데이터로 고객의 욕망까지 읽는 시대,
데이터 드리븐 디자인씽킹을 말하다

Data Driven Design Thinking

ⓒABLEARN, 2022.

이 책은 저작권법의 보호를 받는 저작물로 무단전재, 복제, 배포를 금합니다.

이를 위반 시 민사 및 형사상의 법적 책임을 질 수 있습니다.

ISBN 979-11-979450-1-4 | 02-589-2021 | edu@ablearn.kr | ablearn.kr

1. 디자인씽킹과 데이터

1.1. 디자인씽킹이란?

이 브랜드들의 공통점이 있다면 무엇일까? 다양한 고객사와 서로 다른 비즈니스 형태와 구조를 갖고 있어서 딱히 유사점이 있어 보이진 않는다. 하지만 이들 모두의 공통점은 신제품 기획 및 기존 제품 개선에 디자인씽킹이라는 방법을 쓰고 있다는 것이다. 비단 이들 뿐만이 아니다. 우리나라에서도 스타트업부터 대기업은 물론, 공공기관에서도 디자인씽킹 방법을 사용하여 사용자 중심적인 제품, 상품, 정책을 제공하고 있다. IDEO의 CEO 팀 브라운(Tim Brown)의 말처럼 소비자의 요구를 충족시키기 위하여 디자이너의 감수성과 작업방식을 이용하는 사고방식이라는 디자인씽킹이란 대체 어떤 것일까?

| 와비파커 이야기

지난 9월 29일 뉴욕 증권거래소(NYSE)에 상장하고, 기준가 대비 30% 넘게 급등하여 60억 달러(약 7조 1200억 원)를 넘어선 10년 정도밖에 안된 스타트업이 있다. 그들은 바로 대학원생 몇 명이서 창업한 안경계의 넷플릭스라고 불리는 '와비파커'이다. 사업 초기 이들에게는 뛰어난 기술력이나 화려한 서비스가 존재하지 않았다. 모든 사업가들이 알고 있다고 할 수 있는 합리적인 가격과 만족스러운 서비스를 제공했을 뿐이다. 와비파커의 창업자들은, 대학원생들이 구매하기에는 턱 없이 비싼 미국 안경 가격에 불만이었고, 충분히 착용해보지 못하고 구매하여 실패하게 만드는 경험에 화가 났다. 그래서 안경 가격을 95달러로 낮추고, 5개의 안경을 골라서 5일간 체험해볼 수 있게 서비스를 제공했다. 그렇게 이들은 고객의 불편 한 경험을 명확하게 해결하여 초고속으로 성장할 수 있었다.

와비파커의 성공 이유는 무엇일까? 13세기부터 존재했던 안경이라는 아이템을 갖고 어떻게 21세기 혁신적인 사업 모델을 만들 수 있었을까? 성공에 영향을 끼친 요인을 여러 가지 존재하겠지만, 바로 창업자들이 안경을 사용하는 사람들의 불편한 경험(Pain point)을 정확하게 정의했고 이를 집중적으로 해결하는 서비스를 제공했기 때문이다. 이것이 바로 디자인씽킹이라고 할 수 있다. 사용자의 경험을 정확하게 정의하고 해결해주는 방식이다. 디자인씽킹의 정의를 말해 보라 하면 아래와 같은 여러 가지 답변이 나온다. 다 정답이라고 볼 수도 있다. 하지만 다른 문제 해결 방법론에 비해 디자인씽킹이 지니고 있는

차별점을 바탕을 정의해보면, 디자인씽킹은 "공감을 기반으로 고객의 불편한 점을 명확히 정의하여 창의적으로 문제 해결하는 사고방식"이라고 볼 수 있다.

디자인씽킹 요약정리

- ✓ 창의적 문제 해결 방식
- ✓ 문제정의를 탁월하게 할 수 있는 방법
- ✓ 폭발적으로 아이디어를 창출할 수 있는 수단
- ✓ 빠르게 프로토타이핑을 해 볼 수 있는 프로세스
- ✓ 고객 경험을 테스트해보기 가장 적합한 도구

1.2. 디자인씽킹의 한계와 데이터의 필요성

디자인씽킹에서 가장 중요한 파트인 고객 공감 덕분에 기술과 서비스가 아닌 고객의 불편한 점을 우리는 직시할 수 있었고, 많은 조직은 이 방법을 통해 혁신적인 경험을 설계하여 성공적인 결과를 만들어냈다. 이처럼 상당한 전문가와 엄청난 도구 없이도 문제 해결이 가능한 디자인씽킹은 마치 단점은 전혀 없어 보였다. 그렇지만 세상에 완벽한 것은 없다. 업계에서도 디자인씽킹에 관해서 결과물 예측이 어렵다거나 너무 추상적이라며 회의적으로 보는 시각도 생겨나기 시작했다. 실리콘밸리의 Natasha Jen이라는 디자이너는 강연에서 "디자인씽킹은 헛소리다! (Design Thinking

is Bullshit!)”라며 디자인씽킹을 혹독하게 비판했다. 그녀의 주장을 요약하면 디자인씽킹은 명확하게 정의하기 어렵고, 너무 심플하게 문제를 바라보며, 전문가의 심도 깊은 고찰과 식견을 무시한다고 한다는 것이었다.

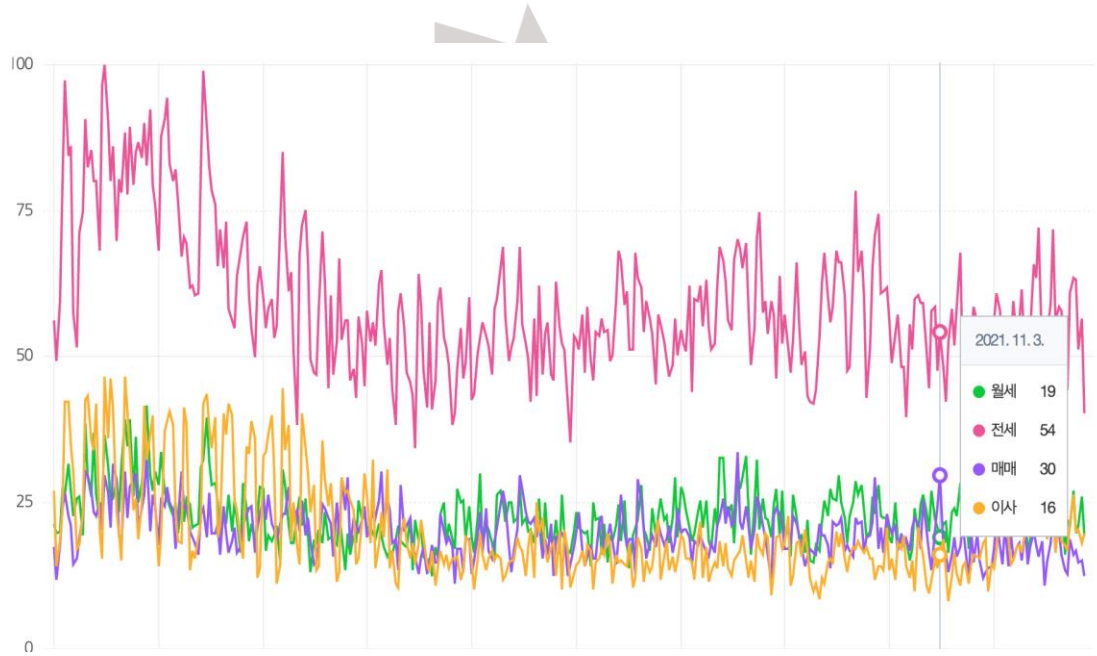
정말 디자인씽킹이 그러할까? 디자인씽킹의 고전 사례로 유명한 IDEO의 쇼핑카트 개선 프로젝트를 살펴보자. IDEO는 마트 안에 있는 쇼핑카트에 여러 불편한 점이 존재한다는 가정하에 팀을 꾸렸다. 그런데 그 팀 구성원은 MBA 학생, 언어학자, 마케터, 심리학자, 생물학 전공자 등 쇼핑카트와는 전혀 거리가 먼 사람들로 구성했다. 보통 이런 프로젝트를 시작한다면 기계공학 또는 산업디자이너 등으로 팀원을 구성할 것이다. 하지만 IDEO는 오히려 편견 없이 고객을 공감하기 위하여 전문가로 구성된 팀을 만들지 않는다. 디자인씽킹은 문제와 관련된 근본적인 추측들에 대해 편견 없이 질문해보라는 것이라고 하는 스탠포드 디스쿨(Stanford D-School)의 철학을 따른다. 전문가의 식견보다는 편견 없이 고객을 공감할 수 있는 환경을 더 선호한다.

정말로 전문가의 고찰과 식견 없이도 우리는 문제를 해결할 수 있을까? 하나의 사례를 더 살펴보자. 20대 초반 여성들의 주거 문제를 해결하겠다고 가정하자. 네이버 검색어 통계치를 알려주는 네이버 데이터랩[1]을 이용해보겠다. 총 네 가지 키워드(월세, 전세, 매매, 이사)를 대상으로, 20대 초반 여성들이 2021년 동안 검색한 양을 살펴보려 한다. 과연 어떤 키워드를 가장 많이 검색했을까?

[1] datalab.naver.com

실제로 40대 이상이 과반수인 집단에게 이 질문을 하면, 월세 또는 이사 키워드를 대부분 꼽는다. 이제 사회 초년생인 여성들은 모아둔 돈도 없을 것이기에 보유한 현금에 맞는 주거형태인 월세를 찾거나 잦은 이사 때문에 그럴 것이라고 예상한다.

하지만 결과는 전혀 다르다. 전세가 월등히 많이 검색되고, 때로는 월세보다 매매가 더 우위에 있다. 즉, 20대 초반부터 매매에도 많은 관심이 있고 매월 많은 지출을 해야 하는 월세보다는 전세를 선호하고 있다. 정부 차원의 다양한 전세 대출 정책이 존재하기에 더 그렇다.



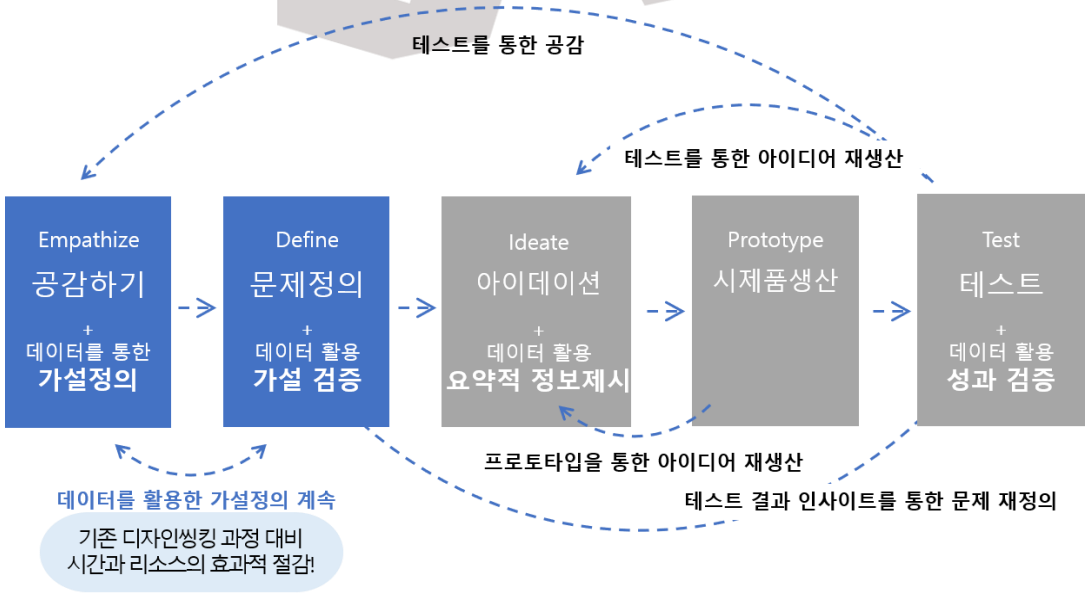
이처럼 편견은 무섭다. 우리의 눈을 가리고 귀를 막는다. “내가 겪어 봐서 잘 알아", “내가 그 분야에서만 10년 넘게 일했어. 내가 고객의 니즈는 제일 잘 알아". 디자인씽킹을 이러한 편견에 의심할 수 있는 기회를 제공한다. 물론 전문가라고 편견이 있고 비전문가라고 편견 없이 공정한 시각을 갖는 것은 아니다. 그렇기에

누가 문제를 해결하는지가 중요한 것이 아니다. 어떤 배경을 가진 사람들이 참여하건 디자인씽킹에서 중요하게 생각하는 공감을 철저하게 수행해야 한다. 진짜 문제를 정의해야한다. 절대로 디자인씽킹은 문제를 심플하게 보지 않는다. 심각한 고찰과 전문가의 식견도 무시하지 않는다. 하지만 제대로 공감을 수행하지 않으면 그러한 중요 부분을 무시해버리는 부작용이 발생될 수 있다. 디자인씽킹 회의론자들은 그 부작용을 보고 비판하고 있던 것이다.

