



Telling a Story with Data

Communicating effectively with analytics

BY THOMAS H. DAVENPORT > PHOTOGRAPHY BY DAVID CLUGSTON
> TYPE ILLUSTRATIONS BY ELIZABETH BADDELEY

애널리틱스와 데이터는 전 세계 주요 조직들의 의사결정 과정을 획기적으로 변화시키고 있다. 그러나 애널리틱스의 문제점 중 하나는 설명하고 이해하기 어렵다는 점이다. 분석가와 의사결정자 간에 의사소통이 잘 이뤄지지 않는다고 알려져 있을 정도다. 따라서 분석 역량이 효율적으로 사용되지 못하고 의사결정자들이 자신의 직관이나 경험에 의존하는 경우가 많다. 그러나 데이터와 전술을 가지고 이야기할 때 계량분석가(Quantitative Analyst)가 사용할 수 있는 다양한 접근법이 있다. 이는 의사결정자가 분석가들과 신뢰 있고 이로운 관계를 맺는 데 기여할 수 있다.

$$f(x) = \lambda^\alpha / \Gamma(\alpha) X^{\alpha-1} e^{-\lambda x}, x > 0. C(A \Rightarrow B) = T$$

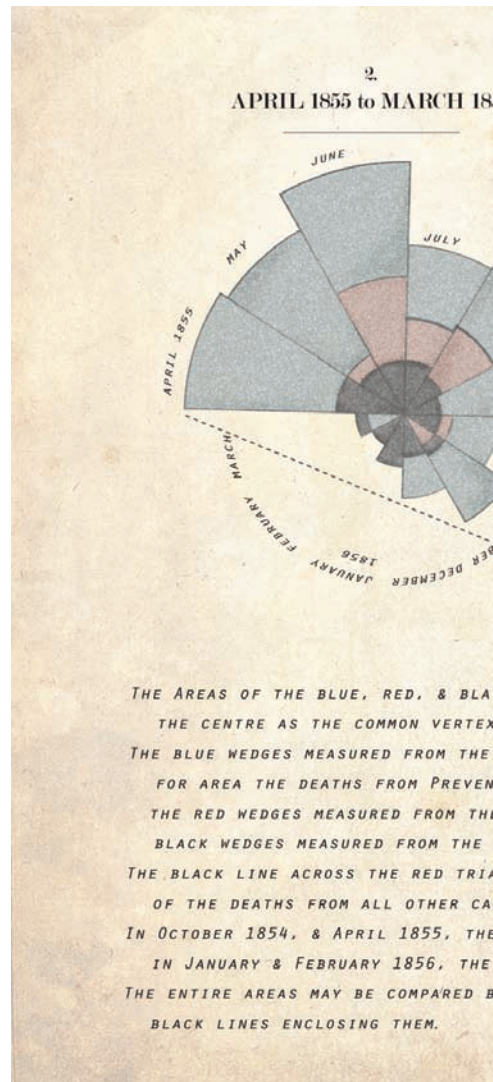
오늘날 조직들은 애널리틱스와 “빅데이터”에 크게 열광하고 있다. 그러나 의사결정자가 애널리틱스와 그 의미를 바로 이해해야만 비로소 행동을 바꿔 의사결정 시 분석적인 접근법을 적용할 것이다. 자신의 작업이 실행에 옮겨지는지- 의사결정을 변화시키고 행동에 영향을 미치는지-의 여부에 주의를 기울이는 계량분석가는 이 문제에 대해 많은 시간과 노력을 들인다. 반면, 이 문제에 주의를 기울이지 않는 분석가들은 결과가 “스스로 대변한다”고 생각하여 의사소통에 대해 크게 신경 쓰지 않는다. 이러한 태도는 현재에도 과거에도 일반적으로 효과적이지 못했다.

