 데이터를 분석해 각 선수의 스타일과 패턴을 식별하고 경기 시작 전 해당 선수들 간의 과거 경기 결과를 분석하는 빅데이터 분석 기반의 인공지능 경기분석 소프트웨어다. 경기 중 승리 요인에 대한 각 성과를 측정 웹사이트에 실시간 업데이트를 한다.

슬램트래커의 모멘텀(Momentum) 기능은 실시간 경기 진행 상황을 한눈에 볼 수 있도록 정리해 주는 서비스이다. 경기와 관련된 주요한 상황과 전환점이 이루어진 배경을 시각적으로 분석할 수 있도록 해준다. 또한 키투더매치(Keys to the Match) 기능은 과거 대회 기록과 실시간 현재 데이터를 동시에 활용하여 특정 시합에 임하는 선수가 최우선으로 해야 할 3가지 임무가 무엇인지 제시한다.⁴⁾

3) Dartfish Korea, [https://www.dartfish.co.kr] 4) GameMeca.com, [https://www.gamemeca.com]



(Source : <https://erpreadyblog.wordpress.com>)



(Source : german-design-award.com)

(3) SAP의 매치인사이트(Match Insight), 독일

2014년 브라질월드컵에서 우승한 독일 축구대표팀 선전에는 실시간 빅데이터 분석 기술의 뒷받침이 있었다. 독일 축구대표팀은 독일 대표 정보통신(IT) 기업 SAP과 손잡고 경기력 향상을 위한 데이터 분석 기술을 활용하기 시작했다. 당시 독일 축구대표팀 코치의 부탁으로 개발한 SAP 매치 인사이트는 선수들 움직임을 실시간으로 분석해 전략을 짤 수 있도록 했다. 훈련 경기 중 골키퍼는 양쪽 무릎과 양손, 어깨에 총 6개의 센서를 달고 나머지 선수들은 양쪽 무릎과 어깨에 4개의 센서를 부착했다. 센서 1개당 전송되는 데이터는 1분에 1만2,000여 개. 90분 경기당 총 5,000만개 데이터가 생성되는 셈이다.

이러한 센서를 통해 수집된 운동량, 속도, 심박수, 슈팅 동작, 공의 방향 등 대규모 데이터가 SAP 매치 인사이트를 통해 실시간 분석된다. 분석 결과는 감독과 선수단의 태블릿PC로 실시간 전송돼 신속한 의사결정을 할 수 있도록 인사이트를 제공했다.⁵⁾

(4) Pa-Sport의 액팀 인덱스(Actim-Index), 영국

Actim Player Rankings 2006/07 - Top 100		
As at: Tue 15 May 07 09:06		
Name	Team	Index Score
1 Didier Drogba	Chelsea	628
2 Cristiano Ronaldo	Manchester United	617
3 Dimitar Berbatov	Tottenham Hotspur	611
4 Wayne Rooney	Manchester United	580
5 Frank Lampard	Chelsea	572
6 Mikel Arteta	Everton	567
7 Cesc Fabregas	Arsenal	565
8 Gareth Barry	Aston Villa	561
9 Benni McCarthy	Blackburn Rovers	535
10 Gabriel Agbonlahor	Aston Villa	517

(Source : fantasyfootballscout.co.uk)

5) 매일경제, [<https://www.mk.co.kr/news/ports/6184866>]

액팀 인덱스는 프리미어십과 챔피언십 공식 선수 평가 지표다. 액팀은 잉글랜드 프리미어 리그, 풋볼 리그, 스코틀랜드 프리미어 리그 공식 데이터의 브랜드 이름이며, 해당 데이터는 경기 중 발생하는 골, 유효 슈팅, 코너킥, 파울, 오프사이드, 패스, 태클, 차단, 클리어 등 모든 이벤트들로부터 추출된다.

액팀 인덱스는 경기에서 핵심적으로 기여를 한 선수가 누구인지 밝히고, 포지션이나 국적, 연령 등에 관계없이 과학적으로 선수들의 랭킹을 매겨, 팬들에게 최고의 선수가 누구인지 알려줄 목적으로 만들어졌다. 살포드 대학(University of Salford) 통계학자들에 의해 고안되어 2004-2005년 시즌부터 프리미어 리그와 챔피언십에 도입되었다.⁶⁾

(5) Sky Sports의 Player Rating, 영국

Sky Sports는 영국의 Sky Limited가 운영하는 텔레비전 채널로, 1991년 개국하였다. 원래는 Sky 텔레비전과 BSB(영국위성방송)의 합병 이전, BSB 계열의 The Sports Channel이 존재했으나 합병 이후 Sky Sports로 채널명을 바꿨다.⁷⁾

스카이 스포츠는 축구, 골프, 럭비, 크리켓, F1, 복싱, NFL, NBA 등 인기있는 스포츠 종목에 대한 라이브 중계와 최신 스포츠 뉴스를 방송하는 채널로서, 다양한 스포츠 선수들의 통계와 기록을 수집하여 이를 각 선수별 랭킹으로 표현하기 위한 분석 알고리즘을 활용했다.



(Source : skysports.com)

(6) 프로존(Prozone), 영국

Prozone은 영상 및 데이터 분석 기술을 통해 선수와 팀의 성과, 전략에 대한 인사이트를 제공하고, 농구, 럭비, 축구, 테니스 등 다양한 스포츠 종목을 포함한다.

Prozone의 분석 서비스는 선수 추적(Player Tracking), 전략과 전술의 분석 등이 있고, 경기장이나 코트에 있는 모든 선수의 움직임을 추적할 수 있기 때문에 선수들의 움직임 하나 하나를 자세히 분석할 수 있다. 또한 Prozone은 코치 등 지도자들이 데이터를 이해하고 이를 통해 선수의 선발과 경기 전략을 세우는 데 필요한 보고서를 지원하고 시각화 기능을 통해 차별화된 서비스를 제공한다.



(Source : sportsbusinessjournal.com)

6) 네이버 블로그, [<https://blog.naver.com/topquark>] 7) 나무위키, [<https://namu.wiki>]

빅데이터와 AI시대에 살고 있는 우리들에게 절실히 필요한 직관력 및 상상력, 독서량

김선일 | 빅헤브솔루션(주) 대표이사



필자는 보통 아침 6시30분쯤 집을 나서 자가용으로 출근한다. 운전석에 앉은 후 스마트폰으로 전날 유럽 및 미국에서 일어난 경제 뉴스를 듣기 위해 유튜브를 실행한다. 이때 똑똑한 유튜브는 빅데이터 분석을 통해 내가 매일 같은 시간대에 보는 ‘한국경제TV #당잠사’를 상단에 표시해 준다. 이미 유튜브 백엔드(backend)에서는 필자의 행동 패턴을 빅데이터 분석 완료했고, ‘추천 콘텐츠’를 제공한 것이다. 이는 참으로 편리하면서도 무서운 세상인 듯하다. ‘한국경제TV #당잠사’가 상단에 표시되는 것을 기술적으로 분석해 보면, 유튜브는 필자의 행동 패턴 데이터를 일정수준 쌓은 후, 이 데이터를 정제 및 분석한 다음, 분석 결과 정보에 기초해서 필자의 행동을 예상하고 선택 가능성이 높은 콘텐츠를 우선적으로 제공해 준 것이다. 그리고 그 예측이 맞다면, 가중치를 더해 다음 시간대에도 동일한 콘텐츠를 제공할 것이다.

데이터와 정보, 그리고 잘못된 데이터의 실패 사례

‘정보(情報, information)’는 수집한 자료(데이터, data)를 실질적으로 도움이 되도록 정제된 유의미한 지식 결과물이라 할 수 있다. 다르게 말하면, 원시(raw) 데이터로부터 일련의 과정을 거쳐 얻은 결과물이 ‘정보’라 할 수 있다. 그렇다면 자료(데이터)를 수집하고 정제하고 시각화하는 과정을 왜 하는 것일까? 이러한 지루한 빅데이터 분석 작업을 수행하는 최종 목표는 지금보다 더 나은 결과물을 도출하기 위한 의사결정을 내리기 위함이다. 그러나 반대로 잘못된 데이터나 정보가 주어진다면, 잘못된 의사결정을 불러올 수 있다.

실제로 구글에서 데이터를 잘 못 활용해, 실패한 사례가 있다. 구글의 온라인 전염병 관리·예방 시스템인 ‘구글플루(Google Flu)’는, 감기에 걸린 환자가 입력한 검색어로 독감 확산을 예방하는 서비스였다. 하지만 ‘구글플루’의 예측 정확도는 구글의 주장과 달리 높지 않았다. ‘구글플루’가 독감과 관련하여 수집한 빅데이터가 실제 독감 전염과 관련성이 적었기 때문에 이러한 오류가 발생했다고 많은 전문가들이 지적했다. 결국 구글은 2015년 ‘구글플루’ 서비스를 중단할 수 밖에 없었다.

데이터 기반 비즈니스 경영

이번에는 비즈니스 경영 분야에서 빅데이터 성공 신화로 널리 알려진 사례를 살펴보자. 2011년 개봉한 영화 『머니볼』을 기억하는 독자가 많을 것이다. 이 영화는 대표적인 데이터 기반 비즈니스의 성공 사례로 볼 수 있는 메이저리그 야구 구단의 성공 신화를 다루고 있다. 영화 『머니볼』의 스토리 핵심은 다음과 같다. 자본이 충분하지 않은 최하위 야구 구단(오클랜드 애슬레틱스)에서 예일대 경제학과 출신 통계 전문가를 고용한다. 이 전문가가는 선수들의 기록 통계 가운데, 다른 사람들이 잘 알지 못했던 ‘출루율’이 중요한 요소임을 깨닫고, 출루율이 높지만 저평가된 선수들을 영입한다. 구단 고위 임원들은 자신들의 전통적인 방식과 다른 데이터 기반 의사결정에 초기에는 반대했으나, 결국 이를 수용하게 되고, 이로써 2002년 메이저리그에서 20연승을 달성하게 된다.