그렇다면 어떻게 디자인씽킹의 부작용을 막을 수 있을까? 정답은 바로 데이터이다. 앞선 네이버 데이터랩 사례에서도 본 것처럼 데이터는 거짓말하지 않는다. 데이터는 편견이 존재하지 않는다. 데이터는 전문가들이 수십 년간 경험을 통해 얻어낸 인사이트를 상대적으로 아주 빠르게 얻어낼 수 있다. 즉 우리는 디자인씽킹의 단점을 보완하고 장점을 더 살리기 위해 데이터를 공감의 수단과 재료로 사용해야만 한다. 데이터는 우리에게 맥락 정보를 갖고 합리적인 추론을 하게 한다. 정성적인 방법을 통해서 공감을 할 때보다 고객의 가려운 점(Pain Point)를 쉽게 추출하고 정확하게 구별하게 한다. 게다가 정성적인 방법을 통해 공감을 시도할 때 흔히 저지르는 긍정 오류(False Positive)와 부정 오류(False Negative)를 방지할 수 있다.

정말 데이터를 통해 디자인씽킹이 가능할까? 10년 전 겨울 필자가 겪었던 사건을 하나 보자. 어느 날 한밤 중에 군고구마를 먹고 싶었다. 그런데 이상하게 그 해 겨울에는 동네를 아무리 돌아다녀도 군고구마 장사하시는 분이 보이지 않았다. 과연 왜 이러한 현상이

나타나는지 알고 싶어서, 데이터를 분석하기 시작하였다. 소셜 분석을 통해 군고구마의 연관 키워드를 살펴봤다. 상위 키워드 중, '직화냄비' 라는 키워드가 있었다. 에어프라이어가 대중화되기 이전에는 이 직화냄비가 집에서 군고구마를 만들 수 있던 제품이다. 그런데 마침 직화냄비는 그해 홈쇼핑에서 대박이 났다. 즉, 대체 수단이 생겨서 장사하시는 분들이 사라졌다고 볼 수 있다. 또한 공공데이터를 통해 고구마 가격의 추이를 봤더니 최근 몇 달간 가파르게 상승한 것을 알 수 있었다. 원가 상승에 따른 충분한 수익이 남지 않으니 군고구마 장사하시는 분들이 그 해 겨울에는 다른 장사 아이템을 찾게 된 것이다. 이처럼 데이터를 통해 우리는 공감하고 문제정의할 수 있다. 이를 아이디어 단계, 테스트 단계에도 적용할 수 있다면 디자인씽킹의 전체 프로세스를 더 특별하게 만들 것이다. 이러한 디자인씽킹을 위한 데이터 활용을 이제 조금 더 자세히 살펴보자.



Design Thinking with Data Process 5단계

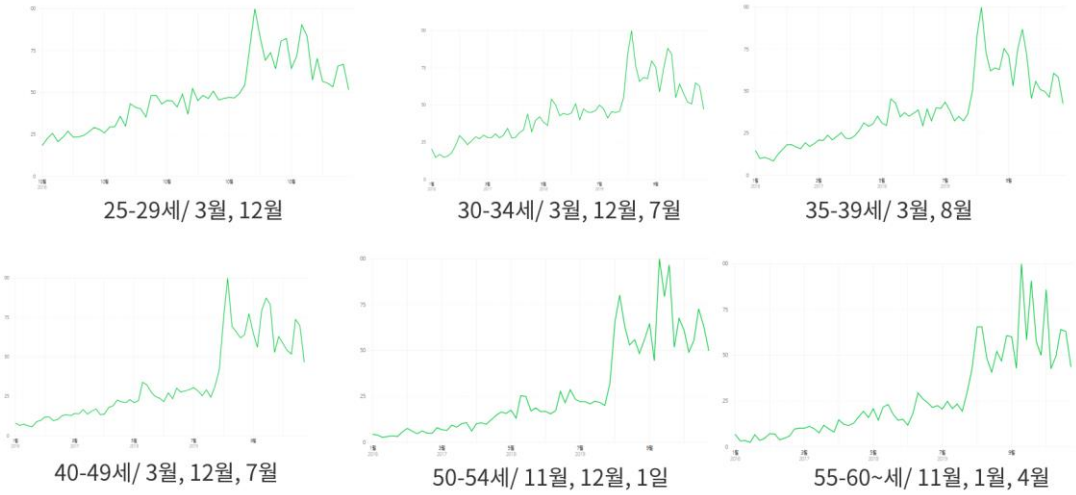
2. 사례로 알아보는 데이터 드리븐 디자인씽킹

2.1. [공감] 데이터 페르소나(고객) 정의하기

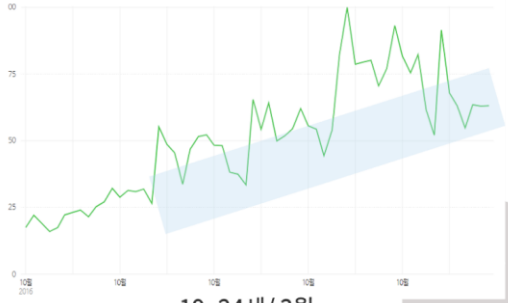
우리 주변의 데이터를 활용하여 페르소나(고객)를 쉽게 정의할 수 있다. 대표적으로 소셜 데이터와 설문조사 데이터 활용 사례를 살펴봄으로써 더 자세히 알아보고자 한다.

코로나바이러스가 장기화되면서 밀키트가 폭풍 성장하고 있다. 2017년부터 2021년까지 다년간 시장 규모가 100배가량 성장했다. 이에 스타트업 기업 S 사는 밀키트 시장의 고객을 파악하기 위해 소셜 데이터를 기반으로 페르소나를 작성했다. 먼저, S 사는 연령대별 키워드 검색 및 언급량 비교를 통해 밀키트를 많이 찾는 연령대를 파악했다.

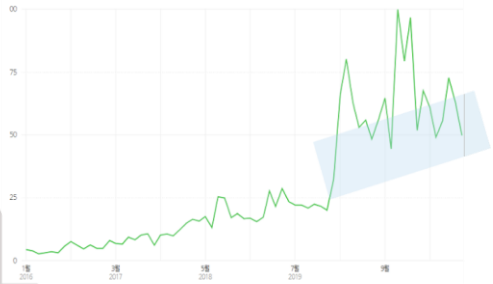
네이버에서 제공하는 포털사이트 검색어 분석 사이트인 네이버 데이터랩을 통해 살펴보도록 하자. 가장 먼저 주제어(밀키트)와 동의어(반조리, 간편식)의 2021년 1월부터 6월까지 데이터를 살펴보았다. 분석 결과, 밀키트에 대한 검색량은 전체 연령대에서 모두 상승 추이에 있으나, 공통적으로 2021년 6월에는 급격한 하락세가 나타났다.



또한 밀키트의 화제성 분석을 통해 10대 미만을 제외하고 20대부터 60대까지 언급량에 따른 관심 추이가 꾸준히 늘어나고 있으며, 그중에서도 상대적으로 관심도가 증가하고 있는 것은 19~24세, 50~54세 연령대임을 알 수 있다. 그 뿐만 아니라 가장 많은 검색량을 보인 시기가 19~24세의 경우 3월, 50~54세의 경우 11월, 12월, 1월 등 겨울철에 급증하는 것을 확인했다.



19~24세/ 3월



50-54세/ 11월, 12월, 1일

S 사는 이러한 데이터 분석 과정을 통해 그중에서도 20대 초반 연령대는 다른 모든 연령군과 달리 최근 관심도가 꾸준히 유지되며, 장기적으로 증가하는 추세라는 결과를 도출했다. 이후에는 19~24세 연령대가 밀키트를 많이 찾고 관심도가 급증하는 이유를 알아보기 위해 데이터 맥락 분석을 했다. 맥락 분석은 19~24세 연령대를 '대학생'으로 정의하고 키워드(대학생, 밀키트)를 구글에 검색하여 아래와 같은 정보를 파악했다.

밀키트 데이터 분석을 통해 얻은 인사이트 요약 정리

- ✓ 소소한 성취를 중시하는 MZ세대에게 밀키트를 통한 요리는 뿌듯함이라는 의미를 가진다.
- ✓ 요리법을 알려주는 유튜브 영향으로 밀키트의 접근도가 높다.
- ✓ MZ세대에게 요리는 여가활동이며 즐기는 콘텐츠다.

- ✓ 코로나로 인해 집에서 음식을 쉽고 빠르게 조리할 수 있어 활용성이 높다.
- ✓ 밀키트로 식사를 대체할 경우 배달비, 남은 음식 처리, 조리 시 물가 등 비용 걱정을 덜어준다.

이 외에도 S사는 네이버 데이터랩의 '냉동/간편 조리 식품' 데이터와 카카오 데이터 트렌드[2]를 통해 모바일 사용자, 여성, 경기-서울 지역의 대학생 집단의 밀키트 화제성이 높음을 확인했다. 이후 소셜데이터 분석 툴인 썸트렌드[3]를 활용하여 가장 많은 검색량을 차지한 3월의 인스타그램 채널에서 발생한 연관어 리스트를 비교하여 '홈파티', '먹스타그램' 등의 키워드가 많이 언급됨을 알았다. 일련의 과정을 통해 S사는 아래와 같은 페르소나를 도출할 수 있었다.