HISTORICAL EXAMPLES OF COMMUNICATING RESULTS, GOOD AND BAD

계량분석 결과를 효과적으로 제시하는 방법은 오래 전부터 사용되어 왔다.¹ Florence

Nightingale의 사례를 살펴보자. Nightingale은 간호사라는 직업을 창시하고 병원 위생관리 방법을 개혁한 사람으로 잘 알려져 있지만, 그녀는 계량기법의 초기 사용자이기도 하였다. 크림전쟁이 한창이던 1854년 10월, Nightingale이 38명의 자원 간호사들과 함께 터키에 있던 영국 국군병원에 파견되었을 때, 그녀는 간이병원의 끔찍한 위생상태를 목격하였다. 병원에서 사망하는 사람들 대부분은 전쟁에서 입은 상처 때문이 아니라 유행병, 전염병, 풍토병으로 사망하였다. 1855년 2월 기준, 병원에서 치료받은 사람들의 사망률은 43%였다.² Nightingale은 병원의 기본 위생관리를 개선하는 것 외에도 통계학을 이용하여 여러 문제를 해결할 수 있다고 생각하였다. 그녀는 입원, 부상, 질병, 치료, 사망에 대해 매일 세세하게 기록하기 시작하였다.

그러나 Nightingale이 이룩한 가장 큰 혁신은 결과를 제시하는 방법에 있었다. 그녀는 수치에 기반한 증거자료의 중요성을 잘 인지하고 있었지만, 모든 사람이 수치표에 관심을 가지고 있지 않다는 점(심지어 그 당시에는 수치표가 흔하지 않았다!)과 일반 독자들이 표를 읽지 않아 증거를 놓칠 수도 있다는 사실 또한 이해하고 있었다.