또한, IT 대기업 가운데 빅데이터 분석으로 큰 성장 모멘텀을 만든 회사는 어디일까? 심중팔구 미국 아마존(Amazon)을 떠올릴 것이다. 아마존은 데이터 분석 기술에 기초한 ‘구매 추천 서비스’, ‘재고 관리 시스템’, ‘예측 배송 시스템’을 구현해 매출이 급증할 수 있었고, 아마존의 데이터 기술 활용 성공 사례는 여러 IT 대기업에 큰 자극을 주었다고 해도 과언이 아닐 것이다.



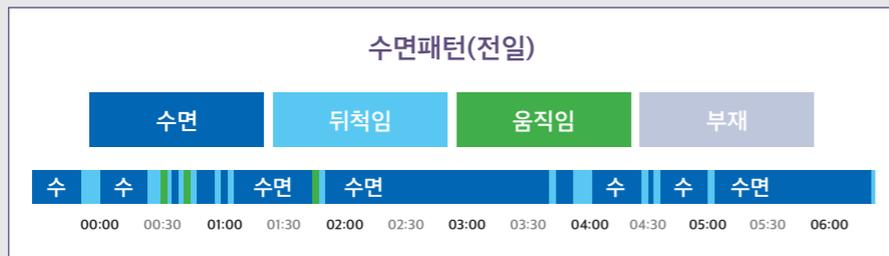
과거 데이터가 부족한 시대에는, 기업에서 주요 의사결정이 고위직 임원들의 직관력(감)이나 경험, 연륜에 의지하는 경우가 많았다. 그러나 최근에는 데이터를 수집하고 정제한 후, 의사결정하는 것이 주류가 되었으며, 기업 고위직 임원들도 데이터 기반 의사결정에 반박하기 쉽지 않다. 데이터가 비즈니스 경영의 중요한 기준이 되었다. 다만 데이터 수집량이 부족하거나 데이터 해석이 어려운 경우에는 초인적인 직관력과 통찰력, 경험, 연륜이 요구된다. 특히 이는 학습하지 않은 낯선 데이터를 다룰 때 더욱 그렇다.

디지털 헬스케어 기술 현황 및 향후 전망

2008년 핏비트(fitbit) 출현 이후, 디지털 헬스케어 시장이 개화되었다. 핏비트는 스마트 밴드로부터 얻은 디지털 헬스케어 정보에 기초해, 사용자의 건강, 운동량, 수면 패턴 등을 분석하여 서비스를 제공한다. 이 가치를 인정한 구글은 핏비트를 2.1억 달러에 인수했다.

이후 핏비트를 시작으로 수많은 디지털 헬스케어 기업이 등장하고 있다. 대표적으로는 △웨어러블 디바이스로 애플워치, 갤럭시워치를 비롯한 스마트 워치와 밴드, 스마트 벨트를 생산하는 웰트, △매개형 디바이스로 일론 머스크가 2016년 설립한 뉴럴링크(Neuralink)와 같은 BCI(Brain Computer Interface, 뇌 컴퓨터 인터페이스) 칩, △삽입형 디바이스로는 전자태그(RFID)를 내장한 작은 마이크로 칩이 있으며, 이는 주로 사람보다는 동물 특히 애완동물 체내에 삽입되어 관리되고 있다. △섭취형 디바이스로는 의료용 캡슐 내시경인 필캠(PillCam)을 들 수 있는데, 필캠은 10시간 동안 초당 35프레임으로 촬영 및 전송이 가능하다. △그 외로는 수면분석에 특화된 슬리프 사이클(Sleep Cycle), 솜즈(Somzz), 사이렌케어(SirenCare) 등이 있다.

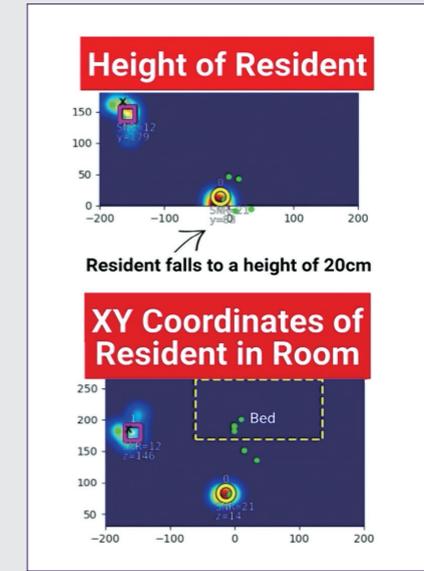
필자가 개발한 「사이렌케어」는 레이더(Radar) 센서를 활용하여 사람의 움직임과 운동량을 측정하고, 이를 빅데이터 분석하여 '낙상 감지' 및 '수면 패턴 분석' 서비스를 제공한다. 사이렌케어는 웨어러블의 불편함을 해소하는 차별화된 기술을 제공한다. 빅데이터 분석은 주로 Python pandas와 scikit-learn로 활용하며, 파일 내용의 정제 등은 awk, bash, perl 등을 사용한다. 데이터 시각화는 주로 matplotlib을 사용하며, 지도 시각화에는 Folium 라이브러리를 사용한다. 그리고 올해부터는 기존의 빅데이터 분석 기반 낙상 감지 및 수면 패턴 판별 알고리즘을 기계 학습(Machine Learning, 머신러닝)을 활용한 AI모델로 대체하는 시도가 진행 중이다.



[그림 1] 사이렌케어 앱에서의 수면 패턴 타임라인



[그림 2] 사이렌케어 앱에서의 수면상태 누적 차트



[그림 3] 사이렌케어 레이더 센서로 사람의 위치 및 낙상, 움직임 정도 판별

최근 불면증환자가 급증하여, 헬스케어 빅데이터 분석 기술이 수면 분석에 많이 활용되고 있는데, 우선 수면의 특징에 대해서 살펴보자. 수면은 포유류와 조류에서 확인되며, 특히 인간의 수면은 논렘(NON-REM)수면과 렘(REM)수면이 약 90분 간격으로 4~5회 정도 반복한 후 잠에서 깨어난다. 그래서 흔히 논렘수면을 깊은 잠, 렘수면을 얇은 잠이라고 부른다. 수면의 75%는 논렘수면이며, 나머지 25%는 렘수면이다.

그리고 우리들은 '꿈'을 렘수면 동안에만 꾸는다고 알려져 있다. 정확히 말하면 '꿈은 렘수면 중에 뇌가 활동하기 위해 일어나는 일종의 환각'이라 할 수 있으며, 이때 우리 몸이 완전히 이완되어 있지 않기에 종종 움직임이 발생한다. 또한, 논렘수면 중에 뇌의 노폐물이 처리된다고 하니, 잠을 잘 자는 것이 건강과 일상생활에 꼭 필요한 활동임을 잊지 말자. 한편 몽유병은 3~4단계 논렘수면에서 발생하는데, 논렘수면은 전술한 바와 같이, 꿈을 꾸지 않기 때문에 꿈과는 상관없다.

수면은 뇌의 기능을 회복하는 작용이 있다. 평균적으로 7시간 수면을 취하는 사람이 가장 장수한다고 알려져 있으며, 최근 국내에서도 불면증 환자가 110만명에 달할 정도로 급증하고 있는데, 특히 60대 환자(약 25만명)가 차지하는 비중이 높다. 이는 생리적인 요인 뿐만 아니라 경제적 요인도 큰 원인으로 지목된다. 참고로 신생아의 수면시간은 16~18시간, 사춘기 때는 약 8시간, 60대 이후가 되면 4단계 깊은 논렘수면이 거의 사라진다고 알려져 있다<참조: 수면의 과학, 사쿠라이 다케시>.

필자가 개발한 수면 패턴 빅데이터 분석을 통해 확인할 수 있는 점은 다음과 같다. 특정 개인의 수면 데이터를 분석하면 일정한 패턴이 나타나는데, 일정한 시간에 수면을 취하고 기상을 하며, 일정한 뒤척임 움직임량과 패턴이 분석된다. 그리고 이러한 패턴은 보통 일정한 범위 내에 수렴되는 경향이 있으며, 게다가 빅데이터 분석을 통해 이러한 경향이 연령별이나 거주지역별로도 유사성을 띄는 것으로 확인되었다. 물론 더 상세한 개인정보를 얻을



경우, 더 세밀한 그룹화와 범주화가 가능할 것으로 생각된다.

이와 같이 건강의 기초가 되는 수면 패턴을 빅데이터 분석을 통해 분석한다면, 건강 지표와 대체 방안을 마련할 수 있으며, 더 건강한 일상생활을 유지할 수 있을 것으로 기대된다. 또한, 낮에 신체와 뇌를 많이 사용할수록, 해당 영역의 뇌가 깊은 수면을 취한다고 한다. 그러므로 불면증이 있다면, 술, 담배, 카페인을 피하고 낮에 활동을 많이 하는 것을 추천한다.