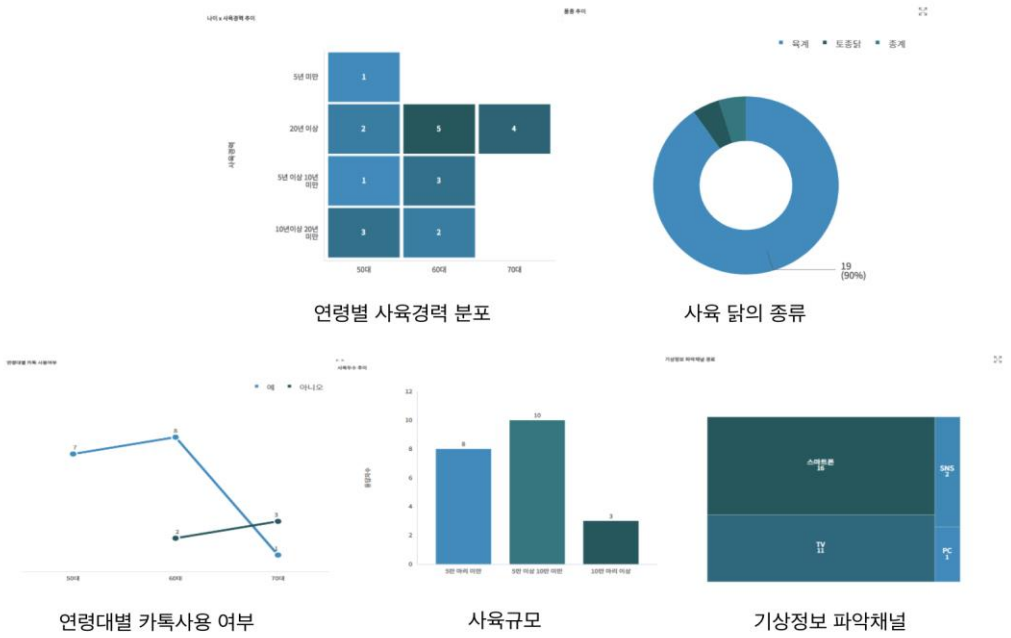
- 이름: 김서현
- 성별: 여
- 나이: 22세
- 직업: 대학생
- MBTI: ENFP
- 사용 기기: 아이폰
- SNS 채널: 인스타그램, 유튜브, 틱톡
- 라이프 스타일
  - 학교 근처 자취방에서 혼자 생활함
  - 요리를 하고 싶지만 재료비나 남은 음식 처리 측면에서 걱정됨
  - 대학 동기들과 함께 근교로 글램핑을 가거나, MT를 자주 감
  - 밀키트 사용에 대해 최근 관심이 높아짐

소셜 데이터 기반의 밀키트 핵심 페르소나

[2] datatrend.kakao.com

[3] some.co.kr

이렇게 데이터 기반의 고객 페르소나가 필요한 것은 비단 기업 사례뿐만 아니다. 데이터를 활용한 페르소나 정의는 국민들의 어려운 문제를 해결하고 더 나은 정책을 만들고 싶어하는 공공기관에서도 쉽게 찾아볼 수 있다. 매해 폭염 피해가 집중되는 충남 부여군의 축산 농가 중에서 가장 피해가 집중될 농가를 군집 분석을 통해 파악한 프로젝트가 있다. 사용한 데이터는 '2018년 충남 부여군 축산 및 가금류 농가 현황'으로, 부여군 축산업 현황에서 확인할 수 있었다. 물론 데이터만 있다고 하여 페르소나를 정의할 수 있는 것은 아니다. 데이터에서 확인할 수 있는 여러 가지 근거들을 바탕으로 모집단을 대표하는 '핵심 페르소나'를 뽑아내는 것이 관건이다. 이번 프로젝트에서는 데이터를 농가별 닭 사육 규모, 농가 주 사육업종, 농가 평균 연령을 기준으로 군집화하여 분석했다.



데이터 분석 결과

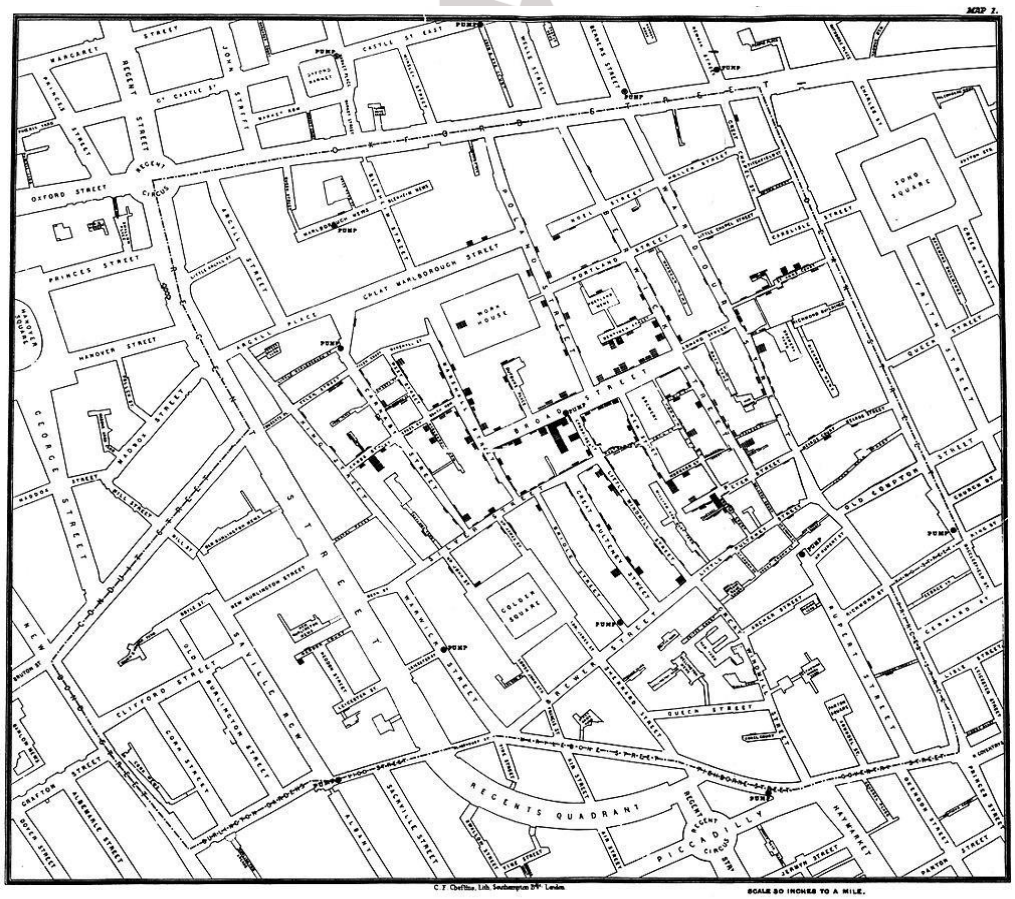
그 결과 비교적 소규모에 속하는 8천~1.4만 마리 육계 농가와 관리자의 연령이 80세 이하인 그룹을 인터뷰 및 관찰 조사를 위한 우선순위 대상으로 도출했다. 우선순위 집단이 된 80세 이하 소규모 농가의 핵심 니즈를 더욱 자세히 파악하기 위해 정성 조사를 실시했다. 정성 조사는 그룹인터뷰, 1:1 심층 인터뷰, 현장 관찰 인터뷰로 진행됐다. 그 결과 우선순위 집단 즉, 본 프로젝트 고객의 특징은 아래와 같이 정리할 수 있었다.

- 이름: 김양계
- 연령: 60대(사육경력 20년 이상)
- 기상정보 파악 채널: 스마트폰(1위), TV(2위)
- 폭염방지시설: 팬(1위), 안개 분무(2위)
- 특이사항: 평균 연령대가 높지만, 스마트폰을 이용한 정보접속에 용이

부여군 축산업 데이터 기반의 핵심 페르소나

2.2. [문제정의] 데이터로 문제 원인 찾아내기

데이터를 통한 창의적 사고방식, 그리고 문제해결은 언제부터 시작되었을까? 사실 역사를 거슬러올라가면 무려 150년도 전부터 시작한다. 앞서 데이터로 고객을 정의했던 사례와 같이 데이터로 문제를 이해하고, 문제의 패턴을 발견하고, 문제에 공감했던 사례가 있다. 바로 1854년 콜레라(Cholera)가 창궐했던 시기의 런던 유행병 학자였던 존 스노우(Dr. John Snow) 박사에 대한 이야기다.



존스노우가 지도상에 표시한 콜레라 환자

1800년대 런던에서는 콜레라가 주기적으로 창궐하고 있던 때였다. 당시까지만 해도 콜레라는 공기로 옮겨진다고 알려져 있었다. 하지만 존 스노우는 환자들의 증상을 살펴보며 '공기를 통한 전염'이라고 보기에는 도저히 설명되지 않는다고 믿었다. 공기를 통한 전염이라면 증상이 기관지 쪽에 있어야 하는데 다들 내장에서 증상을 보였기 때문이다. 이때 그는 문제를 해결하기 위해 데이터라는 새로운 시도를 해보기로 한다.

| 존스노우 박사와 콜레라 이야기

그는 우선 콜레라 전염 양상을 관찰하기 위해 신문 기사 자료를 모았다. 그리고 보도에서 활용한 기초 데이터를 확보했다. 하지만 데이터는 감염에 따른 사망자 수와 주소가 전부였다. 존 스노우는 단순히 사망자의 인적사항과 인원뿐인 데이터를 보고 난감했다. 난감함을 뒤로 하고 그는 사망자 주소를 지도에 표시하기 시작했다. 즉 그만의 데이터 차트를 제작한 것이다. 그 차트를 통해 존 스노우 박사는 거리의 특정한 펌프를 중심으로 콜레라가 전염되고 있다는 패턴을 발견했다. 또한 그는 가설의 확신을 위해 직접 길거리로 라 나가 사망자가 발생한 가구의 공용 펌프 사용 유무 및 펌프 종류를 물었다. 이러한 데이터 정리 작업에서 발견한 결과를 바탕으로 존 스노우 박사는 이후 콜레라는 펌프가 매개체가 되어 물을 통해 전염되는 수인성 질병이라고 문제를 정의할 수 있었다.

진짜 문제를 정의했다면 그 이후에는 이를 해결하기 위한 가장 쉽고 빠른 아이디어를 고안하는 것이다. 이를 위한 아이디어는 가장 쉽고 단순하게도 펌프의 손잡이를 제거하는 것으로 결정했다. 손잡이 제거 이후 콜레라의 확산은

거짓말처럼 억제됐다. 존 스노우의 이러한 행동은 오늘날 중요한 의미를 가진다. 콜레라가 물을 통해 전염된다는 증거를 보여주었을 뿐만 아니라 감염 지도를 통해 전염병의 원인을 파악하여 '역학'이라는 새로운 의학 분야를 만들어낸 것이다. 이후 그의 감염지도는 전염병 연구의 기본이 되었다.

존 스노우의 콜레라 사례는 우리에게 앞서 언급한 디자인 씽킹의 5단계를 명확히 따르되, 그 과정에서 데이터를 활용하는 것이 문제를 정의하는 데에 이점을 가져다 준다는 점을 제시한다.

2.3. [아이디어] 데이터로 상품 콘셉트 기획하기

최근 MZ세대들 사이에서 '플렉스(FLEX)' 라는 단어를 많이 사용한다. 주로 값비싼 물건을 구매하거나 호화스러운 생활을 보여줄 때 등장하는 말이다. 최근 기업에서는 이러한 MZ세대의 플렉스 트렌드에 발맞춰 상품을 기획하고 있다. 상품 기획 단계에서 다양한 데이터를 활용할 수 있는데, 트렌드와 특정 연령층의 관심사를 쉽게 파악할 수 있는 소셜데이터가 대표적이다. 소셜 데이터로 MZ플렉스 상품을 기획하는 단계를 살펴보자.

먼저 소셜 데이터를 수집할 채널 파악이 필요하다. 대학 내일 연구소가 2020년 10월 23일부터 28일까지 6일간 구조화한 설문지를 활용한 'Z세대의 인스타그램 · 페이스북 · 트위터 활용법' 보고서에 따르면 Z세대가 이용하는 채널 중 직접적으로 제품 및 서비스 구매와 연관된 것은 인스타그램과 트위터라고 한다. 두

채널을 중심으로 키워드(플렉스)에 대한 감성어 분석을 진행하고 실제 MZ세대 대상의 설문조사를 통해 검증 과정을 거쳐보았다.