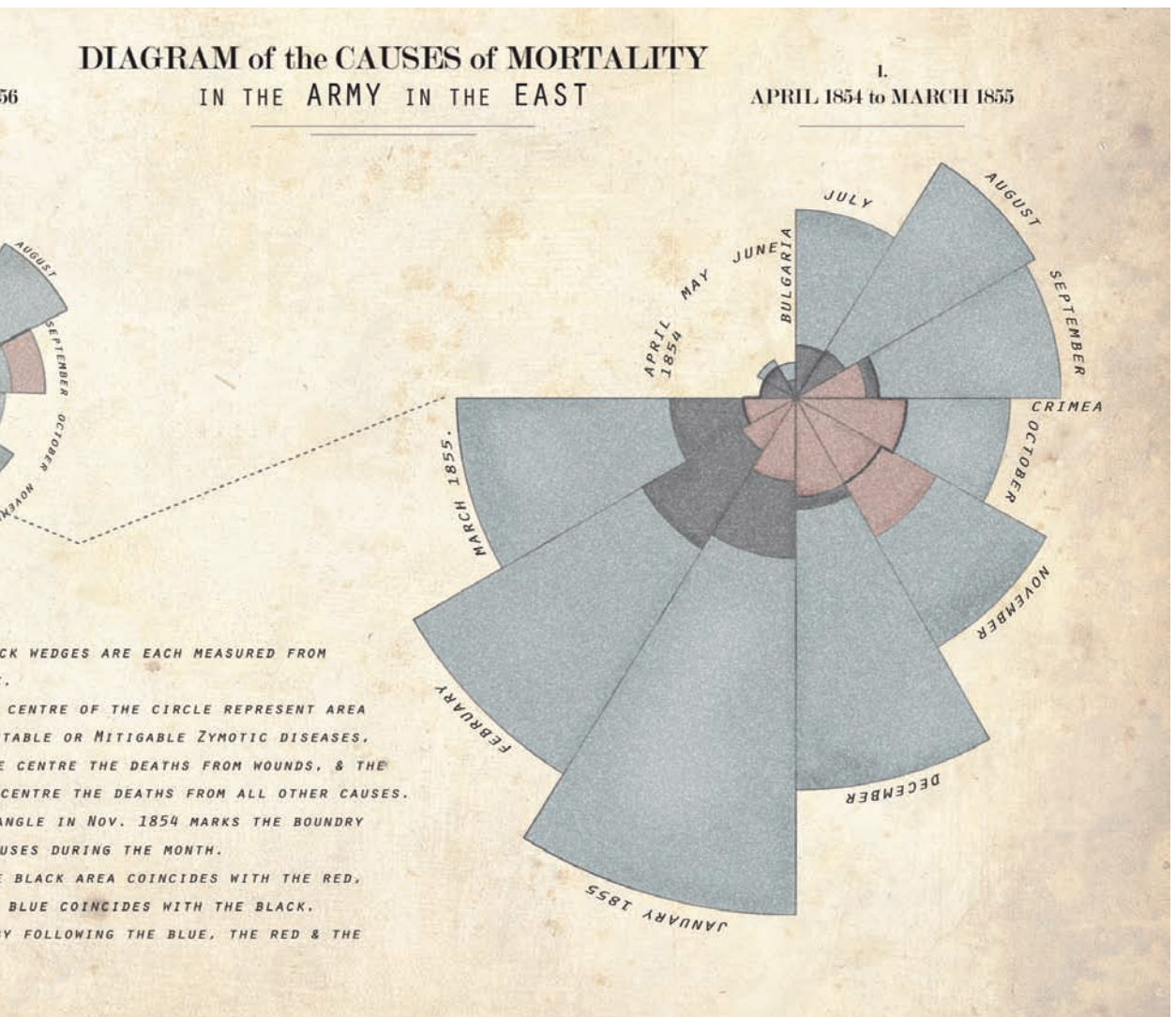


(A⇒B)/T(A), PP(XG=k)=PP(X)PP(G=k|X),R

그녀는 독자들이 통계로 제시하는 메시지를 놓치지 않기를 원했기 때문에 다이어그램을 이용하여 비위생적인 환경으로 인한 불필요한 죽음과 개혁의 필요성을 극적으로 표현하였다. 오늘날에는 데이터를 제시할 때 흔히 쓰이는 기법이지만 그 당시에는 상당히 새로운 방법이었다.

혁신적이었던 그녀의 다이어그램은 썰기 모양의 파이차트(Pie Chart)였다. Nightingale은 다양한 색을 사용하여 원인별 사망률이 매달 어떻게 변하는지 효과적으로 보여주었다. 수치와 다이어그램으로 제시된 증거는 반론의 여지가 없을 정도로 분명하게 표현되었다.

Nightingale은 수치를 읽거나 이해할 수 있는 능력이 낮은 의원들을 대상으로 이 창의적인 다이어그램을 이용하여 국군병원 의료서비스의 실태와 심각성에 대한 보고서를 발표하였다. 사람들은 부상병들이 병원에서 치료받는 대신 죽어가고 있다는 말을 듣고 충격을 받았다.



$$SS(f, \lambda) = \sum \{y_i f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt, EIP^{(\infty)} = EI($$

Nightingale이 체계적으로 수집한 데이터에서 볼 수 있듯이 결과적으로는 사망률이 급락하였다. 크림전쟁이 끝난 1856년 6월, 영국에 돌아온 그녀는 자신이 유명인이 되어 영웅으로 칭송 받고 있다는 것을 알게 되었다. 그녀는 1859년에 영국 왕립통계학회의 회원이 되었고—최초의 여성 회원이었다—1874년에 미국 통계협회의 명예회원이 되었다.³

결과를 잘 전달하지 못한 사례—그리고 이 주제가 얼마나 중요한지 상기시켜주는 사례—로는 Gregor Mendel을 들 수 있다.⁴ 유전(遺傳)학의 아버지인 Mendel은 1884년 사망하기 몇 달 전에 “과학 연구를 하면서 큰 만족을 느꼈습니다. 그리고 곧 전 세계가 제 업적을 인정하게 될 것이라고 믿습니다.”라고 말하였다. 그가 자신의 연구결과를 잘 전달하였다면, 어쩌면 그가 살아있는 동안에 그의 업적이 좀더 빨리 인정받았을지도 모른다.

성직자였던 Mendel은 완두의 특정 형질의 유전에 대해 연구하였다. 1856년과 1863년 사이에 Mendel은 유전법칙을 증명하기 위해 수만 개의 완두를 교배하고 기록하였다. 그는 한 세대에서 다른 세대로 유전정보가 전해질 때 일정한 법칙을 따른다는 것을 밝혀내었고 이 법칙은 추후에 그의 이름을 따서 명명되었다. Mendel의 업적은 20세기 초에야 비로소 인정받았다. 추후에 이 법칙들을 재발견하면서 현대 유전학의 기반이 마련되었다.