의료 기술과 AI 기술의 발전으로 기대 수명이 증가하면서 노인 인구의 점진적인 증가와 노인 인구 비율 증가는 피할 수 없는 현실이 되었다. 디지털 헬스케어 빅데이터 기술과 AI 기술은 이러한 상황을 지혜롭게 대처하는 훌륭한 수단으로 기대받고 있다. 그러나 유럽과 미국과는 달리 일본과 특히 한국은 노인 빈곤율이 높아 노년층에서 헬스케어에 투자하는 비율이 낮다. 예를 들어, 사이렌케어의 낙상 감지 센서 제품은 북·서유럽에선 이미 3만대 이상 판매되었는데, 이는 많은 노년층이 자신의 건강과 안전을 위해 헬스케어 제품 도입에 적극적이라는 점을 보여준다. 한편 일본 및 한국에서는 판매량이 아직까지 낮은 편이다. 이는 북·서유럽의 이민법 강화로 인한 인건비 상승 및 일손 부족 등의 요인에 따라 헬스케어 제품 및 서비스의 필요성을 인식하는 데 차이가 있기 때문일 것으로 보인다. 특히 아시아 시장에서는 노년층의 경제적 요인이 큰 역할을 하는 것으로 판단된다. 하지만 이 문제 또한 점진적으로 일손 부족, 인식 개선, 소비력 향상과 함께 차츰 해결될 것으로 기대된다. 미국의 401K 연금 제도와 같이 혁신적인 은퇴 자금 마련 제도가 국내에도 하루빨리 마련되어야 할 것이다.

또한 최근 흥미로운 점은 AI 기술이 신약 개발에도 활용되고 있으며, 전문가들은 이를 통해 신약 개발 시간을 획기적으로 단축시킬 수 있을 것으로 예상하고 있다. 의료 분야에서는 또한 협동 로봇 및 온디바이스 AI의 확산으로 의료 질이 향상되고 있어, 이는 인간의 수명 연장에 큰 기여를 할 것으로 예상된다.

위와 같이 빅데이터 분석 기술과 AI 기술의 발전으로, 디지털 헬스케어 기술은 점진적으로 발전할 것으로 예상되며, 이에 따라 인간의 건강 증진과 수명 연장이 기대된다. 또한, BCI 기술과 로봇 기술 발전으로 SF영화에서처럼 인체 일부를 기계로 대체할 수 있는 날이 멀지 않을 것으로 예상된다.

피할 수 없는 AI시대

마크 저커버그 메타 CEO는 2022년 11월 4일 메타 주가가 90.78달러로 최저점을 찍은 이후, 효율성을 강화하기 위해 약 1년간 2만명 정도의 대규모 해고를 실시했고, 그 비용을 AI에 집중적으로 투자한 결과, 매출이 두 자릿수

로 증가하고 순이익은 69% 증가했다. 이는 다른 기업들에게도 강한 영향을 미쳐 M7(애플·마이크로소프트·알파벳·아마존·테슬라·엔비디아·메타)를 위시한 주요 대기업에서도 인력 구조 조정을 실시해, AI로 대체함으로써 비용을 절감하고 매출을 증가시켰다.

한편 최근 AI분야에서 가장 핫한 회사인 엔비디아의 젠슨 황 CEO는 “AGI(인공일반지능; 인간과 유사한 다양한 작업을 수행할 수 있는 AI)가 5년 내에 실현될 것으로 예상된다”고 발언한 바 있다. 이는 곧 사람과 AI가 서로 협조하고 경쟁하는 세상이 도래할 것으로 예상되며, 앞으로 데이터 수집과 의사결정이 더욱 가치있는 일이 될 것임을 시사한다. 특히 기업 경영진과 정부 고위 관료들은 반드시 이러한 능력을 갖추는 데 도전해야 할 것이다.

AGI가 출현하더라도 분명한 사실은 인간만이 가진 ‘상상력’, ‘직관력’, ‘무의식’, ‘잠재의식’ 등의 능력을 AGI가 완벽하게 실현할 것으로는 생각되지 않는다. 그리고 아직 우리 인간은 ‘꿈’과 ‘무의식’의 본질에 대해 정확하게 이해하지 못하고 있으므로, 당분간은 분석가의 ‘상상력’, ‘직관력’, ‘잠재의식’, ‘인지력’, ‘예지력’ 등의 능력은 시장에서 필요로 할 것이다.

수확체감의 법칙

보통 사람들은 ‘미루는 버릇’이 있으며, 특정한 상황에서 무엇을 해야 할 지 결정하지 못할 때, 흔히 데이터를 추가로 수집하려는 경향이 있다. 그러나 데이터가 지나치게 많아지면 ‘과적합’ 문제가 발생할 수 있으며, 이는 또 다른 기회비용을 낭비하게 된다. 따라서 데이터 분석을 진행하면서 일정한 윤곽이 드러나고 의사결정이 확실해 진다면, 굳이 추가 데이터를 수집할 필요는 없을 것이다.

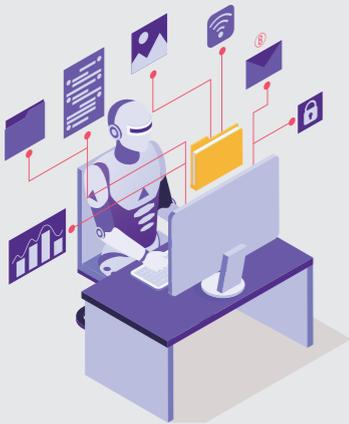


기계 학습(Machine Learning, 머신러닝)에서도 이와 유사한 문제가 발생한다. 데이터가 부족하면 제대로 학습하지 못해 ‘과소적합(underfitting)’ 현상이 나타나고, 반대로 데이터가 과다하거나 오류가 많으면 너무 편향되어 ‘과대적합(overfitting)’ 현상이 발생한다. 따라서 이 두 현상의 중간인 ‘최적적합’을 찾아야 한다.

경제학에서 ‘수확체감의 법칙(law of diminishing returns)’이라는 이론이 있다. 이론에 따르면 반드시 수많은 데이터를 수집한 이후에 행한 데이터 분석이 최상이고 가장 경제적이라고 할 수는 없다. 대다수 사람들은 타인의 비난을 의식해 완벽한 결정을 내리는 일에만 집착한 나머지, 스스로 압도당해 의사결정을 내리지 못하는 경우가 많다. 미국 아마존에서는 제품 기획 단계에서 80% 정도 기획안이 준비되면 사업화를 추진하는 경향이 있다. 나머지는 일을 추진하면서 임기응변으로 문제를 해결하는데, 이는 ‘수확체감의 법칙’을 경험 등으로부터 깨달은 결과이다. 콜린 파월 전 미국 국무장관 또한 “100퍼센트 확신이 느껴지기엔 충분한 정보가 모일 때까지 기다리지 마라. 그때 즈음이면 항상 너무 늦기 때문이다. 40 ~ 70 퍼센트의 정보를 가지고, 직감(instinct)을 믿고 나아가라”라고 말 한 바 있다. <참조:https://govleaders.org/powell.htm>

모든 의사 결정은 불완전한 정보를 기반으로 이루어지지만, 올바른 의사결정을 내릴 수 있는 어느 정도 충분한 정보만 수집하고 결정을 내린 후 나아가고, 또 다른 문제에 직면하게 될 때에는 침착하게 상황을 파악하고, 임기응변으로 그 상황에 맞게 올바른 의사결정을 하면 될 것이다. 손자병법에서도 완벽한 계략의 지구전보다는 졸속이어도 속전속결로 승부를 보는 것을 권장하고 있지 않은가?<참조:손자병법 작전편>

또한 세상에는 이상적인 데이터만 있는 것이 아니라, 일정량의 잡음(noise) 데이터가 반드시 혼재한다. 따라서 데이터의 특징과 경향이 충분히 드러나고 의사결정이 가능한 단계에 이르렀다면, 분석 과정을 종료할지 추가 데이터를 수집할지 결정해야 한다. 위와 같이, 데이터 수집 및 정제, 분석, 의사결정은 ‘종합 예술 기술’에 가까운 까다로운 작업이라 할 수 있으며, 이를 위해서는 다양한 분야의 배경지식과 도메인 지식, 경험이 반드시 필요하다. 또한, 데이터 분석 과정에서 어려움을 겪는다면, 망설이지 말고 전문가의 도움을 받는 것을 강력히 권장한다. 합리적인 비용으로 시간을 절약하고 고충을 손쉽게 해결할 수 있다. 또한, 가능한 한 혼자보다는 다양한 지식과 경험이 있는 분석가로 구성된 팀으로 분석 작업을 진행하는 것이 좋다. 필자는 팀으로 분석 작업을 해서 뜻밖의 큰 성과를 이뤘던 경험이 있다.



데이터 리터러시와 독서의 중요성

최근 「데이터 리터러시(literacy, 문해력)」라는 용어가 유행하고 있다. 데이터 리터러시는 단어 그대로 데이터를 읽고 이해하는 것에 기초해 분석 결과를 다른 사람에게 잘 전달하는 능력을 의미한다. 데이터 리터러시를 향상시키기 위해서는 해결하고자 하는 문제를 명확히 정의할 수 있어야 하며, ‘생각, 질문, 추론하는 능력’이 반드시



시 필요하다. 따라서 이러한 능력을 키우기 위해서는 많은 독서량이 뒷받침되어야 한다.