감성어 분석은 상품 및 서비스, 기관이나 단체, 사회적 이슈, 사건 등에 관하여 소셜미디어 등에 남긴 의견을 수집하고 분석하는 것을 의미한다. 이는 사람들의 감정의 상태 및 태도에 대한 변화, 의견과 평가, 선호 등을 파악하는 데에 용이하다. 키워드(플렉스)의 감성어 분석은 '좋다, 강력하다, 구매하다, 미치다, 귀엽다, 예쁘다, 사랑스럽다, 돈 벌다, 찐이다, 소중하다, 비싸다'의 키워드를 도출할 수 있었다. 이러한 단어들을 바탕으로 가설을 세워보면, 플렉스는 '돈 벌어서 구매하는 비싸지만 미치도록 놀랍고 소중하고 사랑스럽고 예쁘고 좋은 찐이라고 볼 수 있는 제품'인 것이다.

다음으로는 이 가설을 검증하기 위하여 실제 MZ세대 대상의 설문조사를 실시했다. 그들은 '플렉스' 하면 떠오르는 것으로 '명품(지갑, 시계), 전자기기, 신발, 호캉스 등'이라고 답했다. 간단한 감성 연관어 분석을 통해 정확히 정의되지 않는 '플렉스'의 뜻을 알아보았다. 연관어란 두 개의 단어가 주어진 문맥에서 서로 얼마나 연관되어 있는지를 의미한다. 따라서 연관어 분석은 연관 알고리즘을 통하여 단어 사이의 관계를 파악해 시각화하는 분석을 뜻한다. 이러한 연관어 분석을 통해 상품 기획 시 참고할 수 있는 플렉스의 특징에 대해 살펴볼 수 있었다.

'플렉스' 감성 연관어 TOP10 [왼쪽부터 2021년 6월, 5월, 2월 순]

또한 데이터를 활용해 상품을 성공적으로 기획한 사례도 있다. 새로운 소비 트렌드에 영향을 미치고 있다고 평가되는 다이슨의 에어랩이다. 59만 9천 원의 고가에도 불구하고 품귀 현상을 보이는 다이슨 에어랩이 인기를 끄는 비결은 바로 '데이터'에 있었다. 다이슨은 신제품 출시를 앞두고 전 세계 80개의 미용실에서 420명 이상의 전문 미용사가 1만 시간에 걸쳐 진행한 11억 건이 넘는 데이터를 분석했다. 11억 건의 데이터 분석 결과 미용사들은 둥근 브러시를 사용해 모발의 모양을 잡는 동시에 드라이어의 공기 흐름을 이용해 잔머리를 안쪽으로 밀어내는 패턴을 확인했다. 즉, 헤어드라이어의 각도가 매우 중요하다는 것이다. 다이슨 데이터 분석팀은 이를 기반으로 페르소나를 아래와 같이 설정했다. 페르소나를 기반으로 반원 모양의 플라이어웨이 노즐을 개발했다. 또한, 압력 차이로 인해 물체 표면에 모발이 달라붙는 현상인 '코안다 효과'를 이용했다. 그 결과 드라이어만으로 머리카락을 매끈하게 정리해주는 다이슨 에어랩이 출시된 것이다. 그리고 이 에어랩은 수 억 건의 데이터와 사용자 니즈 분석을 거쳐 페르소나를 아주 구체적이고 상세하게 '가정에서 직접 머리 손질을 할 때

드라이어 각도 조절이 어려워 머리카락 표면이 부스스해진 경험이 있음. 부스스함을 정리하기 위해 고온의 기기를 사용해 머리카락 표면을 정리하게 되고, 이것이 모발 손상으로 이어지는 경험을 한 여성’으로 정의할 수 있었다.

다이슨 애어랩 스타일러. [사진 제공 = 다이슨]

여기까지 데이터 드리븐 디자인씽킹 사례를 살펴보았다. 하지만 데이터를 활용한다는 것이 막막할 것이다. 비전공자도 손쉽게 데이터를 활용할 수 있는 방법을 다음 챕터에서 알아보자.

3. 데이터를 활용한 디자인씽킹 실전 적용 방법

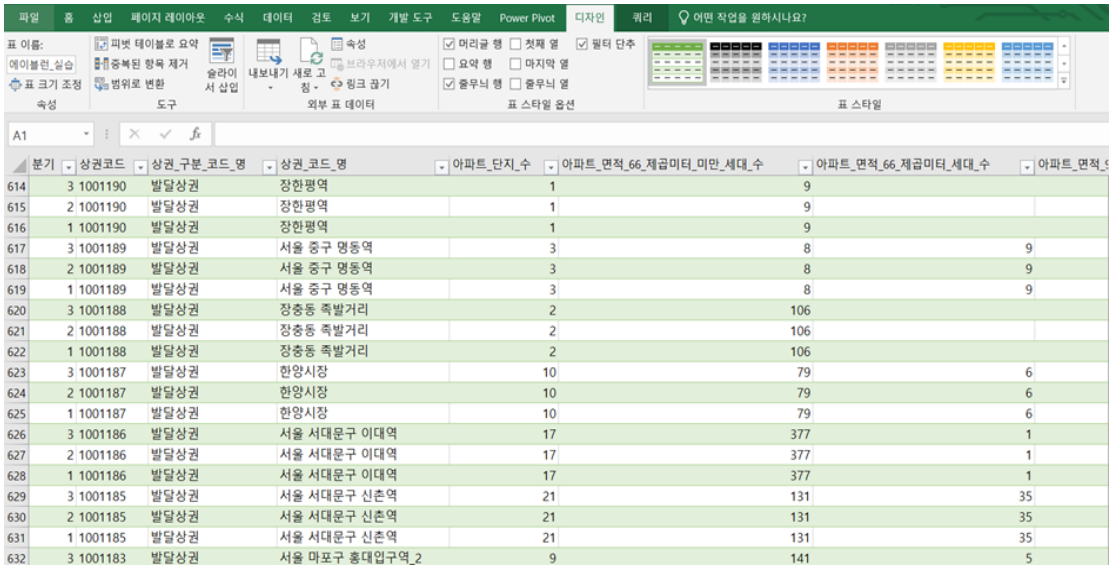
데이터기반의 디자인씽킹을 위해 노코드로 데이터를 활용하는 다양한 방법을 소개하고자 한다. 엑셀을 활용한 데이터 수집, 키워드를 활용한 소셜 데이터 탐색과 분석, 나아가 시각화를 통한

탐색적 데이터 분석까지 모든 과정을 코딩없이 할 수 있는 방법을 살펴보자.

3.1. 엑셀을 활용한 노코드 데이터 수집

전문적인 데이터 분석 툴의 활용법이나 코딩을 모르는 비전공자도 엑셀을 활용하면 노코드로 데이터를 불러오고 정제하는 것뿐만 아니라 다양한 분석기법을 활용하여 인사이트를 얻는 것까지 쉽게 가능하다. 데이터 기반의 디자인씽킹을 위해 노코드로 데이터를 활용하는 다양한 방법을 소개하고자 한다. 우선 엑셀을 활용한 데이터 수집이다.

실제 실무에는 수많은 데이터들이 존재하지만 그 중에서도 제대로 된 데이터 분석을 위해 잘 '정제' 되어 있는 데이터는 흔치 않다. 대부분 파일별로, 폴더별로 분산되어 있거나 보고서형식에 맞게 정리되어 있을 뿐 데이터 분석을 위해 원본 데이터(Raw data) 형태로 열과 행, 레이블을 적절히 갖추는 것까지 준비하기는 어렵기 때문이다. 이를 위해 데이터 정제의 과정이라고 할 수 있는 ETL, 즉 추출(Extract), 변환(Transform), 로드(Load) 과정이 필요하다. 이 과정을 엑셀 파워쿼리를 활용하여 수행할 수 있으며 이를 통해 조직 내 존재하는 비정형, 반정형 형태의 데이터를 정형 데이터 및 테이블로 변형하여 고객 문제에 대하여 공감하여 니즈를 발견하는 재료로 사용할 수 있다.



	분기	상권코드	상권_구분_코드_명	상권_코드_명	아파트_단지_수	아파트_면적_66_제곱미터_미만_세대_수	아파트_면적_66_제곱미터_세대_수	아파트_면적_
614	3	1001190	발달상권	장한평역	1	9		
615	2	1001190	발달상권	장한평역	1	9		
616	1	1001190	발달상권	장한평역	1	9		
617	3	1001189	발달상권	서울 중구 명동역	3	8	9	
618	2	1001189	발달상권	서울 중구 명동역	3	8	9	
619	1	1001189	발달상권	서울 중구 명동역	3	8	9	
620	3	1001188	발달상권	장충동 족발거리	2	106		
621	2	1001188	발달상권	장충동 족발거리	2	106		
622	1	1001188	발달상권	장충동 족발거리	2	106		
623	3	1001187	발달상권	한양시장	10	79	6	
624	2	1001187	발달상권	한양시장	10	79	6	
625	1	1001187	발달상권	한양시장	10	79	6	
626	3	1001186	발달상권	서울 서대문구 이대역	17	377	1	
627	2	1001186	발달상권	서울 서대문구 이대역	17	377	1	
628	1	1001186	발달상권	서울 서대문구 이대역	17	377	1	
629	3	1001185	발달상권	서울 서대문구 신촌역	21	131	35	
630	2	1001185	발달상권	서울 서대문구 신촌역	21	131	35	
631	1	1001185	발달상권	서울 서대문구 신촌역	21	131	35	
632	3	1001183	발달상권	서울 마포구 홍대입구역_2	9	141	5	

엑셀 파워쿼리를 활용한 데이터 정제 예시

개인 또는 기업이 이미 보유하고 있는 데이터를 활용하는 방법도 있지만, 내가 분석하고자 하는 목적에 부합하는 적절한 데이터를 가지고 있지 않은 경우도 있을 수 있다. 이런 경우에는 웹상에 존재하는 무수히 많은 데이터를 활용할 수 있다. 웹이라 불리는 월드 와이드 웹(World Wide Web, WWW)상에 존재하는 정보들을 끌어오는 방법은 여러가지가 있는데, 대표적인 예가 직접적으로 웹페이지내 테이블, 텍스트 형태 등으로 존재하는 데이터를 복사해서 가져오는 것이다. 하지만 정보의 업데이트가 불가능하고, 대량의 빅데이터를 가져오기에는 적절치 않다는 단점이 존재한다.

이것을 극복하기 위해서는 봇이나 소프트웨어 에이전트의 한 형태인 '웹 크롤러'가 필요하다. 웹 크롤러를 통해 우리는 '웹 크롤링'이 가능하며, 웹 크롤링을 통해 방문한 사이트의 모든 페이지의 복사본을 생성하여 데이터를 가져오는 것과 동시에 주기적으로 해당 URL을 방문하여 복사된 데이터를 최신 상태를 유지할 수 있다.

다만, 이 웹 크롤러는 앞서 이야기했다시피 봇이나 소프트웨어 에이전트의 일종으로 인식되므로 크롤러를 개발하기 위해서는 코드를 통한 프로그래밍 작업이 반드시 필요했다. 즉, 웹 상의 데이터를 가져오기 위해서는 개발자나 IT 전문가의 도움을 받아야만 데이터 수집과 준비가 가능했다는 것이다. 하지만 엑셀에 내재되어 있는 웹크롤링 기능을 활용하면 코딩 없이 간편하게 웹 상의 빅데이터를 수집하고 불러올 수 있다. 이를 통해 웹에 존재하는 다양한 데이터를 활용하여 데이터 기반 공감 단계에서 문제 도출에 필요한 재료로 사용할 수 있으며, 반복적인 수집 작업을 자동화 해두어 효율적인 데이터 드리븐 디자인씽킹 작업이 가능하다.