Mendel이 연구결과를 전달하는 방법이 그의 실험만큼 효과적이었다면 얼마나 좋았을까? 그는 연구결과를 잘 알려지지 않은 모라비아의 한 과학저널에 발표하였다. 이 저널은 유럽과 그 외 지역에 있는 130개의 과학기관에 배포되었지만 당시 큰 반향을 일으키지 못했고, 그 후 그의 논문은 35년간 세 번밖에 인용되지 않았다.⁵ 심지어 같은 학문 분야의 영향력 있는 사람들마저도 복잡하고 상세한 그의 논문을 이해하지 못하였다. Mendel의 직업이 성직자가 아닌 과학자였다면 자신의 연구결과를 널리 알릴 수 있었을 것이고 이를 외국에서 발표했을지도 모른다. Mendel은 Darwin 등의 사람들에게 자신의 논문을 보내면서 타국의 과학자들과 접촉을 시도하였다(그가 보낸 논문 중 극소수만이 발견되었다). 알려진 바에 따르면 Darwin은 Mendel의 논문을 펴보지도 않았다고 한다.

Mendel은 그의 발견이 역사를 얼마나 변화시킬지 알지 못한 채 사망하였다. Mendel의 논문은 당시의 기준으로는 기발하고 전혀 없는 것이었지만 과학자들은 30년 후에야 이 논문을 이해하기 시작하였다. 자신의 계량분석 결과가 그렇게 오랜 시간 동안 묻혀지기를 원하지 않는다면, 그 결과를 어떻게 전달할지에 대해서 상당한 관심을 기울여야만 할 것이다.

$$y - f(x_0))^2 / [X = X_0] = \sigma_{\varepsilon}^2 + [E\hat{f}(x_0) - f(x_0)]^2 + E[\hat{f}(x_0)$$

TWO RECENT EXAMPLES

좀더 최근에는, 애널리틱스를 사용하는 조직 중에 분석의 질 뿐만 아니라 이를 전달하고 채택시키는 데에도 노력을 기울인 주목할만한 사례가 있다. 비즈니스 분야에서 분석 결과가 효과적으로 전달되고 널리 채택된 예로는 FICO 신용점수의 사례를 들 수 있다. 이 점수는 FICO라는 기업이 부여하는 300에서 850 사이의 점수인데, 특정 시기의 한 개인의 재정상태를 한 눈에 확인할 수 있다.⁶ 대출기관은 신용카드 발급, 자동차 대출, 모기지 등 개인과의 신용거래 시 감수해야 할 리스크가 무엇인지를 알고 싶어한다. 많은 대출기관에서 FICO 신용점수를 이용하여 채무자의 신용리스크를 결정한다. FICO 점수를 이용하여 대출기관은 특정 시기에 얼마를 어떤 조건(이자율 등)으로 대출할 지를 결정한다. 이는 애널리틱스를 실행에 옮긴 성공적인 사례이다. 미국의 거의 모든 대출기관에서 이를 활용하고 있고 그 외 지역에서도 이용자가 증가하고 있다.

FICO 점수는 리스크평가를 개선시켜 미국 신용시장의 효율성을 극적으로 향상시켰다. 이 점수를 통해 대출기관은 전적으로 채무자의 대출 이력에 근거하여 대출이 상환될 수 있을 지의 여부를 파악할 수 있게 되었다. 신용거래와 무관한 기업(자동차 보험회사, 통신사, 임대주, 금융서비스 기업의 자회사 등)들이 FICO 점수를 검토하고 이 정보를 이용해 특정 고객과 거래할 것인지 여부를 결정하며 소비자 등급별로 다른 요율을 책정하는 경우도 증가하고 있다.