故찰리 멩거, 워렌 버핏, 일론 머스크, 빌 게이츠, 사티아 나델라를 필두로 국내외 주요 CEO 및 부자들 중에 독서광이 참 많다. 이문열 작가는 대학 입학할 즈음에 이미 1만권의 책을 읽었다고 하며, 빌 게이츠는 90년대부터 매년 몇 차례 워싱턴주 후드 운하 근처 작은 오두막에서 일주일 동안 독서와 사색에만 집중하는 시간(생각 주간)을 가졌을 정도로 독서광이다(유튜브에서 관련 동영상을 쉽게 찾을 수 있다). 부자들은 대부분의 일을 다른 사람에게 맡기지만, 유일하게 ‘독서’만큼은 남에게 맡길 수 없기에 스스로 많이 읽게 되었으며, 그 결과 지식과 사리에 더 밝아져 더 큰 부자가 되었다. 또한, 독서 시간을 통해 부자들은 자신의 지식, 사고력, 창의력, 상상력을 키워 나가고 있다. 숙련된 독서가의 뇌와 초보 독서가의 뇌를 비교해 보면, 당연히 숙련된 독서가의 뇌의 각회, 전두엽 등이 더 활발히 활동하는 것을 관찰할 수 있다<참조: 「책 읽는 뇌」, 매리언 울프>. 따라서 폭넓고 깊은 지식, 사고력, 경험은 우수한 데이터 분석 및 의사 결정의 밑바탕이 된다는 점을 명심하자. AI 시대에서 살고 있는 우리들의 부와 미래를 위해, 오늘부터라도 매주 한 권의 종이책을 읽는 것은 어떨까?



본고를 요약하면 아래와 같다.

- ① 데이터 기반 비즈니스 경영 사례 알아보기
- ② 디지털 헬스케어 기술 및 빅데이터 분석 기술 그리고 AI 기술 진보로, 인간의 수명 연장이 기대된다.
- ③ ‘수확체감의 법칙’을 토대로 의사결정하라.
- ④ 우리는 늘 부족한 정보에 기초해서, 분석 및 의사결정할 수 밖에 없다.
- ⑤ 도움이 필요하다면, 전문가의 도움을 받으라.
- ⑥ 가능한 한 혼자보다는 다양한 구성원으로 형성된 팀으로 분석 작업을 진행하라.
- ⑦ 더 좋은 분석 결과와 의사결정을 위해, 꾸준한 학습과 많은 독서를 하라.

통계로 바라보는 세상이야기

신동헌 | 도서출판 지일북스 대표

통계 프리즘을 통해 살펴보는 세계 속 다양한 이슈들

질병관리청에서 발표하는 ‘국민건강영양조사 식품섭취조사’ 결과에 따르면, 2021년 하루 쌀 섭취량은 총 127.6g, 끼니당은 49.7g으로 나타났는데, 8년 전인 2013년 하루 쌀 섭취량 172.9g, 끼니당 65.5g이었던 것에 비하면 상당히 감소한 것을 알 수 있습니다. 또한, 통계청 사회조사 결과에 따르면 2022년 “우리 사회가 안전하다고 생각하느냐?”는 질문에 X세대부터 Z세대까지는 ‘개인정보 유출’ 문제가 가장 ‘안전하지 않다’라고 답했으나, 건강에 민감한 베이비붐세대나 시니어세대의 경우에는 ‘신종 질병’을 가장 불안전하게 느끼는 것으로 확인되었습니다. 아울러 환경오염을 막기 위한 행동 중 93.3%가 재활용품 분리배출을 하는 것으로 나타났습니다.

1분 1초가 아깝다? 지금은 “분초사회”

‘분초사회’라는 말을 들어보셨나요? 트렌드모니터의 ‘트렌드코리아2024’에 따르면, 웨이팅, 가사대행 앱 등 ‘시간 절약 서비스 관련 U&A 조사’ 결과, 현대 사회의 ‘시간’은 가장 큰 자원이라고 생각한다고 응답한 비율이 82.4%, 나는 시간이 곧 ‘돈’이라고 생각한다는 응답은 77.7%로 나타났습니다. 또한, ‘영상 콘텐츠 빨리 감기 시청 습관 관련 조사’에 따르면, 영상 콘텐츠 시청 경험이 있는 19~59세 성인남녀 1,000명 중 69.9%는 영상을 빨리 감기로 시청한 경험이 있다고 응답했습니다. 빨리 감기로 영상을 시청하는 이유는 ‘빨리 결론을 알고 싶어서’가 41.6%로 가장 많았고, ‘봐야 할 작품에 비해 주어진 시간이 짧아서’가 36.5%로 뒤를 이었습니다.

워라벨 추구로 확보된 여가, 혼자서 TV 시청?

한국고용정보원의 직업가치관검사 분석 결과, 직업 선택 시 가장 중요하게 고려하는 직업가치는 일과 삶의 균형(워라벨)이 4.23점으로 가장 높게 나타났습니다. 다음으로 직업안정이 4.09점, 경제적 보상이 4.07점으로 높았습니다. 또한, 문화체육관광부의 ‘국민여가활동조사’에 따르면, TV 시청(60.8%), 산책 및 걷기(43.5%), 모바일 콘텐츠·OTT 시청(43.3%)이 가장 많이 참여한 여가 활동 상위 3개로 나타난 반면 최하위는 친구 만남(22.0%)이었는데요. 대체로 실내 공간에서 개인적으로 할 수 있는 여가 활동에 참여한다고 볼 수 있습니다. 특히, 여가 활동의 동반자는 혼자서 여가를 즐기는 사람이 50.5%로 가장 많았고, 가족(34.0%), 친구·연인(13.2%)이 뒤를 이었습니다.

국내 등산 인구 약 2천4백만 명

산림청에서 전국 만 19세 이상 79세 이하 성인남녀를 대상으로 실시한 ‘2022년 등산 등 숲길체험 국민의식 실태조사’에 따르면 한 달에 한 번 이상 산에 간 등산 인구는 2021년(47.9%) 대비 10.1%p나 증가한 전체 성인 남녀의 58.0%로 약 2,392만 명으로 나타났습니다. 이 중 거의 매일 등산하는 사람은 107만명, 일주일에 두 번 이상은 359만 명, 일주일에 한 번은 590만 명, 한 달에 한두 번은 1,015만 명이었으며, 두 달에 한두 번 등산하는 사람은 321만 명이었습니다. 등산 인구의 연령별 구성비는 60대 이상이 30%로 가장 많았고, 50대 23%, 40대 20%로 그 뒤를 이었습니다. 등산을 하는 이유로 ‘건강’이라고 응답한 사람이 75%로 가장 높았습니다.

13세 이상 인구의 절반 이상 독서 안 해!

통계청에서 발표한 2023년 사회조사 결과에 따르면 지난 1년 동안 책을 읽은 사람은 48.5%로 나타났는데요. 이는 지난 2021년에 비해 2.9%p 증가한 수치이긴 하지만, 독서 인구 1인당 평균 독서권수는 오히려 2년 전에 비해 0.4권 감소한 14.8권으로 나타났습니다. 지난 2013년 62.4%, 2019년 60.6%의 독서 인구와 비교하면 갈수록 줄고 있는데요, 60세 이상은 29.2%만이 독서를 했다고 응답했습니다. 책과 가까워질 수 있는 방법으로 E-북을 활용해 보는건 어떨까요. 만약 눈이 피로해 책을 읽기가 버겁거나 이동 중에도 편하게 독서를 하고 싶다면 오디오북을 활용하는 방법도 있고, 책 읽는 것에 부담을 느끼는 분들이라면 이색 도서관을 방문해 보는 것도 좋겠습니다.

내가 다닐 회사, 5년 후에도 살아남을 수 있을까?

통계청의 ‘2022년 기업생멸행정통계’를 살펴보면, 산업별 신생기업 생존율은 1년 생존율의 경우, 전기·가스·증기 산업 90.3%, 보건·사회복지 산업 85.2%, 수도·하수·폐기 산업 78.8%로 나타났습니다. 5년 생존율 역시 전기·가스·증기 산업이 75.0%로 가장 높고, 보건·사회복지 산업 56.6%, 수도·하수·폐기 산업 50.4%를 기록하였습니다. 산업별 신생률과 소멸률을 살펴보면, 2021년에는 전기·가스·증기 산업이 24.2%로 가장 높은 신생률을 보였고, 2022년에는 정보통신업이 19.0%로 가장 높은 신생률을 보였습니다. 반대로 예술·스포츠·여가 산업은 2020년과 2021년 각각 16.9%와 13.9%로 가장 높은 소멸률을 보였습니다.

우리나라 임금근로자 평균소득 얼마?

통계청에서 지난 2월에 발표한 2022년 임금근로자일자리 소득 결과, 평균소득은 353만 원으로, 이는 전년대비 20만 원(6.0%) 증가한 것으로 나타났습니다. 중위소득 역시 2021년 250만 원에서 2022년 267만 원으로 6.9% 늘었는데요. 평균소득과 중위소득 모두 2017년 관련 통계 집계 이래 가장 높은 증가 폭을 기록했습니다. 연령대별로는 40대 근로자의 평균소득이 438만 원으로 가장 높았는데요. 50대(415만 원), 30대(379만 원), 20대(255만 원), 60세 이상(243만 원) 등이 그 뒤를 이었습니다. 19세 이하의 평균소득은 100만 원이 채 되지 않는 92만 원으로 평균소득증가율 역시 가장 낮은 2.6%에 그쳤고, 평균소득증가율이 가장 높은 연령대는 60세 이상(8.1%)였습니다.

여러분의 직장생활 만족도는 어떠한가요?