	A	B	C
1	순위	국가/령	GDP (백만달러)
2		세계	83844988
3	1	미국	20807269
4	2	중화인민공화국	14860775
5	3	일본	4910580
6	4	독일	3780553
7	5	영국	2638296
8	6	인도	2592583
9	7	프랑스	2551451
10	8	이탈리아	1848222
11	9	캐나다	1600264
12	10	대한민국	1586786
13	11	러시아	1464078
14	12	브라질	1363767

	세계[4]	83,844,988
1	미국	20,807,269
2	중화인민공화국	14,860,775
3	일본	4,910,580
4	독일	3,780,553
5	영국	2,638,296
6	인도	2,592,583
7	프랑스	2,551,451
8	이탈리아	1,848,222
9	캐나다	1,600,264
10	대한민국	1,586,786
11	러시아	1,464,078
12	브라질	1,363,767

웹데이터를 엑셀로 가져와 불러온 모습 예시

공공데이터 포털 및 지자체 데이터 개방사이트 등을 통해 공공 빅데이터를 활용하는 방법도 있다. 공공데이터는 국가에서 보유하고 개방하는 데이터를 일컫는 말로, 거시 경제 지표부터 각종 생활소비 분야에서 생성되는 인구 및 문화 데이터까지 다양한

데이터가 존재한다. 실제로 실시간 교통정보 내비게이션, 화장품 성분 및 가격정보 비교, 갑자기 아플 때 유용한 병원찾기 등을 목적으로 각종 공공데이터 기반 서비스 모델이 다수 개발·활성화되고 있는 것은 대표적인 활용 사례라고 할 수 있다. 또한 데이터가 많을수록 보다 풍부하게 가설을 수립하고 검증할 수 있다는 점을 고려하면 공공데이터를 사용할 가치는 충분하다.

이러한 공공데이터는 이곳 저곳 산재되어 있는 다른 웹 데이터들과 달리, 데이터 공개 및 제공이라는 뚜렷한 목적성을 가지고 카테고리별로 분류되어 있으며 데이터를 활용하기 용이하도록 다양한 형태로 데이터를 제공한다. 필요한 형태에 맞게 직접 데이터를 다운로드하거나, 오픈 API(Application Programming Interface: 응용 프로그램 프로그래밍 인터페이스)를 활용하여 빅데이터를 불러오는 방법도 있다. 마찬가지로 엑셀을 활용하여 코딩없이 공공 빅데이터를 불러와 활용할 수 있다.

기본정보

데이터명	국민연금 가입 사업장 내역 **상세설명**		
서비스유형	REST	**심의여부**	자동승인
신청유형	개발계정 | 활용신청	**처리상태**	승인
활용기간	2021-10-14 ~ 2023-10-14		

서비스정보

데이터포맷	XML
End Point	http://apis.data.go.kr/B552015/NpsBplcInfoInqireService
API 환경 또는 API 호출 조건에 따라 **인증키가 적용되는 방식이 다를 수 있습니다.** 포털에서 제공되는 **Encoding/Decoding** 된 인증키를 적용하면서 구동되는 키를 사용하시기 바랍니다. * 향후 포털에서 더 명확한 정보를 제공하기 위해 노력하겠습니다.	
일반 인증키 (Encoding)	MVRSqdCaFlZECgujz4qZQFINpq%2FqI7I%2F1QTmtHgd%2F38vJ861iVzYMGiLH6cqLa8evD2SNS4%2Bu576TbR%2BszmKCA%3D%3D	
일반 인증키 (Decoding)	MVRSqdCaFlZECgujz4qZQFINpq/qI7I/1QTmtHgd/38vJ861iVzYMGiLH6cqLa8evD2SNS4+u576TbR+szmKCA==	

공공빅데이터 API 활용 정보 예시

위와 같은 방법들을 통해 데이터가 준비되었다면 데이터를 활용하여 다양한 분석을 진행해 볼 수 있다. 데이터 분석은 단순히

데이터의 기술 통계값을 구해 데이터의 특성을 파악하는 것뿐만 아니라 가설을 통해 우리가 설정한 목표나 과제를 과학적으로 검증하고 해결하는 과정을 포함한다. 이중 '가설을 수립' 하는 일은 데이터 분석에서 특히 중요한 부분이다. 가설이 올바르게 수립되지 못한다면 문제해결을 위해 상당한 물적, 인적 자원을 낭비하게 되기 때문이다. 하나 많은 실무자들이 '문제' 를 정확히 파악하고 고민하기 보다 그저 '주어진 데이터' 만 보고 가설을 수립해버리는 경우가 많다. 이러한 잘못을 범하지 않으려면 '주어진 데이터' 간에 특징과 영향요인들을 충분히 검토할 수 있어야 한다.

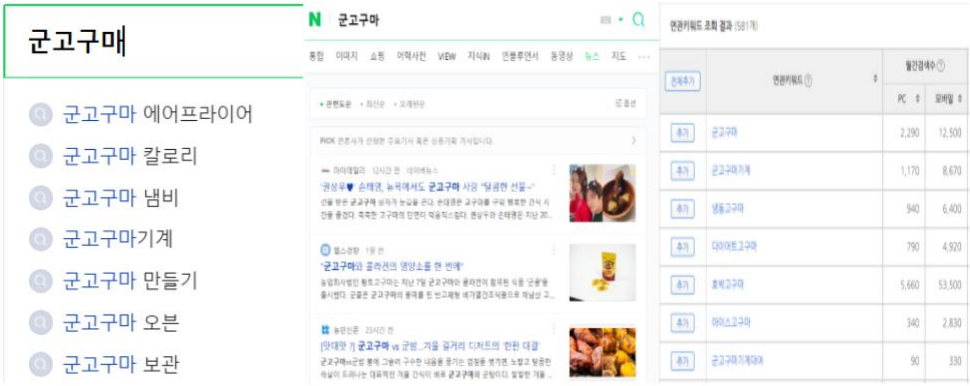
이렇게 데이터간의 상관관계와 인과관계를 파악하기 위해 필요한 분석 기법이 바로 상관분석과 회귀분석이다. 이를 통해 우리는 서로 다른 데이터 간의 상관성을 분석하거나 숨겨진 의미를 찾아내고, 앞으로의 관계를 예측하는 일 등이 가능하다. 상관 분석과 회귀 분석을 활용하여 우리가 데이터 드리븐 디자인씽킹을 통해 해결하고자 하는 문제와 최종 목표에 부합하는 적절한 가설을 보다 객관적이고 과학적인 방법으로 설정해볼 수 있는 것이다. 가설이 설정되었다면 이제 다양한 통계검정방법들을 통해 가설을 검증하고 그 결과를 활용한다. 기술통계 값을 구해보는 것 뿐만 아니라 Z검정, T검정, 분산분석, 정규화 및 표준화나 상관분석, 회귀분석, 그리고 다중회귀분석까지 이 모든 일련의 과정들을 엑셀을 통해 누구나 코딩 없이 가능하다.

3.2. 소셜 데이터를 통한 탐색과 분석

핵심 키워드를 활용한 데이터의 탐색 및 검색을 활용하는 방법으로도 디자인씽킹 과정에서의 다양한 인사이트를 얻을 수 있다.

데이터 탐색의 시작은 '키워드'다. 어떠한 핵심 키워드를 선택하느냐에 따라서 원하는 결과를 얻을 수도, 얻지 못할 수도 있다. 예를 들어 군고구마 같은 구체적인 주제는 일반적으로 원하는 검색 결과를 얻기 쉬운 편이지만, B2B 산업이나 트렌드는 어떨까? 당초 키워드를 어떻게 설정하고 찾을 지부터 어렵고 막막한 경우가 있을 수 있다. 그런 점에서 내가 어떤 목적으로 어떤 데이터를 확인하고 싶은지 명확히 하고, 필요한 데이터를 얻기 위해 어떠한 검색 키워드를 사용할지 정하는 것이 중요하다. 키워드는 내가 찾고자 하는 주제, 데이터로 확인하고 싶은 목적과 밀접하게 연관된 핵심적인 단어일수록 좋다.

좋은 검색 키워드를 찾기 위해서 다양한 방법을 사용할 수 있다. 네이버 연관 키워드 검색 시, 자동 완성되는 목록을 확인해 본다거나 뉴스 기사를 통해 연관 키워드를 확인하는 방법 혹은 네이버 키워드 광고 도구를 활용하여 연관 키워드를 조회해보는 방법 등이 있다.



네이버 연관어 키워드 조회결과

이러한 방법들을 통해 우리가 획득해야 하는 3가지 키워드의 유형이 있다. 첫째는 얻고자 하는 데이터 주제와 가장 밀접한 연관을 지니고 가장 양질의 데이터를 보여줄 수 있을 핵심 키워드이고, 둘째는 이 핵심 키워드와 관련된 보다 다양한 데이터를 찾기 위한 핵심 키워드의 동의어 및 포함어다. 셋째는 부정적인 단어 및 광고성 데이터를 제외할 수 있는 제외어 키워드이다. 핵심 키워드, 동의어/포함어 그리고 제외어를 제대로 파악해야 우리가 알고자 하는 주제에 최적화된 데이터를 획득할 수 있다.

필요한 키워드를 모두 선별하고 준비했다면 이제 이 키워드들이 왜 생겼고, 군고구마라는 핵심키워드의 무엇과 관련 있는지를 맥락적으로 파악하기 위한 배경·현상 조사가 필요하다. 키워드와 관련된 원문 정보를 분석하는 방법과 연관어 비교분석을 통해 유추하는 방법을 통해 맥락을 파악할 수 있다.

조사의 시작은 검색이다. 적절한 키워드를 검색한 뒤, 검색을 통해 얻어낸 정보들을 분석해 본다. 효율을 높이기 위해서는 먼저 시장과 산업 등 거시적인 분석을 한 후에 그보다 작은 현상들을 분석하는 순서로 조사하는 것이 일반적으로 좋다. 예를 들어 '군고구마', '군고구마 트렌드', '군고구마 2022년' 등과 같은 키워드를 활용하여 검색되는 다양한 기사를 살펴보는 것이다. 기사를 하나하나 열어보지 않더라도 제목을 통해 흐름을 유추해볼 수 있다.

네이버 뉴스 키워드 조회결과

맥락분석을 더 잘하고 싶다면 타겟에 대한 구체적인 정보나 원하는 목적에 대한 이슈 키워드를 추가해서 입체적인 분석을 해야 한다.