FICO 점수는 Bill Fair라는 공학자와 Earl Isaac이라는 수학자가 1956년에 FICO(당시에는 Fair, Isaac으로 알려져 있었다)를 창립하면서 만든 것이다.⁷ 처음에 두 사람은 각각 400달러를 투자하여 신용평가 모델을 만들기 시작하였다. 그들은 1958년에 미국의 주요 50개 대출기관에 편지를 보내 신용평가라는 개념에 대해 설명할 기회를 달라고 요청하였다. 그들은 단 한 곳으로부터 답장을 받았다. 그러나 그들은 은행, 신용카드사와 결제대행업체, 보험회사, 유통업체, 신용조사기관 등의 잠재 고객에게 이 개념과 모델을 설명하기 위해 노력하였다. 그들은 FICO 점수를 의사결정 과정에 포함시킬 수 있도록 지원하는 컨설팅 서비스와 소프트웨어를 개발하였다. 21 세기가 시작될 무렵 그들은 전 세계로 사업영역을 확장하였고 고객들에게 신용점수를 열람할 수 있도록 하는 동시에 이를 개선시킬 수 있는 맞춤형 자문서비스를 제공하기 시작하였다.

분석연구는 제대로만 전달된다면 비즈니스 분야에서뿐만 아니라 학계에서도 큰 반향을 일으킬 수 있다. 예를 들어, 부부들의 결혼생활 유지 가능성을 예측하는 분석모델에 대해 들어본 적이 있을 것이다.

$$\sigma^2 = \sigma_{\epsilon}^2 + \text{Bias}^2(\hat{f}(x_0)) + \text{Var}(\hat{f}(x)) + \text{IP}(\hat{f}(x)) + \text{fx}(x) \cdot \sqrt{2\pi}$$

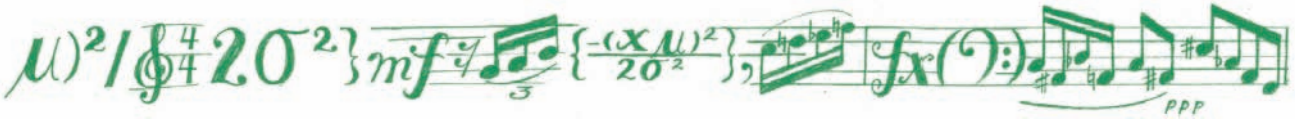
Washington 대학의 심리학교수 John Gottman은 이에 대해 Oxford 대학의 응용수학자 James Murray와 공동으로 연구하였다. Gottman은 가설과 데이터—여러 쌍의 부부에 대한 영상기록과 관찰기록을 수집하였다—를 제공하였고 결혼생활을 성공적으로 만드는 요소가 무엇인지에 대해 오랜 기간 동안 관심을 기울였다. Murray는 비선형 모델에 대한 전문적인 지식을 제공하였다. 이를 통해 나온 모델은 관찰된 행위에 근거하여 부부에게 점수를 부여하는데, 결혼의 미래 성공률을 94% 정확도⁸로 예측할 수 있었다.

이 모델은 Gottman, Murray 등이 저술한 'The Mathematics of Marriage: Dynamic Nonlinear Models'라는 저서를 통해 공개되었다. 원래 이 저서의 주요 목표 독자는 동료 학자들이었다. 그러나 많은 학자들과는 달리 Gottman은 연구대상이었던 사람들의 행위에 영향을 끼치고 싶었다. 그는 이에 관한 여러 저서와 논문을 출간하였고, (자신의 부인과 함께) The Gottman Relationship Institute(www.gottman.com)를 창설하였다. 이 연구소는 부부들과 치료사들을 위해 교육 클래스, 관계개선을 위한 비디오, 그리고 기타 여러 의사소통 수단을 제공한다. 또한 이 모델을 이용하여 연구자들이 다양한 상황 하에서의 부부들의 반응에 대해 모의 실험을 할 수 있게 되었다. 이처럼 모델링을 통해 “만약에”라는 사고(思考) 실험을 할 수 있게 되었고, 이러한 사고 실험은 위기에 놓인 부부들을 위한 새로운 과학적 개입 전략을 개발하는 데 도움을 줄 수 있었다. Gottman은 부부들을 대상으로 역사상 최대 규모의 무작위 임상실험을 수행하였는데, 여기에는 만 쌍이 넘는 부부들이 참여하였다. 그는 연구가 실제 부부들에게 얼마나 도움이 되었는지에 대해 다음과 같이 말하였다:

지난 8년간 나는 뛰어난 내 아내와 함께 이러한 아이디어들을 종합하고 우리의 이론을 이용하여 부부들과 그 자녀들에게 도움이 되고자 하였다. 우리는 이러한 개입들이 큰 변화를 만들어낸다는 사실을 알게 되었다. 우리는 위기에 처한 부부들에게 이틀간의 워크샵과 9개의 결혼상담 세션을 제공하였고, 이들 중 75%가 관계를 회복하였다.⁹

이는 효과적인 커뮤니케이션과 행동의 사례이다.

이러한 결과에도 불구하고 애널리틱스에 대한 전달은 전통적으로 계량분석가의 교육과 밀접한 관계가 있는 주제로 간주되지 않았다. 대부분의 학자들, 특히 연구와 강의가 분석에만 크게 치우쳐있는 학자들은 분석방법에 큰 관심을 보여왔지만 효과적인 전달방법에 대해서는 큰 관심을 기울이지 않았다. 다행히도 이 같은 상황은 변하기 시작하였다.



Harvard Statistics Department의 학장이자 최근 Harvard의 Graduate School of Arts and Sciences의 학장으로 임명된 Xiao-Li Meng 은 그의 목표가 “통계를 효과적으로 전달하는 사람”을 양성하는 것이라고 밝혔다. 또한 그는 다음과 같이 말했다:

흥미롭게도, 와인을 만드는 방법을 알지 못해도 와인감정가가 될 수 있다는 철학에 근거하여 우리는 고급 와인을 만드는 데에만 초점을 맞췄을 때보다 더 많은 차세대 와인 제조업자를 양성해 냈다.¹⁰

당신이 분석가이든 애널리틱스의 소비자이든(다른 표현을 사용하자면 분석적인 와인 제조업자이든 와인 소비자이든) 커뮤니케이션은 중요한 주제이다. 물론 분석가들은 더 많은 행동을 유도하기 위해 연구결과물을 더욱 흥미롭고 주의를 끌 수 있게 만들 수 있다. 애널리틱스의 소비자—예를 들어 분석 프로젝트를 의뢰한 매니저—는 결과물을 흥미롭고 이해가 용이한 형태로 제공할 것을 주장해야 한다. 계량분석을 읽는 독자들이 지루함을 느끼거나 혼란스러워 한다면 아마도 그것은 그들의 잘못이 아닐 것이다. 애널리틱스의 소비자들은 계량분석가들과 협력하여 결과물을 쉽게 이해하고 사용할 수 있도록 만들 수 있을 것이다. 물론 결정을 내리고 결과물에 따라 행동을 취하는 쪽은 대체로 애널리틱스의 소비자들이다.

COMMUNICATING THE BASICS OF AN ANALYSIS

분석한 결과물을 전달할 때의 핵심은 문제와 그 배경, 모델, 사용한 데이터, 분석에 사용된 변수간의 관계를 설명하는 것이다. 변수간의 관계가 확인되면, 그 관계의 의미가 해석되고, 언급되고, 문제와 연관하여 제시되어야 한다. 결과물을 명확하게 제시할수록 계량분석에 기반하여 결정하고 행동할 가능성이 높아진다. 이것이 결국 분석의 궁극적인 목표인 것이다.

결과물을 제시할 때에는 연구과정의 개략적인 설명, 결과의 요약과 결과가 의미하는 바에 대한 설명, 행동을 위한 제언—이 순서대로 이뤄질 필요는 없지만—이 포함되어야 한다. 대개의 경우 요약과 제언으로 시작하고 결과물은 관련자들과의 회의에서 전달하거나 공식적인 문서로 전달하는 것이 좋다.