지난 2월 한국리서치에서 전국 만 18세 이상 남녀 1,000명을 대상으로 어릴 적 장래희망과 실제 직업을 조사한 결과, 일치하는 사람은 11%에 불과했습니다. 2023년 근로자들의 근로 여건 만족도는 35.1%로 2021년 대비 0.1%p 상승한 것으로 나타났습니다. 전반적인 근로 여건 만족도는 지난 2017년부터 꾸준히 상승하고 있는데요. 가장 만족도가 높은 부분은 ‘하는 일에 대한 만족도’(39.0%)로 나타났으며, 근로시간(38.3%), 근무환경(37.2%), 일과 가정의 양립(35.0%) 순으로 나타났습니다. 지난해 연간 취업자는 2,841만 6천명으로 전년대비 32만 7천명(1.2%) 증가하였고, 고용률은 62.6%로 전년대비 0.5%p 상승했는데요, 고용률 조사 이래 역대 최대치를 기록했다고 합니다.

통계로 알아보는 슬기로운 통근/통학 생활!

통계청의 '인구총조사'에 따르면, 2020년 기준, 서울특별시 평균 통근·통학 시간은 37.2분으로, 전국 평균 시간인 30분보다 약 7.2분 높았고, 그 다음으로 경기(35.3분), 인천(35.0분), 부산(30.3분) 등 순으로 나타났습니다. 또한, 전국 12세 이상 인구 4,534만 명 중 61.8%인 2,801만 명이 통근 또는 통학을 하고 있다고 하는데요. 이 중 2,329만 명은 통근자(51.4%)이며, 473만 명은 통학자(10.4%)라고 합니다. 주요 교통수단으로 직장인은 승용차·승합차(50.8%)를, 다음으로 걸어서(13.8%), 시내·좌석·마을버스(11.4%) 등이 그 뒤를 이었습니다. 평균 통근, 통학 소요 시간은 2020년 30분으로 집계되었는데요. 2015년(30.9분) 대비 소폭 감소한 것으로 나타났습니다.

현재와 미래의 1인가구는?

통계청의 '2022 인구주택총조사'에 따르면, 1인 가구는 계속해서 증가하는 추세입니다. 2022년 1인가구는 750만 2천 가구였는데요. 전체 가구 중에서 1인가구가 차지하는 비중은 2017년 28.6%에서 2022년 34.5%로 5년간 5.9%p 확대되었습니다. 연령대별 1인가구 비중을 살펴보면, 20~34세가 28.8%로 가장 큰 비중을 차지했고, 65세 이상(26.3%), 50~64세(24.2%), 35~49세(20.0%), 20세 미만(0.7%)의 순이었습니다. 1인가구로 생활하는 주된 사유로는 본인의 직장(34.3%), 본인의 독립(26.2%), 가족 사유(17.0%), 가족 사별(15.5%), 본인의 학업(4.9%) 등으로 나타났으며, 직업은 전문가가 21.5%로 가장 많았고, 사무직(19.1%), 서비스직(13.4%)이 뒤를 이었습니다.

저소득 고물가에 지갑 닫는 국민들!

지난 2월 29일 통계청에서 발표한 2023년 4/4분기 가계동향조사 결과에 따르면, 소득 1분위 가구의 월평균 소득은 117만 8천원으로 전년동분기 대비 4.5% 증가, 소득 5분위 가구의 월평균 소득은 1,080만 4천원으로 3.6% 증가한 것으로 나타났습니다. 반면, 소득 1분위 가구의 월평균 소비지출은 128만 3천원으로 전년동분기대비 1.6% 감소, 소득 5분위 가구의 월평균 소비지출은 491만 2천원으로 7.9% 증가하였는데요. 소득 5분위에서 소비지출이 가장 많이 늘어난 것에 비해 소득 1분위에서는 소비지출이 유일하게 감소한 것으로 나타났습니다. 참고로, 대한민국 가구를 20%씩 5개로 나눴을 때, 소득 최상위 20%를 5분위, 최하위 20%를 1분위라고 할 수 있습니다.

인구절벽 대한민국, 아기 울음소리 듣기 힘들어요

지난 2월 통계청에서 발표한 <2023년 출생·사망통계>에 따르면, 2023년 출생아 수는 전년 대비 1만 9천 2백명(-7.7%) 감소한 23만명으로 나타났는데요. 이는 1970년 이후 가장 낮은 수치였습니다. 또한 여성 1명이 평생 낳을 것으로 예상되는 평균 출생아 수인 '합계출산율' 역시 작년보다 0.06명 감소한 0.72명으로 나타났는데요. 특히, 4분기 합계출산율은 분기 기준 사상 처음으로 0.6명대를 기록하였습니다. 연령대로 보면 45세 이상 연령층의 출산율은 유사하지만 45세 미만에서 출산율이 감소하였는데요. 전년에 비해 30대 초반의 출산율이 6.8명, 20대 후반이 2.6명 감소했으며, 1월에서 12월까지의 월별 출생아 수 역시 모두 전년동월대비 감소한 것으로 나타났습니다.

농어업 인구 감소 심상치 않다!

최근 통계청 2023년 농림어업조사 결과에 따르면 지난해 농가 수는 전년대비 2만 4천 가구 감소한 99만 9천 가구였습니다. 농가인구 수 역시 전년대비 7만 7천 명이나 감소한 208만 9천명이었습니다. 농가 인구 가운데 65세 이상 고령인구가 차지하는 비율도 52.6%로 지난해 우리나라 전체 고령인구 비율(18.2%)의 3배에 달하는 수준으로 농촌이 늙어가는 속도가 유독 빠른 것으로 나타났습니다. 어가의 상황은 더 심각했는데요. 통계청 농림어업조사 결과, 2023년 어가는 4만 2천 가구로 전년대비 8백 가구 감소하였는데요, 2013년 14만 7천명이었던 어가인구 역시 꾸준히 줄어 2023년 8만 7천명을 기록했습니다. 이는 2013년 대비 40.9%나 감소한 수치였습니다.

그냥 여기서 살자! 인구 이동자 수 49년만에 '최저'

지난 1월 통계청은 인구이동량 등을 파악할 수 있는 2023년 국내인구이동통계 결과, 지난 2023년 국내인구 이동자수는 총 612만 9천 명으로 1974년 529만 8천명 이후 49년만에 최저치로, 전년 대비 0.4% 감소한 수치로 나타났습니다. 인구 100명당 이동자 수를 나타내는 인구이동률 역시 12.0%로 1972년 이후 최저를 기록하였는데요. 연령별 이동률은 20대(22.8%)와 30대(20.1%)가 높고, 60대 이상 연령층에서는 낮게 나타났습니다. 인구이동 감소 이유는 고령화와 교통망의 발달 등을 들 수 있을 것 같습니다. 국내인구이동이란 주민등록법에 따라 거주지 변경 시 제출하는 전입신고서를 기초로 읍면동 경계를 넘어선 이동자를 집계한 결과라고 합니다.

우리 공감으로 나눠요. 모두가 함께 하는 세상을 위해!

지난 4월 20일은 장애인의 날인데요. 보건복지부에서 매년 발표하는 장애인 현황에 따르면 2023년 국내 등록장애인 수는 264만명으로 집계되었습니다. 이는 전체 인구의 5.1% 수준이었는데요. 이중 여성장애인은 111만명(42.1%), 남성장애인은 153만명(57.9%)이었으며, 15개 장애유형 중 가장 많은 유형은 지체장애(43.7%)였고, 다음으로 청각장애(16.4%), 시각장애(9.4%), 뇌병변장애(9.1%), 지적장애(8.7%) 등의 순이었습니다. 2019년 37.3%였던 경제활동참가율은 2022년에 38.1%를 기록하였고, 지난해에도 37.4%로 꾸준한 증가 추세를 보이고 있습니다. '장애인 고용 의무화 제도'가 시행된 지 30년째인 2020년 장애인 의무고용 사업체의 장애인 고용률이 처음으로 3%를 넘었으며, 2022년 고용률은 3.12%로 나타났습니다.

봉사활동 하고 싶은데, 시간이 없다면?

통계청이 2023년 11월에 발표한 '2023년 사회조사 결과'에 따르면 지난 1년 동안 자원봉사활동 경험이 있는 사람은 10.6%에 불과했는데요. 코로나19가 한창이던 2021년보다 2.3%p 증가하긴 했지만 10년 전인 2013년의 19.9%와 비교하면 자원봉사참여 경험자가 상당히 감소했음을 확인할 수 있었습니다. 자원봉사에 참여하는 이유로 '학교나 직장 등에서의 단체활동에 동참하기 위해서'라고 답한 사람이 25.9%로 가장 많았습니다. 반면 자원봉사활동에 참여하지 않은 가장 큰 이유는 '시간적 여유가 없어서(48.7%)'로 나타났고, '자원봉사에 관심이 없거나(33.6%)' '자원봉사활동 경비가 부담된다(5.3%)'라는 답변이 뒤를 이었습니다.



‘데이터 분석 제목 달기’ 심층 알아보기

강양석 | Deep Skill 대표

데이터 분석 학습을 하는 사람에 비해 정작 실제 분석을 통해 의사결정을 하는 경험을 가진 사람은 많지 않은 것 같다. 이는 실컷 장을 본후 정작 요리를 하지 않는 것에 빗낼수 있다. 조직내에서 이런 현상이 지속되면 불필요한 데이터만 많이 쌓이고 데이터 분석에 대한 실효성에 시작도 전에 의심하는 문화가 생길수 있다. 그렇기 때문에 되든 안되든 실제 분석을 시도해보는 것은 중요하다. 한데, 데이터 분석을 정작 실시하려면 일종의 기획서가 필요하다. 분석 기획서가 그것이다. 얼핏 분석 기획서를 써보라 하면 많은 사람들이 3~4장되는 문서를 한바탕 써내곤 한다. 한데 이런 방식은 뭔가 비효율적이다. 문제는 이렇게 많은 양의 분석기획서를 작성하면서 ‘생각이 정리’되어야 하는데 잘 들여다 보면 기획서 작성을 위한 작성이 되곤 한다는게 문제이다. 용어의 일관성, 문제의 식의 뚜렷함, 해결 방안의 구체성이 결여된 채 뭔가 글은 많아지지만 더욱 모호해진 상태가 돼버리

기 때문이다. 이를 방지하기 위해 내가 주로 사용하는 방식은 ‘제목 달기’이다. 즉, 데이터 분석 기획의 제목만 달아보는 것이다. 저번 호에 이어 이 방법에 대해 좀더 구체적인 사례를 바탕으로 그 요령을 공유해보고자 한다. 대기업 임원부터 핵심인재까지 본 방법을 통해 압축적인 분석 기획의 효과를 맛보았다고 술회했기 때문이다.