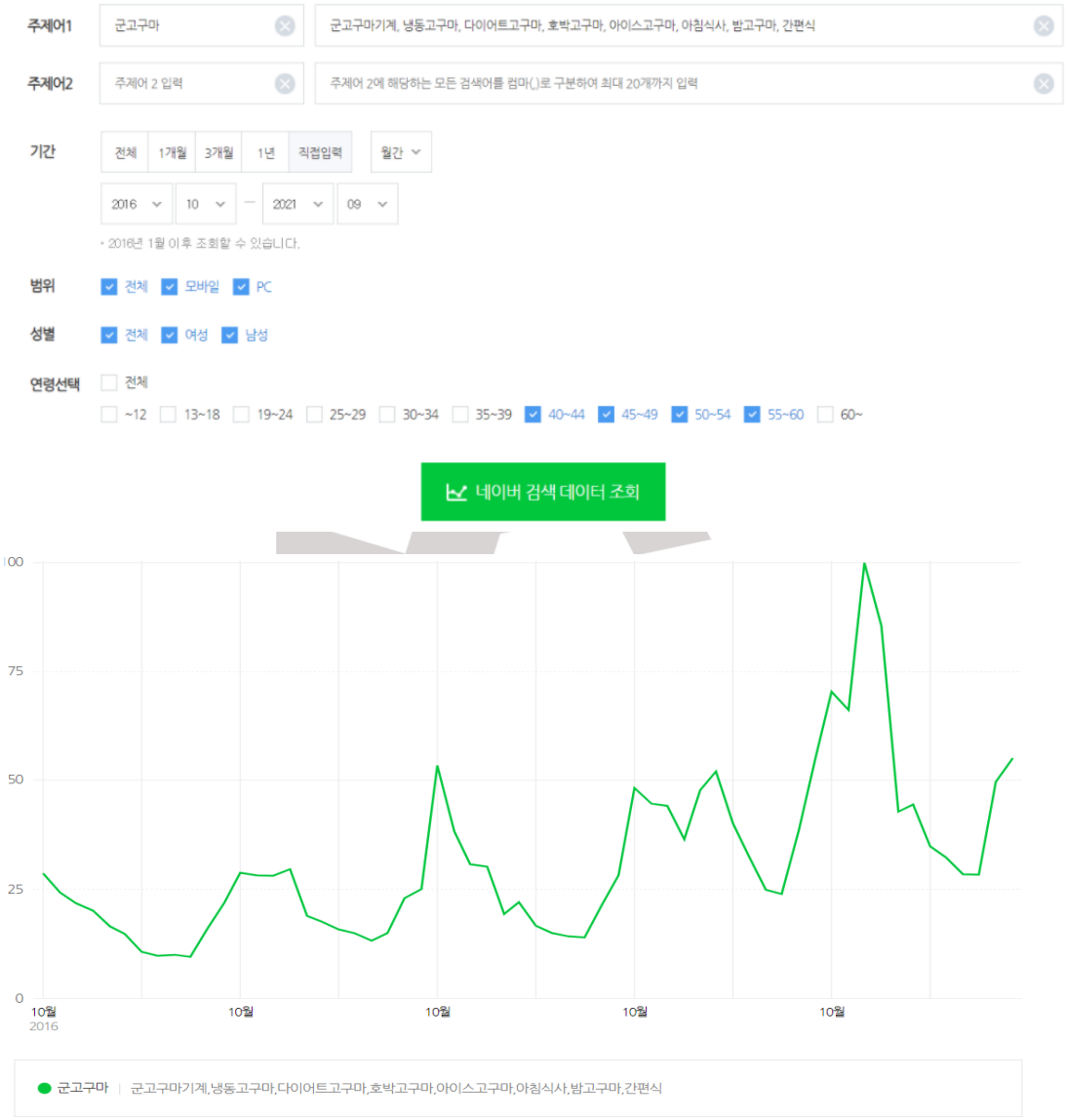
타겟 조사를 할 때는 네이버보다 구글을 활용하는 것이 더 효과적이다. 검색 알고리즘의 특성상, 네이버가 블로그 · 포스트 등 자사 콘텐츠를 노출하는 경향이 있다면, 구글은 사용자가 많이 찾거나, 여러 사이트에서 언급되는 링크를 더 많이 노출하기 때문이다. 이때 타겟의 조건을 키워드로 넣으면 더욱 입체적인 검색 결과를 얻을 수 있다. 성별, 연령, 지역 등 인구통계학적 조건이 대표적이다. 예를 들어 군고구마에 대한 향수가 존재하는 40대, 50대가 군고구마에 대한 관심도가 높을 것이라는 가정 하에 '군고구마 40대 분석', '군고구마 50대 여성 분석' 등으로 검색하면 조금 더 뾰족한 검색 결과를 얻을 수 있는 것이다.

핵심 키워드 그룹군을 추출하고 원문분석을 통해 전반적인 맥락을 파악했다면 해당 주제어를 바탕으로 구체적인 데이터 수치를 확인하여 핵심 주제가 시장(고객)의 관심을 얼마나 얻고 있는 지를 확인해 볼 필요가 있다. 기사나 자료가 아무리 많이 나왔더라도 소비자들이 관심을 갖지 않는다면 큰 의미가 없기 때문이다. 핵심 키워드의 언급량을 확인하여 화제성을 분석해보는 것이다.

이 외에도 키워드 언급량 등 다양한 분석 데이터를 얻을 수 있는 여러가지 툴이 있다. 이러한 툴들을 활용할 때 중요한 점은, 각각의 툴에 따른 장단점이 다르므로 해당 툴의 특징을 파악하여 용도에 맞게 적절히 활용할 수 있어야 한다는 것이다.

1) 네이버 데이터랩

네이버 데이터랩에서는 궁금한 '주제어'를 설정하고, 하위 주제나 동의어에 해당하는 검색어를 함께 입력하면 쉽게 언급량 추이를 볼 수 있다. 다만, 키워드를 너무 많이 넣었을 경우 신뢰성이 저하될 수도 있으니 주의해야 한다. 또한 그래프의 기간 단위를 다음과 같이 일별, 주별, 월별로 설정할 수 있다. 또한 채널이나 타겟의 연령대, 지역 등을 설정할 수도 있다.

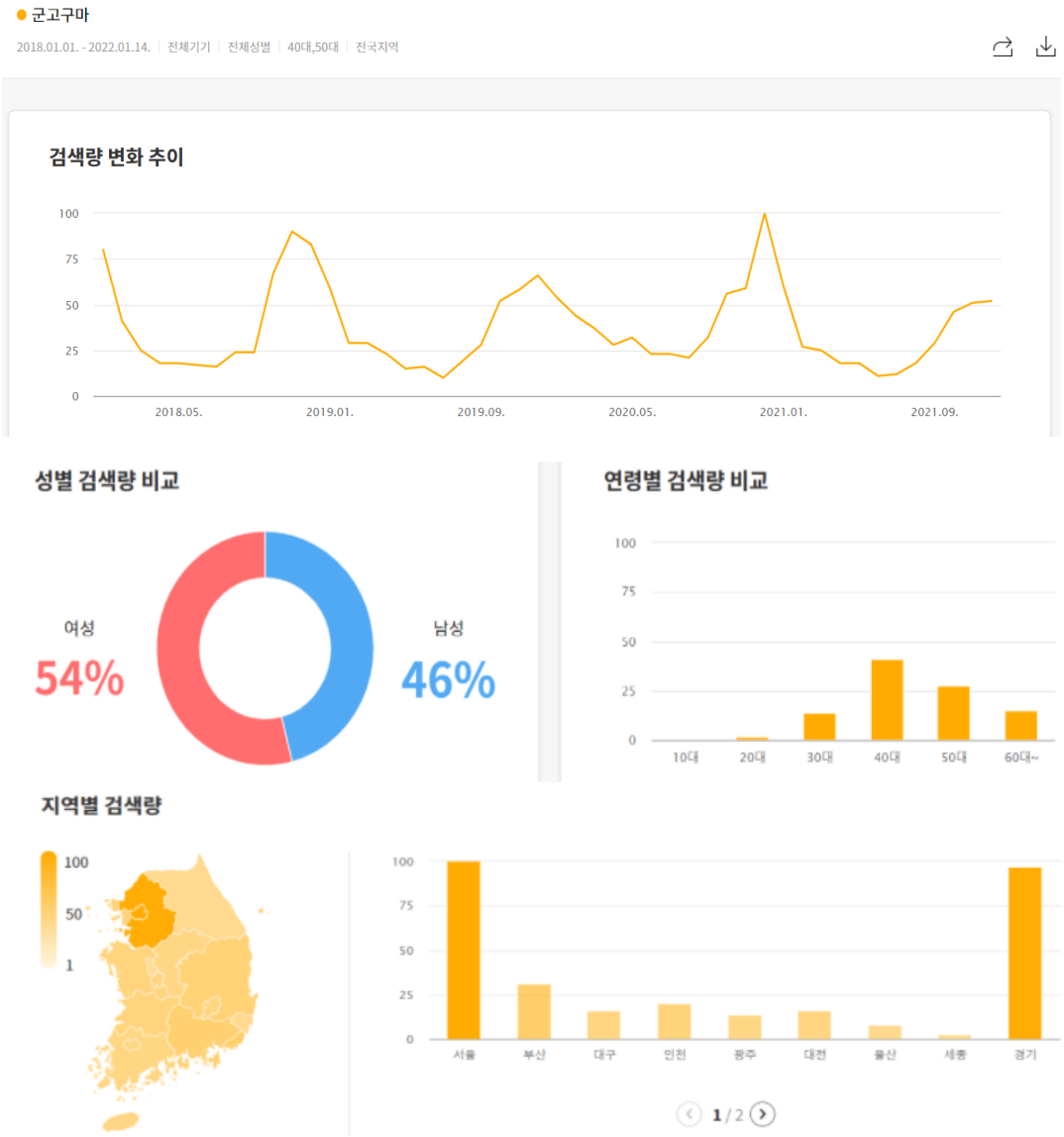


네이버 데이터랩을 활용하기 좋은 조건들

- ✓ 구체적인 연령대를 특징 짓고 싶을 때
- ✓ 해당 주제에 대해 포털사이트에서 사람들이 주로 검색하는 행위가 충분하다고 판단될 경우
- ✓ 1년 이상의 키워드에 대한 전반적인 동향을 필요로 할 때
- ✓ 여러 주제어와 동시에 비교분석 할 때
- ✓ 데이터 분석 결과에 따른 검색량 원본 데이터를 얻고 싶을 때
- ✓ '주로 알고 싶은 정보의 유형이 얼마나 많이 언급되고 있을까?', '언제 많이 언급됐을까?'와 같은 것일 때

2) 카카오데이터트렌드

카카오 데이터 트렌드를 활용하면 최대 5개의 주제어를 동시에 분석해 볼 수 있다. 네이버 데이터랩과 달리 하위 주제어는 포함이 불가능지만, 기타 검색옵션을 상세히 조정 가능하다. 또한 단순히 검색량 트래픽 외에도 해당 검색어에 대한 [성별], [연령별], [지역별] 비교 및 원본 데이터 다운로드가 가능하다.



카카오 데이터 트렌드를 활용하기 좋은 조건들

✓ 연령별 / 성별 / 지역별 자세한 검색조건을 원하는 경우

✓ 조사하고자 하는 데이터가 2018년 이후로 충분한 양이 쌓였다고 판단될 경우

✓ 데이터분석결과에 따른 원본 데이터를 얻고 싶을 때

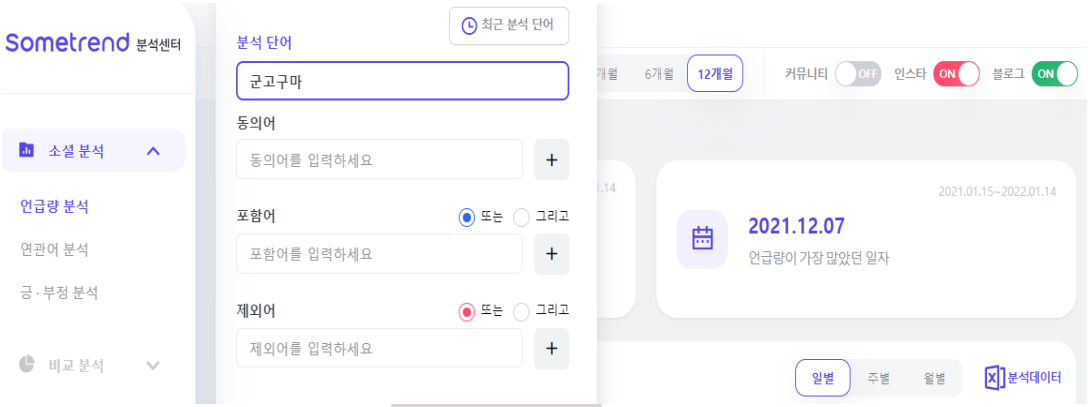
✓ '주로 알고 싶은 정보의 유형이 누가 많이 언급했을까?', '어디서 많이 언급됐을까?' 와 같은 경우

3) 썸트렌드

썸트렌드는 소셜 빅데이터를 분석할 수 있는 유용한 툴이다. 우리는 페이스북, 트위터, 인스타그램 등 SNS나 다양한 인터넷 플랫폼을 통해 본인의 생각을 표현하고 취향을 표출하는 시대에 살고 있다. 소셜데이터는 이러한 SNS 게시글이나 댓글 등의 빅데이터를 분석하여 트렌드를 파악하고 연령대별 혹은 기간별, 성별에 따라 어떤 취향 및 생각을 가지고 있는지 파악할 수 있도록 해준다. 사람들의 생각을 파악하기 위해서 사용했던 기존의 설문조사/ 여론조사 방식은 대상자를 물색하고 원하는 데이터를 얻기 위한 양질의 질문을 설정해야 하며, 이에 대한 응답을 받고 처리하는 등 많은 시간과 노력이 필요한 반면, 소셜 데이터는 어떤 키워드를 통해 어떤 목적으로 데이터를 확인하고자 하는지 명확히 설정할 수만 있다면, 이미 SNS 상에 존재하는 소비자 스스로가 만들어 둔 자료 속에서 우리가 필요한 데이터를 보다 쉽게 찾고 확인할 수 있다.