단순히 흑백의 수치표와 수식으로만 데이터를 제시한다면 연구결과가 무시될 가능성이 높아진다. 기본적인 보고서는 대부분 단순한 그래픽으로 표현될 수 있다: 막대그래프, 파이차트 또는 보다 시각적인 효과가 뛰어난 인터랙티브 디스플레이를 이용하는 경우가 많다. 일부에서는 시각적인 프리젠테이션보다 수치를 통해 데이터를 제시하는 것을 선호하지만 그러한 사람들은 그리 많지 않다.



TELLING A STORY WITH DATA

좀 더 효과적인 분석가들 중에는 데이터로 이야기를 만들 수 있는 사람들이 있다. 분석방법의 세부사항이나 그 수단에 상관없이, 분석결과를 가지고 만든 훌륭한 이야기들에는 공통적인 요소가 있다. 이러한 이야기들은 서술적 묘사가 뛰어나며, 대부분 비즈니스 문제나 목표로 이야기를 이끌어 나가는 경우가 많다. 고객충성도에 관한 이야기는 다음과 같이 시작할 수 있다: “모두 아시다시피, 오랫동안 우리는 충성도가 높은 고객들을 파악하기를 원했고 그들의 충성도를 더욱 높일 수 있는 방법을 발견하고자 하였습니다. 이제 우리는 그 방법을 찾아냈습니다.”

훌륭한 이야기는 청중들이 이해할 수 있는 방식으로 연구결과를 제시한다. 청중들이 계량적 연구와 전문 용어에 익숙하다면 통계학이나 수학 용어—심지어 수식도 가능할 것이다—가 사용될 수 있다. 그러나 대개의 경우 청중들은 전문 용어에 익숙하지 않기 때문에 연구결과는 청중들이 이해하고 공감할 수 있는 개념을 이용하여 제시되어야 한다. 비즈니스 분야에서는 추가 이익이나 절감할 수 있는 비용, 또는 투자 수익률로 설명되는 경우가 많다.

훌륭한 이야기는 취할 수 있는 행동과 이것이 가져올 결과를 결론으로 제시한다. 물론 분석가들은 사전에 주요 이해관계자들과 다양한 액션 시나리오에 대해 논의해야 할 것이다. 계량분석가가 “이렇게 하고 저렇게 해야 합니다.”라고 말하는 것을 좋아할 사람은 없다.

Pixar의 스토리 아티스트였던 Emma Coats는 Twitter를 통해 스토리텔링의 22가지 원칙을 제시하였다.¹¹ 22가지 원칙이 모두 애널리틱스에 직접 적용될 수는 없지만 다음의 4가지 원칙은 적용될 수 있을 것이다.

- “저자(또는 계량분석가)의 입장이 아닌 청중의 입장에서 흥미를 가질만한 것을 염두에 두어라. 그들은 매우 다를 수 있다.”
- “본론보다 결론을 먼저 결정하라. 결론이 어렵기 때문에 먼저 정해야 한다.”
- “일단 작성해야 수정을 시작할 수 있다. 완벽한 아이디어가 머리 속에만 있다면 다른 사람과 공유할 수 없다.”
- “이야기의 핵심이 무엇인가? 가장 경제적으로 전달하는 방법은? 이를 안다면 여기서부터 시작하면 된다.”



이해관계자와 의사소통 시 사용할 수 있는 체계를 갖추는 것도 도움이 된다. 이를 통해 분석가와 의사결정자가 무엇을 해야 하는지가 분명해진다. 예를 들어 George Roumeliotis는 Intuit의 데이터과학 그룹을 책임지고 있는데, 이 그룹은 Intuit가 수집한 엄청난 규모의 온라인 데이터에 기반하여 제품 특성을 분석하고 개발한다. 내부고객을 상대하는 프로젝트에서 Roumeliotis는 다음과 같은 분석방법과 결과 전달방법을 추천한다. 대부분의 단계는 강한 비즈니스 지향성을 가지고 있다.