지난 호에서 설명했듯이, 제목 달기를 위해 우리에게 필요한 개념은 딱 4가지이다. A : 데이터, B : 분석, C : 의사결정내용, D : 비즈니스 임팩트가 그것이다.



그러니 이를 한번에 읽으면 다음과 같이 데이터 분석기획 제목이 되는 것이다. ‘어떤 데이터를 어떤 분석하여, 어떤 의사결정을 함으로써, 어떤 가치를 만들어 낼것이다.’라고 말이다. 언뜻 보면 쉬워보이지만 생각보다 이 넷을 어울리게 작성하는 사람은 많지 않다. 왜 그렇다는 것인지 한번 사례들을 보면서 체험해 보자.

사례 1 | 엄격한 글쓰기가 안되는 경우(금융사 사례)

- A 지역별 수신고추이와 4대 시중은행 평균금리, 지역금융기관의 금리의 데이터를 활용하여
- B 기간별, 지역별로 분석하여
- C 지역별 수신고의 지역은행 및 시중은행 금리 간의 상관관계를 분석하여
- D 수신고의 변동 가능성을 대비해, 지역별 AAA금융 수신고의 안정성을 높이고자 함

이 제목을 읽어보면 약간 혼란스러운 곳이 있다. 정확히 분석의 대상(objet)과 현상이 어떻게 구분되는지 분간이 안되기 때문이다. 이런 현상은 이런 실습을 해보면 왕왕 볼수 있다. 글을 길게 쓸때는 잘 드러나지 않지만, 글을 짧게 쓸 때 나타나는 모미중에 하나이다. 자신이 어떤 대상의 어떤 현상을 어떤 쟁점으로 분석하고자 하는지가 ‘일단 글 자체의 모호함’ 때문에 잘 안보이는 것이다. 명확한 글쓰기가 중요한 이유는 생각이 정리되지 않은채 데이터에 손을 대는 것은 위험하기 때문이다. 왜냐하면 자칫 데이터가 생각을 끌고 가는 현상에 빠져 분석을 해도해도 갈피를 못잡는 경우가 발생할수도 있기 때문이다. 그래서, 자신이 생각이 정리되었는지 아닌지 또는 적어도 그 글을 다른 사람이 이해하는지 아닌지를 알기 위해서라도 글을 명확히 써보는 연습이 모든 이야기의 시작중에 시작이 된다.

사례 2 | 목적이 사실은 없는 경우(금융업 사례)



- Ⓐ OO은행 고객의 이체건수 데이터로
- Ⓑ 고객의 AAA예금 보유규모와 AAA은행 이용빈도를 분석하여
- Ⓒ 보유규모와 이용빈도에 따라 고객군을 구분(Heavy유저, Medium 유저, Light 유저, Inactive유저) 및 도출하고
- Ⓓ 고객군별 마케팅방안 발굴을 통해 AAA은행 활성화를 도모한다.

이 제목달기는 어떤 점을 개선할 필요가 있을까? 언뜻 보면 글도 잘 이해가 가고 분석의 대상과 현상을 잘 구분한듯 보인다. 그럼에도 한가지 작지만 중요한 뒀지 아닌 뒀지를 걸자면 바로 D 부분의 ‘마케팅 방안 발굴’이라는 표현에 있다. A, B, C는 뭔가 섬세하게 잘 구성하였지만, 정작 D가 마케팅 방안 발굴이라 적히면 분석의 목적이 지나치게 광범위해 지는 것이다. 예를 들면, MAU 또는 DAU 증가와 같은 구체적인 목적을 필요로 한다는 것이다. 왜냐하면 고객을 어떤 특성에 따라 분류하는 것 자체(C부분)는 그 자체로 목적이 될 수 없기 때문이다. 실제로 분석을 해보면 뚜렷한 목적 하에서 기준이 생기고 그 기준이 분석 실무에 강하게 영향을 준다는 것을 알 수 있다. 그러므로 조직의 상위 직급자일수록 이 D부분을 명확하고 구체적으로 정해주는 것에 공을 들여야 한다. 마케팅 방안 발굴 그 자체는 너무 광범위한 분석 목적이다.

사례 3 | 상관관계와 인과관계를 혼동하는 경우



- Ⓐ OO예금 우수고객 스코어링 데이터를 통해
- Ⓑ 이탈한 우수고객의 주요 원인 항목을 분석하여
- Ⓒ 핵심원인 파악을 통한
- Ⓓ 우수고객 이탈 방어 대책 마련.

이 제목에서 가장 중요한 단어는 단연 ‘원인’이라는 단어이다. 원인을 밝힌다는 것은 흔히 말해 인과관계 분석을 한다는 것이다. 그런데, 인과관계는 여간해서는 과거 데이터 추이에서는 밝히기 어렵다. 상관관계는 경향을 다루고, 인과관계는 영향을 다루는데 그 영향력의 크기는 아이러니컬하게도 그 원인이 없었더라면 결과는 어느정도였을까?라는 모호한 질문에 답이 되어야 하기 때문이다. 흔히 말해, 결혼이 행복에 얼마나 기여했을까의 영향력을 알기 위해서는 결혼을 안했다더라면 행복감이 얼마일지를 알아야 결혼의 행복 기여를 정확히 알 수 있다는 말이다. 그래서 이런 복잡한 관계의 비교 가능성을 확보하기 위해 과거 데이터 분석으로 쉽게 얻어지기가 어렵다.

이 기획은 이 둘을 구분하지 못한 표현이라고 볼 수 있다. 다시 말해, 과거 고객의 이탈 원인 경향성이 반드시 이탈 방지책이 될 수는 없다는 것이다. 이럴 경우 그러면 어떻게 해야할까? 인과추론을

위한 분석 기획을 해야 맞을 것이다. 인과와 상관을 구분하는 것은 데이터 과학에서 굉장히 중요한 주제이다. 그러므로 이런 쟁점을 간단한 제목달기에서만이라도 느낄 수 있다는 것은 나름 의미가 크다 하겠다.

사례 4 | 분석결과와 결론이 다를 때



- Ⓐ 기존 판매된 고객 데이터로
- Ⓑ 고객의 할인과 사용분석을 통해
- Ⓒ 추가상품 유치가 가능한
- Ⓓ 판매의 효율성 및 많은 상품 유치로 매출증대

마지막으로 분석 기획 시 자주 드러나는 실수 유형이 분석의 결과와 결론이 연관이 없는 경우이다. 실은 분석 결과와 결론이 어떤 차이인지도 인지하지 못하는 경우도 많다. 분석 결과는 데이터 분석 내에서의 결과이고 이를 어떤 의사결정으로 끌어내느냐는 별개의 문제이다. 지금 이 생각 프레임워크를 중심으로 설명하면 B부분과 C부분이 엇박자가 나는 경우라고 하겠다. 이게 무슨 말이고 하니, 기존 제품의 데이터 분석으로는 새로운 추가 상품 니즈가 완성되지 않는다는 것을 의미한다. 그도 그럴 것이 어느 편의점에서 할인율에 따른 우유의 판매량 민감도 분석을 통해 초콜릿의 추가 판매 가능성을 알수는 없다는 뜻이다. 언뜻 들어도 쉽지 않아 보이는 이 분석 기획 사례가 은근히 많이 나오는 실수 중 하나가 되겠다. 물론 우유를 사는 고객과 초콜릿을 사는 고객이 실질적으로 동반 구매한다는 사전 분석이 있으면야 가능하겠지만 이 분석 기획 내용만으로는 해당 내용을 알 수가 없기 때문에 보강이 필요하다는 것이다.

자, 정리해 보자. 데이터 분석을 기획하라하면 많은 사람들이 어떤 분석기법을 쓸까를 고민하는 흔적이 많다. 하지만, 분석 기법 보다 더 중요한 쟁점들이 기획단계에서 제법 많다. 앞서 살펴본 현상과 대상의 구분, 상관관계와 인과관계의 구별, 결과와 결론의 정렬, 목적의 모호함 등은 그 중 대표적인 사례만을 소개해 보았다. 이렇게 짧게 분석 기획을 하다보면 오히려 이런 더 근원적인 쟁점에 다가갈 수 있다는 장점이 있다. 그러니, 너무 긴 기획서를 쓰려고 하기 보다 짧더라도 수미일관하고 뚜렷한 방향성이 있는 기획을 해보도록 노력해 보자.