썸트렌드의 키워드 검색은 키워드와 검색기간을 설정할 수 있을 뿐 아니라 동의어, 포함어, 제외어를 통해 검색 키워드를 구체화할 수 있다. 예를 들어, 동의어로 ‘bigdata’를 입력하면, ‘빅데이터’ 또는 ‘bigdata’를 언급한 게시글들 모두 분석한 결과를 확인할 수 있다. 포함어로 소셜을 입력하면, ‘빅데이터’와 ‘소셜’이 함께 언급된 게시글을 분석한 결과를 확인할 수 있고, 반대로 제외어에 소셜을 입력하면, ‘빅데이터’와 ‘소셜’이 함께 언급된 경우를 제외하고 남은 게시글을 분석한 결과를 확인할 수 있다. 광고성 글을 제외하기 위해, 이벤트 등을 제외어로 입력해 주는 것도 광고 및 홍보성 자료를 거르기 위한 방법 중 하나다.

다만, 시간에 따른 추이를 나타내는 라인그래프를 제대로 보기 위해 목적에 따라 채널별/시간별로 보는 것이 적합하다. 채널 및 기간을 어떻게 설정하느냐에 따라 동일한 데이터 조차도 검색결과가 크게 달라질 수 있기 때문이다. 또한 같은 검색 결과라도 그래프 추이의 [보기]를 일/주/월 중 어느 것으로 하느냐에 따라 다른 인사이트를 볼 수 있다.



또한 썸트렌드의 소셜분석을 통해 키워드 언급량과 연관어 분석 뿐만 아니라 키워드에 대한 긍정적 반응, 부정적 반응과 같은 부분을 알 수 있는 감성어 분석도 가능하다. 이는 소셜데이터가 SNS 상에 자유롭게 표현된 개인의 생각과 취향을 통해 얻어지는 데이터이기 때문이다. 키워드를 입력하고 긍, 부정어 분석을 하면 해당 키워드에 대한 긍정적인, 혹은 부정적인 키워드들을 확인할 수 있다. 이 때 중요한 것은 말의 뉘앙스를 염두에 두어야 한다는 것이다. 예를 들어 단순히 좋다/싫다와 같이 명확한 표현이 아닌, 크다/ 작다 혹은 높다/ 낮다와 같이 상황에 따라 표현의 긍부정이 달라질 수 있는 경우가 있다. 또는 ‘ㅋㅋㅋㅋㅋ’ 라는 표현과 같이 단순히 긍정적인 웃음을 표현하는 것이 아닌 ‘어이없음’ 등과 같은 부정적인 뜻을 내포하고 있는 경우도 많이 있다. 이런

경우에는 글의 원문을 확인하여 단어가 사용된 뉘앙스를 파악해야 하며, 썸트렌드에서 원문 또한 한번에 확인이 가능하다.

3.3. 시각화를 통한 탐색적 데이터 분석

탐색적 데이터 분석은 가설 등을 미리 설정하고 그에 맞는 데이터 수집과 통계를 분석하는 것이 아닌, 데이터를 수집하고 수집한 데이터를 토대로 시각화 하여 각각 수집한 데이터의 패턴을 도출하여 결론 혹은 현상을 발견할 수 있는 데이터 분석 방법이다. 데이터가 들어왔을 때 다양한 각도에서 관찰하고 이해하는, 한마디로 데이터 자료를 직관적으로 바라보는 과정이다. 탐색적 데이터 분석은 이론적 배경이나 선행 연구에 근거하기 보다는 데이터가 보여주는 결과 자체를 그대로 받아들이게 되므로 이론 생성 과정에 가깝고, 데이터 지향적인 성격을 띈다고 할 수 있다. 가설도 없고, 모형도 없이 원데이터(Raw data)를 가지고 유연하게 데이터를 탐색하고, 데이터의 특징과 구조로부터 얻은 정보를 바탕으로 통계모형을 만드는 분석방법이다. 수집한 데이터가 들어왔을 때 이를 있는 그대로 다양한 각도에서 관찰하고 이해하는 과정이라고 할 수 있다.

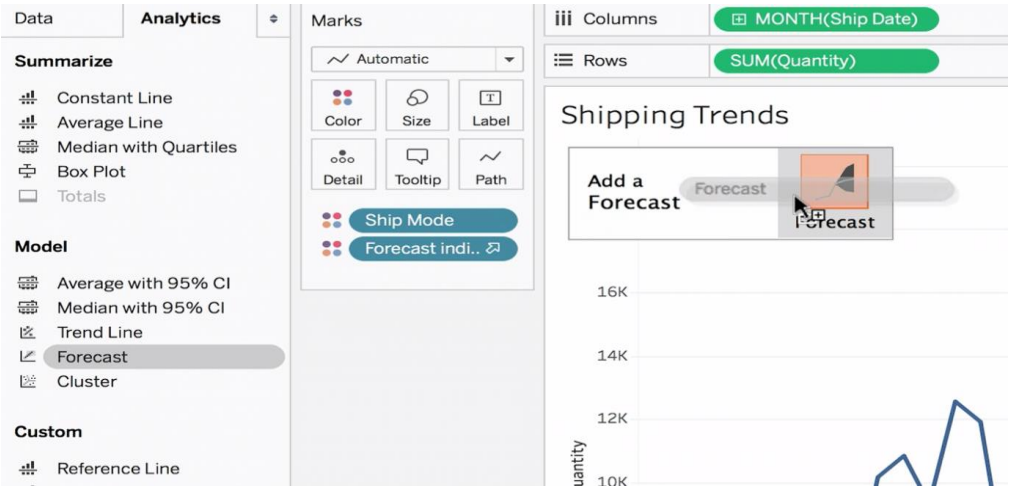
가설 검증을 목적으로 데이터를 수집하고 분석했던 기존의 통계학적 방법으로는 데이터가 지나치게 많아지게 되면 핵심 의미를 파악하기 어렵다. 이때 탐색적 데이터 분석은 가설을 정하고 데이터를 수집하는 것이 아닌 열린 상태로 데이터를 수집하고 시각화, 패턴 확인을 하기에 이러한 단점을 보완할 수 있다. 데이터의 분포 값을 검토함으로써 데이터가 말하고자 하는 바를 더욱 이해하기 쉬우며 잠재적인 문제를 발견한다면 해결 혹은 수정도 가능하다. 이에 따라 본격적인 데이터 수집과 분석에 들어가기 전에 데이터의 수집 범위 등을 정할 수 있으며 원하는 정도내에서 깊게 탐색할 수도 있다. 이러한 과정은 데이터를 다양한 시각에서 바라보게 하고 새로운 패턴이 도출되었을 때, 기존의 가설을 수정하거나 새로운 가설을 수립하는 등 능동적으로 대처하고 데이터를 분석해 나갈 수 있게 한다.

그럼 데이터 시각화는 왜 필요한 것일까. 데이터 시각화는 말 그대로 데이터를 분석한 결과를 한눈에 쉽게 이해할 수 있도록 시각적으로 나타내는 것을 말한다. 데이터 시각화 과정을 통해 숫자 데이터로만 봐서는 미처 발견하지 못했던 인사이트를 발견할 수 있으며, 내가 이해하고 발견한 데이터의 내용과 인사이트, 그리고 데이터를 통해 궁극적으로 말하고자 하는 바를 상대방에게 보다 쉽고 명확하게 전달하기 위한 방식으로 활용할 수 있다. 디지털 트랜스포메이션의 시대에서 수많은 데이터 속 숨겨진 인사이트를 발견하고, 데이터를 통해 말하고자 하는 바를 효과적으로 전달할 수 있게 하는 데이터 시각화 능력은 하나의 큰 무기이자 필수 역량이라고 할 수 있다.

이렇게 데이터를 시각화할 수 있는 다양한 시각화 솔루션이 존재한다. 그 중 비전공자들도 코딩 없이 다양한 시각화 차트를 단순한 드래그앤드롭(drag&drop) 방식과 간단한 조작을 통해 차트를 즉석에서 만들 수 있는 솔루션을 몇 가지 소개하고자 한다.

1) 태블로(tableau)

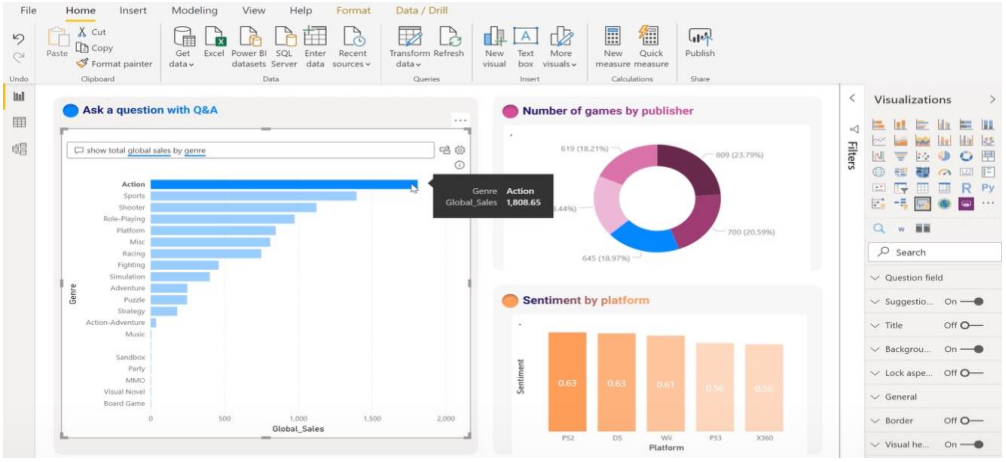
태블로는 데이터 사이언스를 비전문가에게도 보급하겠다는 비전을 제시하고 있는 태블라는 회사에서 제작한 Self-Service BI 솔루션이자, 데이터 시각화 툴이다. 태블로는 drag&drop을 활용한 독보적으로 직관적인 UI를 구축하고 있으며 다수의 무료 교육 동영상을 5개국어로 제공하여 빠른 학습이 가능하다는 특징이 있다. 태블로는 다른 BI 툴에 비해 '시각적 분석'이 용이하고, 차트 하나에 표현할 수 있는 시각화 옵션이 다양하고 자유도가 크기 때문에 시각적으로 표현해낼 수 있는 범위가 넓다. 거의 모든 데이터 원본에 대응이 가능하고, 비주얼적으로 뛰어난 시각화 대시보드를 만들 수 있다는 장점이 있다. 하지만 여러 데이터 원본에 걸친 데이터 분석은 하기 어렵다는 단점도 있다.