스몰 데이터로 새로운 가치 창출하기

구자룡 | 벨류바인 대표

빅데이터 시대에도 여전히 중요한 스몰 데이터

바야흐로 지금은 인공지능의 시대다. 인공지능을 가능하게 한 재료가 바로 빅데이터다. 빅데이터를 활용한 인공지능의 도움으로 이동이 편리해졌다. 내비게이션에는 인공지능이 들어가 있다. 그런데 간혹 의문이 든다. 왜 내가 원하는 길로 안내하지 않을까? 편한 길로 가고 싶은데 빠른 길로 안내한다. 1분을 단축하기 위해 좁고 위험한 길을 가고 싶지는 않다. 비록 1분이 늦어지더라도 5km를 돌아가더라도 간선도로를 이용하여 편하게 운전하고 싶다. 인공지능은 왜 내 마음을 모을까? 어쩌면 당연한 것일지도 모르겠다. 인공지능은 과거의 데이터를 바탕으로 예측을 통해 정보를 제공한다. 나의 마음을 읽으려고 한 적은 없다. 실시간으로 이루어지는 의사결정에서 나의 마음까지 읽고 정보를 제공해 주면 좋겠지만 아직은 그 경지에까지 오르지 못한 것 같다.



이러한 실시간 의사결정이 필요한 상황에서 사람들은 순간적으로 자신에게 질문하고 답변을 하면서 의사결정을 한다. 또한 순간적으로 주변의 여러 상황을 파악하여 의사결정을 한다. 빅데이터가 아니라 스몰 데이터로 판단하고 의사결정을 한다. 인간은 38억 년에 걸친 진화 과정을 거쳤고 15만 년 전에 현생인류(호모 사피엔스)로 분화했고 그 이후 현재까지 질적으로 거의 같은 두뇌 구조로 되어 있다. 원시시대에 정글에서 살아남아야 하는 숙명과 현대사회에 도시 정글에서 살아남아야 하는 숙명은 같다.

원시시대에는 뱀이나 호랑이를 빠르게 인식해야 살아남았다. 발밑에서 바스락거리는 소리를 들으면 빨리 피해야 한다. 설령 뱀이 아니더라도 손해 볼 일은 없다. 피하지 못해 뱀에 물리는 것보다는 나은 선택이다. 현대에도 다양한 위험 요소를 빠르게 간파해야 살아남을 수 있다. 자동차의 굉음이나 폭발음을 들으면 빠르게 피해야 한다. 급발진이든 가스폭발이든 피하면 살 수 있고 설령 나에게 위해가 없더라도 피해서 손해 볼 일은 없다. 우연을 필연으로 간주하는 능력이 필요하다. 빅데이터와 인공지능으로부터 도움을 받을 수 있는 부분은 당연히 도움을 받아야 하지만 그렇지 못한 상황에서는 여전히 눈치가 중요한 생존 수단이 된다. 인공지능이 할 수 없지만 우리는 할 수 있는 능력

에 다시 관심을 기울일 필요가 있다. 바로 스몰 데이터(small data)다.

스몰 데이터는 몸짓이나 말투, 인상 등과 같이 사소한 행동에서 나오는 소소한 데이터다. 소비자의 취향이나 라이프스타일과 같은 개인화된 데이터라고 할 수 있다. 그리고 데이터의 양을 기준으로 살펴보면 기본적으로 적은 양의 데이터다. 단순한 양이 아니라 데이터의 속성(feature, 특징 또는 변수)이 매우 작은 경우다. 예를 들어 어떤 사람의 특성을 나이, 성별, 키, 몸무게, 인상, 말투, 몸짓 등으로 데이터화할 수 있다. 물론 이 경우에도 한 사람이 아니라 5천만 명의 데이터를 확보했다면 속성은 몇 개 되지 않지만 데이터의 양은 5천만 개가 되어 대규모의 데이터가 된다. 일반적으로 빅 데이터는 대규모의 양과 실시간 데이터 수집과 저장을 할 수 있는 속도, 그리고 정형 및 비정형 데이터 등 데이터 유형의 다양성을 함께 가지고 있는 경우를 뜻한다. 반면에 스몰 데이터는 규모 측면에서 양이 크지 않고, 실시간 데이터 수집 및 저장이 어렵고, 데이터의 유형이 다양하지 않은 데이터다. 우리가 일반적으로 접하는 개인화된 데이터가 대부분 여기에 해당한다.

이런 개인화된 데이터를 비즈니스에서 활용하기 위해 그 동안 산업계에서는 다양한 방법으로 데이터를 수집하고 분석하는 방법을 개발해 왔다. 가장 오래되었으며 특별한 기술이 필요하지 않은 방법으로 관찰법이 있다. 물론 지금은 특별한 기술이 들어간 기계의 힘을 빌린 관찰법이 많이 보급되고 있다. 또한 오랜 기간 기술의 발전에 따라 더욱 체계화된 방법으로 인터뷰가 있다. 인터뷰는 면접하는 방법으로 심층면접법, 표적집단면접법, 그리고 은유유도기법(ZMET), 컬러코드 등의 방법들이 있다. 대체로 텍스트 기반의 데이터로 질문을 통해 응답자의 내면의 세계에 들어가 보고자 한 접근방법이다. 아울러 질문지를 체계화한 서베이도 전형적인 스몰 데이터를 수집하는 방법이다. 스몰 데이터 연재를 통해 관찰과 인터뷰, 서베이 등의 데이터를 수집하고 분석하여 통찰을 통해 새로운 가치를 창출하는 방법을 알아보려고 한다. 이번 호에서는 새로운 가치 창출을 위한 관찰법과 인터뷰 방법에 대해 살펴본다.

숨어 있는 소비자의 마음을 읽어내는 방법 : 관찰법

관찰(observation)은 체계적인 방식으로 관심 주제에 대한 정보를 얻기 위해 사람, 사물, 사건의 행동 형태를 기록하는 것이다. 관찰자는 관찰되고 있는 사람들과 질문도 하지 않고 의사소통도 하



지 않는다. 예를 들어 소비자들이 어떤 브랜드를 선택하는지 또는 어떤 TV 프로그램을 보는지에 대해 응답자에게 물어보는 것이 아니라 관찰하는 것이다. 길거리에서 유행하는 패션을 관찰한 데이터나 할인점이나 슈퍼마켓에서 스캐너로 바코드(bar code)를 읽은 데이터를 분석하여 유행하는 색상이나 유형 또는 잘 팔리는 제품이나 브랜드를 확인할 수 있다. 관찰의 결과를 바탕으로 신제품 기획이나 새로운 프로모션 계획을 세울 수 있다.

예를 들어, 여러분은 사무실이나 워크숍에서 많이 사용하던 ‘포스트잇’을 알고 있을 것이다. 이 포스트잇을 사용하는 사람을 며칠 동안 관찰한다고 생각해 보자. 포스트잇을 사용하면서 불편한 점이나 고통스러운 점이 있는지 관찰해야 한다. 먼저 문제를 정의했으니, 구체적으로 데이터를 수집하는 것이다. 수집된 데이터를 바탕으로 포스트잇의 문제점을 분석해 보자. ‘메모를 보관하기 어렵다, 중요한 메모라면 다시 옮기는 작업을 해야 하는 불편함이 있다, 바람에 날릴 정도로 잘 떨어진다, 분실 가능성이 있고 손상되기 쉽다, 메모할 수 있는 범위가 한정적이다, 작성한 글을 지우거나 재사용하기 어렵다.’ 등과 같은 불편이나 고통을 발견했다. 이제 고객의 불편과 고통을 해결하는 새로운 아이디어를 도출해야 한다. 실제 이런 문제에 착안하여 “접착제나 핀 없이 어느 곳이나 붙일 수 있는” 마그네틱 노트(Magnetic Notes)가 개발되었다([그림1] 참조).



[그림1] Tesla Amazing의 Magnetic Notes

그렇다면 어떻게 해야 통찰로 연결되는 관찰을 할 수 있을까? 제대로 관찰하려면 어떤 대상을 유심히 바라봐야 한다. 유심히 바라보면, 이전에는 쉽게 보이지 않던 것이 보이기 시작한다. 대중 훑어보고 지나치는 수동적인 보기가 아니라 관심을 두고 주의 깊게 탐색해야 한다. 아서 코넨 도일의 추리 소설 《보헤미아의 스캔들》에서 탐정 셉록 홈스는 “보는 것과 관찰하는 것은 크게 다르다!”라고 일갈했다. 적어도 비즈니스에서 데이터로 새로운 가치를 창출하고자 한다면 셉록 홈스 같은 관찰 마인드가 필요하다. 즉, 그냥 보는 것이 아니라 관심을 두고 유심히 훑어지게 관찰해야 한다. 오래전 레

오나르도 다 빈치는 제자들에게 “벽돌 사이의 금까지도 스케치하라. 좋은 그림을 그리려면 날마다 밖으로 나가서 흥미로운 것을 발견하고 관찰한 뒤 자세하게 기록하라.”라고 했다. 우리 역시 어떤 문제를 정의하고 흥미를 갖고 이상하거나 변화가 일어나는 것을 감지(sensing)하고 그 속에 무엇이 있는지 훤히 꿰뚫어 보는 통찰(insight)을 해야 한다. 새로운 가치를 만들어 내는 출발점이다.



숨어 있는 소비자의 마음을 읽어내는 방법 : 인터뷰

인터뷰(interview)는 관심 주제에 대해 연구자(interviewer)가 응답자(interviewee)의 생각을 직접적으로 질문을 하고 답변을 듣는 방식이다. 대표적인 방법으로 심층 면접(in-depth interviews)이 있다. 실무적으로 가장 많이 활용되고 있으며 가장 단순한 방법이다. 심층 면접은 구조화되어 있지 않고, 직접적이며, 개인적인 면접으로 응답자(소비자)의 심리 전반에 내재해 있는 지식, 신념, 태도, 동기, 감정 등의 정보를 획득하고자 할 때 유용하다.

예를 들어, 우리 회사가 고객의 니즈를 파악하여 아이들의 시간 단축과 즉각적인 만족감을 제공할 수 있도록 쉽게 조립할 수 있는 블록을 만들어 제공했으나 지속해서 매출이 감소하고 있다고 생각해 보자. 매출 감소가 가장 큰 문제지만 그 문제의 근원이 무엇인지 알 수 없다. 고객 니즈를 파악했고 고객이 원하는 제품을 만들어 주고 있었기 때문이다.

이런 문제에 봉착한 레고(LEGO)는 독일의 한 소년을 찾아가 인터뷰를 진행했다. 그는 레고광일 뿐만 아니라, 열정적인 스케이트 보더였다. “당신이 가장 자랑스럽게 생각하는 물건은 무엇입니까?”라는 질문에 소년은 님고 님고 아디다스 운동화 한 켤레를 가리키며 “이 운동화는 나에게 우승컵이자 금메달입니다.”라고 답변했다. 여기서 우승컵이자 금메달이라고 한 의미는 무슨 뜻일까? 만약 레고에서 판매했던 즉각적인 만족감을 줄 수 있는 쉬운 레고 블록이었다면 이게 우승컵이자 금메달이 될 수 있었을까. 고객이 원하는 것은 시간 단축과 즉각적인 만족감이 아니라 자신이 이 지역에서 최고의 스케이트 보더라는 것을 입증하는 증표였다. 이후 레고는 “블록으로 돌아가자(back to the brick)”라는 전략으로 선회하여 성공을 거두고 당시 세계 최고의 완구 제조사인 마텔(Mattel)

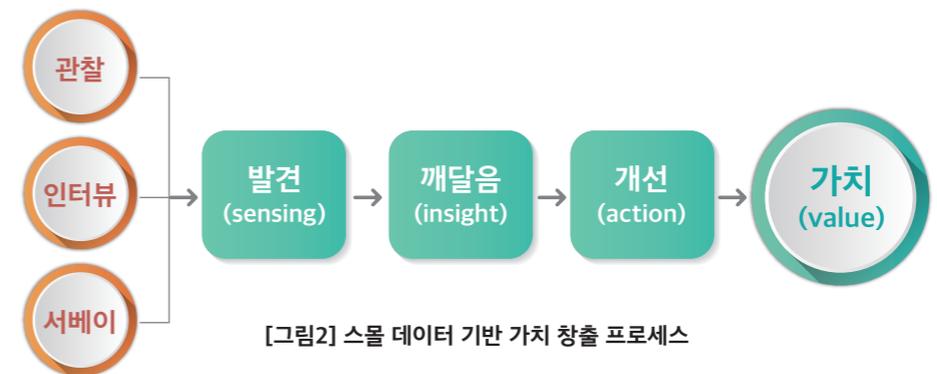
을 넘어서는 성과를 냈다.

인터뷰를 통해 통찰을 끌어내기 위해서는 캐묻기와 같은 면접 기술(skill)도 있어야 한다. 캐묻기(probing)는 의미 있는 응답을 획득하거나 숨어 있는 문제를 찾아내는 데 있어 매우 중요하다. 캐묻기는 왜 그렇게 생각하십니까?, 어떤 내용인지 좀 더 말씀해 주시겠습니까?, 무엇이든지 좀 더 말씀해 주시겠습니까? 등과 같은 질문으로 계속 파고들어가는 방식이다.

예를 들어, 토마토 수프에 대한 신제품을 개발하여 시식하게 한 다음 “이 토마토 수프가 어떻게 개선되었는지를 말씀해 주시겠습니까?”라고 질문을 해보자. 응답자가 “글쎄요, 우선 색상이 좀 변한 것 같네요.”라고 응답했다면 캐묻기로 “색상이 좀 변했다고요?”라고 질문하면 된다. 그리고 “네, 아주 짙어 졌어요.”라는 답변을 받았다면 이어서 “아주 짙어져요?”라는 추가 질문을 하면 된다. 그러면 응답자는 “네, 수프가 아니라 토마토 소스 같아요.”라는 답변을 했다고 가정하자. 심층면접을 통해 우리가 알지 못했던 토마토 소스라는 의견을 수집한 것이다. 이제 우리가 개발한 토마토 수프와 응답자가 말한 토마토 소스가 무엇을 의미하는지 통찰하면 된다.

데이터의 본질은 새로운 가치 창출

관찰이나 인터뷰 또는 서베이를 통해 어떤 이상한 점을 발견, 즉 감지했다면 이를 통해 그 데이터의 이면을 훤히 꿰뚫어 보는 깨달음, 즉 통찰해야 한다. 그리고 통찰을 바탕으로 개선 활동을 해야 어떤 결과를 얻을 수 있다([그림2] 참조). 데이터 기반의 개선 아이디어를 실행으로 옮겨야 새로운 가치가 나왔는지 혹은 그렇지 못한지를 알 수 있다. 데이터 기반 의사결정의 중요성이 강조되고 있지만 정작 실행(action)을 하지 않으면 아무런 변화가 일어나지 않는다. 데이터의 본질은 데이터로 새로운 가치를 만드는 것이다. 궁극적으로는 비즈니스에 변화를 일으키는 것이다. 마그네틱 노트나 레고는 스몰 데이터로 센싱을 한 다음 실행으로 통찰을 증명했다. 스몰 데이터이든 빅데이터이든 중요한 것은 데이터의 규모가 아니라 센싱과 통찰을 할 수 있느냐 하는 것이다.



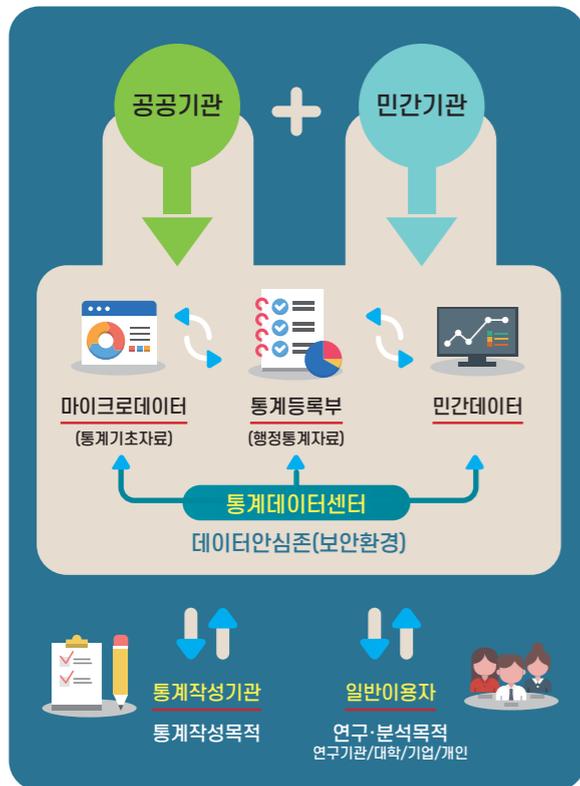
[그림2] 스몰 데이터 기반 가치 창출 프로세스

행정통계자료와 민간자료를 한곳에!

통계데이터센터 서비스

통계데이터센터가 새로운 서비스로
정보화 사회를 선도합니다.

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



- 1 분석 플랫폼 제공 서비스**
 - 분석시스템 · 통계패키지 제공
 - 통계자료(통계등록부 · 통계기초자료) 및 민간자료, 이용자 반입자료 연계 · 분석
- 2 전문가 분석지원 서비스**
 - 분석 경험이 없는 이용자를 위한 데이터 분석 지원
 - 센터 이용 상담 및 데이터 분석 자문
- 3 주문형 분석서비스**
 - 시간 및 거리상 센터 방문이 어렵거나 직접 자료분석을 하기 힘든 이용자를 위한 서비스
 - 센터 이용자료를 활용하여 연계 · 분석 후 이용자가 원하는 형태로 결과를 제공
- 4 명부 서비스**
 - 분석센터로 방문하여 자료분석 및 표본설계를 통해 데이터 반출
- 5 이용자 교육 서비스**
 - 이용자 교육 홈페이지 운영
 - 통계분석 프로그램 및 분석사례 교육
 - 매년 통계데이터 활용대회 개최

※ RDC 제공자료도 이용 가능합니다.

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



보다 심도 있고 다양한 분석을 원한다면
지금 바로 MDIS를 클릭해 보세요.

■ 서비스 소개 (2023년 5월 기준)

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

나. 제공 통계 수 : 21개 주제별 총 357종 통계 제공

(통계청 50종 및 통계작성기관 307종)

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

기준	주요 통계	
통계청	인구·가구	경제활동인구조사, 가계동향조사, 국내인구이동통계, 사망원인통계, 가계금융복지조사, 지역별고용조사, 인구주택총조사, 인구동향조사, 생활시간조사, 사회조사 외 8종
	사업체·농어가	전국사업체조사, 광업제조업조사, 농가경제조사, 기업활동조사, 농림어업총조사, 농산물생산비조사, 경제총조사, 어가경제조사, 운수업조사 외 14종
	행정통계	귀농어귀촌인통계, 영리법인기업체행정통계, 신혼부부통계, 주택소유통계, 중장년층행정통계, 퇴직연금통계, 일자리행정통계, 기업생멸행정통계, 육아휴직통계
통계작성기관	전국다문화가족실태조사, 가족실태조사, 자동차주행거리통계, 직종별사업체노동력조사, 보육실태조사, 기상관측통계, 국민여가활동조사, 외래관광객조사, 한부모가족실태조사, 청소년종합실태조사 외 297종	

■ 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 집사무실 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 제공 사이트를 활용하세요.



통계교육원



sti.kostat.go.kr

국내 유일의 국가통계교육 전문기관

통계 작성 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