태블로 활용 예시(출처: 태블로 홈페이지)

2) 파워비아이(Power BI)

파워BI는 마이크로소프트사에서 출시한 비즈니스 분석 서비스로, 개발자나 전문가가 아니더라도 다양한 데이터를 가져와 구조화하고, 시각화하여 공유할 수 있도록 서비스를 제공하는 대표적인 BI 솔루션이다. 파워BI는 통합 ETL툴(데이터 추출, 변환, 적재가 모두 가능한)로써 다양한 데이터 원본에 연결하여 데이터를 가져올 수 있고, 가져온 데이터를 추출, 변형, 통합하는 등의 전처리 작업을 수행할 수 있다. 또한 이 과정에서 수작업을 최대한 없애고 로딩시간을 단축한다. 파워 BI는 엑셀과 연계하여 데이터를 그대로 이용이 가능하며 자연어 쿼리를 통해 순간 미리보기 검색이 가능하여 초고속 답변을 얻을 수 있다는 장점이 있다. 하지만 외부 데이터의 이용이 어렵다는 단점이 있다.



파워BI 활용 예시(출처: 파워BI 홈페이지)

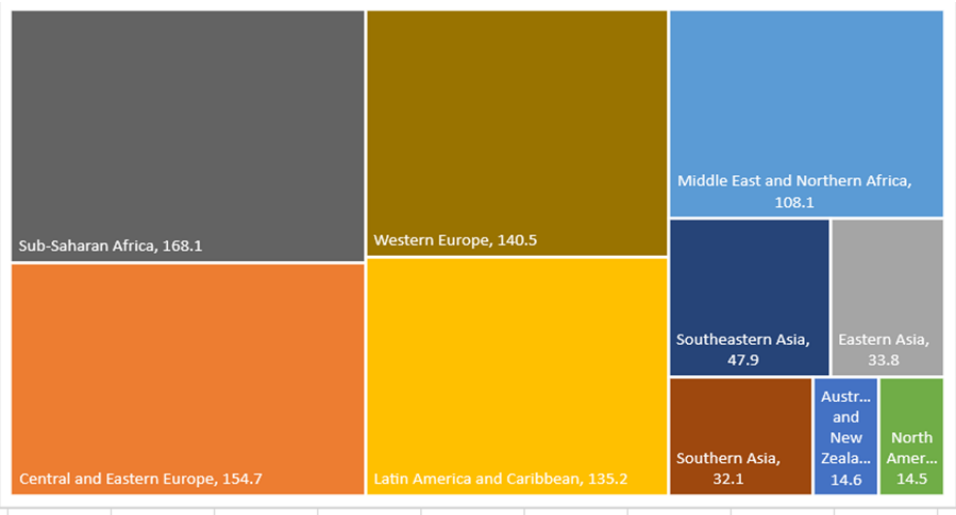
3) 시센스(Sisense)

시센스는 고도의 대화형 사용자 환경을 제공하는 분석 앱을 구축할 수 있는 분석 BI 플랫폼이다. 빅데이터 뿐만 아니라 이질적인 데이터들을 즉시 분석하고 그래프로 보여주는 BI 툴이다. 마우스 클릭과 드래그 몇 번으로 대용량의 빅데이터로도 쉽고 빠르게 복잡한 데이터 모델을 만들고 관리할 수 있으며 뛰어난 시각화 옵션으로 간단하고 즉각적인 분석이 가능하다는 장점이 있다. 하지만 비교적 그래프와 지도의 종류가 적고 공동작업에 적합하지 않다는 단점을 가지고 있다.

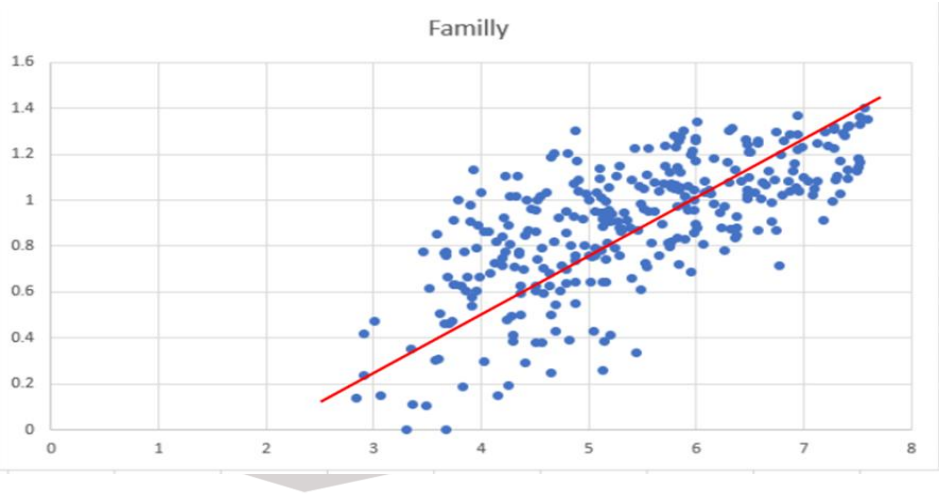
시센스 활용 예시(출처: 시센스 홈페이지)

4) 엑셀(Excel)

엑셀에 자체적으로 내재되어있는 차트 생성 기능을 활용하는 방법도 있다. 비전공자도 엑셀을 활용하면 탐색적 데이터 분석을 통해 꺾은선 차트, 막대형 차트, 분산 차트(산점도), 지도 차트, 파이 차트, 트리맵 차트 등 다양한 방법으로 쉽게 데이터 시각화가 가능하다.



엑셀로 그린 트리맵 차트 예시



엑셀로 그린 분산 차트(산점도) 예시

4. 마치며

올바른 질문에 올바른 답을 찾는다. 이것은 우리가 디자인씽킹을 '일하는 방식'으로 채택하는 궁극적인 이유이자 목표이다. 다만, 대부분의 사람들은 빠르게 '올바른 답'을 찾으려 하는데 집중할 뿐 '올바른 질문'을 찾는 일에는 소홀한 경우가 많다.

이를 테면 "판매가 잘 안된다"라는 문제를 가지고 "어떻게 하면 판매를 잘할 수 있을까?"라고 해결책을 찾는다면 무엇이 문제인지 명확하지 않으므로 해결책을 찾는 것도 어렵다. 그러므로 이러한 사례에서 우리가 접근해야 할 태도는 "잘 안 된다"는 것이 무엇인지를 명확하게 정의하고 이로부터 접근하는 것이어야 한다. 즉, 올바른 문제를 정의해야 한다. 그러나 막상 정의를 내리려고 하면 막막하고 어려울 수 있으므로 올바로 문제를 정의할 수 있도록 하는 '기준'과 '방법론'을 알아야할 필요가 있다.

다만 이 기준과 방법론이라는 것이 더 이상 전통의 디자인 방법론을 고수한 것일 필요는 없다. 우리는 최선을 다해서 '제대로' 기준과 방법론을 세우기 위한 가능한 모든 방안들을 고안해야 하며 이를 위한 데이터 활용은 더 이상 선택이 아닌 필수다. 가령 우리의 고객이 누구냐고 묻는다면 그 기준과 정의는 누가 정할 것인가? 현장에서 고객을 직접 만나는 영업사원도, 온오프라인을 통해 수많은 마케팅 메시지를 뿌리는 마케터도, 경험과 사업에 통달한 대표자도 오늘날처럼 빠르게 디지털 전환이 일어나는 시대에서는 더 이상 어떠한 하나가 완전한 정답이라고 확언할 수 없다. 그러니 가능한 객관적이면서

정성적이고 정량적인 많은 데이터를 모아서 데이터 기반의 의사결정을 해야 한다.

이를 위해 우리는 고객에 대한 정보를 어떠한 데이터로 수집할지, 모아진 데이터를 어떻게 해석하고 판단하여 사용할지를 끊임없이 고민해야 한다. 물론 데이터와 정량적 분석기법을 통해 고객의 문제를 이해하고 니즈를 발견하는 데이터 기반의 공감(Empathy)이라는 것은 쉽지 않을 것이다. 다만 전공지식의 여부와 상관없이 가능한 접근하기 쉽고 활용해볼 수 있는 여러 노코드 방식의 데이터 활용법을 통해 데이터와 친근해지고, 실제 필요한 데이터를 고민하는 훈련 등이 추가된다면 진정한 데이터로 공감하기를 시작하는 첫 걸음으로서는 손색이 없을 것이다.

바야흐로 고객의 욕망까지 데이터로 읽는 시대가 왔다. 누구도 데이터로부터 완전히 분리될 수도, 자유로울 수도 없다. 그렇다면 처음엔 다소 어렵고 막막해 보이더라도 우리는 선택의 갈림길에 서야 한다. 이것을 활용하여 나의 도구로 만들 것인지, 아님 손 놓고 흘려 보낼 것인지 말이다. 데이터 드리븐 디자인씽킹(Data Driven Design Thinking)은 이 갈림길에서 당신이 가장 손쉽게 데이터를 활용하고 혁신적 사고발상을 할 수 있도록 지원하는 훌륭한 촉매제가 될 것이다.

끝.

저자 소개

박진아

'비전공자를 위한 진짜 필요한 데이터 활용 교육'을 모토로 교육 서비스 스타트업을 운영하고 있는 창업가. 서강대 기술경영학 석사과정을 졸업했다. 데이터로 문제를 해결하고 의사결정 하는 실질적인 방법론과 프로세스를 연구하고 교육 콘텐츠로 개발하는 일을 한다. 특히 데이터 기반 문제해결과 관련한 레퍼런스로는 2019년 행안부 주관 국민디자인단을 통해 국내 지역의 폭염피해에 따른 가축폐사 문제를 컨설팅해 장관상을 수상하기도 했다. 세상 돌아가는 소식과 새로운 비즈니스 기회를 찾는 것, 그리고 일을 효율적으로 개선하는 것에 관심이 많다.

김경하

한양대학교에서 정치외교학과 사회혁신을 전공하는 전형적인 문과생이다. 비전문가를 위한 데이터 교육 전문 그룹 에이블런에서 마케터로 일하며 데이터를 어깨 넘어 공부했다. 디자인 씽킹과 데이터로 지역 사회 문제를 해결하는 데에 관심이 많다. 2021년에는 국민 중심의 예보평가 방법 마련을 위한 국민디자인단 프로젝트의 퍼실리테이터로 참여한 경험이 있다.

김도원

국민대학교에서 행정학을 전공했다. 이과 전공이었으나, 수학 과학보단 사회와 역사에 흥미를 느껴 문과로 전향했던 경험이 있다. 비전문가를 위한 데이터 교육 전문 그룹인 에이블런에서 교육지원 및 다양한 교육 콘텐츠 개발 업무를 하며 비전문가의 입장에서 데이터 활용법을 쉽게 풀어낸 경험들이 있다. 2021년에는 소상공인을 위한 디지털 사회혁신 프로젝트를 운영하여 데이터를 활용한 서울시 소상공인 매출 개선에 기여한 바 있다. 사회복지, 노인복지 등 다양한 복지분야 및 사회문제 해결에 관심을 두고 있으며 데이터를 활용한 사회혁신에 관심이 많고, 고양이를 키우는 한명의 집사로서 고양이에도 관심이 많다.

ⓒABLEARN, 2022.

이 책은 저작권법의 보호를 받는 저작물로 무단전재, 복제, 배포를 금합니다.

이를 위반 시 민사 및 형사상의 법적 책임을 질 수 있습니다.

ISBN 979-11-979450-1-4 | 02-589-2021 | edu@ablearn.kr | ablearn.kr