1. 비즈니스 문제에 대한 나의 이해
2. 비즈니스에 미치는 영향을 어떻게 측정할 것인가?
3. 사용할 수 있는 데이터는 무엇인가?
4. 해결책에 대한 초기 가설
5. 해결책
6. 해결책이 비즈니스에 미치는 영향

이 방법을 사용하는 데이터 과학자들은 위키(Wiki)를 만들어 각 단계의 결과를 게시하도록 장려된다. 클라이언트는 위키 상의 내용을 검토하고 그에 대한 의견을 제시할 수 있다. Roumeliotis는 위키가 분석결과를 검토하는 데 사용되는 온라인 도구지만 데이터 과학자들과 클라이언트 사이의 직접적인 커뮤니케이션을 유도하기도 한다고 말한다.

WHAT NOT TO COMMUNICATE

분석가들이 기술 용어에 익숙하기 때문에—사용한 통계학적 방법을 제시하고 실제 회귀계수를 구체적으로 설명하고 R^2 값(사용한 회귀모델을 통해 설명된 데이터의 분산합의 비율)을 밝히는 등—청중들도 그러할 것이라고 생각한다. 그러나 대부분의 청중들은 고도로 기술적인 프리젠테이션이나 보고서를 이해하지 못한다. 어느 여행업체 분석가가 지적했듯이 “아무도 R^2 값에 관심이 없다.”

분석가들은 그들이 분석한 순서대로 결과를 제시하려는 경향이 있다. “우리는 우선 이상치(Outlier)를 제거하였고 그리고 나서 로그변환을 하였습니다. 이를 통해 높은 자기상관관계가 도출되었고 우리는 1년의 시차를 둔 변수(One-year Lag Variable)를 만들었습니다.”

대략 어떤 상황인지 그림이 그려질 것이다. 다시 한번 말하지만 분석결과를 전달받는 청중들은 분석가들의 분석 절차에 대해 궁금해하지 않는다. 그들은 결과와 결과가 전달하는 바에만 관심을 갖는다. 이러한 정보를 보고서나 프리젠테이션의 부록에 수록하면 도움이 되겠지만, 데이터를 가지고 훌륭한 이야기를 할 때에는 이러한 정보들이 방해 요인이 되어서는 안 될 것이다. 청중들이 정말로 원하는 것부터 이야기를 시작해라.

MODERN METHODS OF COMMUNICATING RESULTS

오늘날에는 분석결과를 전달하는데 사용할 수 있는 새로운 도구들이 존재하며, 모든 분석가들은 그 가능성을 인식하고 있어야 한다. 물론 적절한 전달도구는 상황과 청중에 따라 달라질 것이고, 화려함만을 추구하기 위해서 거창한 데이터 시각화 효과를 사용해서는 안 될 것이다.

비주얼 애널리틱스(Visual analytics, 다른 말로 데이터 시각화)는 최근 몇 년 동안 급성장하였다. 막대그래프와 파이차트는 시각적 디스플레이로 표현할 수 있는 것의 극히 일부에 불과하다.

다른 표현 방법으로는 산점도, 행렬, 히트맵(Heat Map), 선 그래프, 버블차트, 트리 맵 등의 다양한 방법이 있다. 어떤 목적으로 어떤 종류의 도표를 사용할지 결정하는 것이 어려워 보일 수 있는데, 일부의 경우 비주얼 애널리틱스 소프트웨어를 이용하면 데이터 변수에 근거하여 적절한 도표를 추천해 준다. 예를 들면 SAS Visual Analytics의 “Autochart”는 이용자들의 결정을 도와주는 도구이다. 이 프로그램은 데이터에 “일자/시간 카테고리 하나 있다면 다른 카테고리가 몇 개이든지 상관없이” 자동적으로 선 그래프를 만든다.¹²

많은 사람들이 정적인 차트를 이용하지만, 비주얼 애널리틱스는 점점 역동적이고 인터랙티브해지고 있다. 스웨덴 교수 Hans Rosling은 그의 유명한 TED Talk¹³을 통해 이 접근법을 유행시켰다. 그는 비주얼 애널리틱스를 이용하여 시간에 따라 변하는 선진국과 개도국 간의 인구 건강 관계를 제시하였다. Rosling은 Gapminder(www.gapminder.org)라는 웹사이트를 만들었는데, 이 사이트는 다양한 인터랙티브 비주얼 애널리틱스를 보여준다. 이는 시간에 따른 데이터의 추이를 제시할 때 더욱 많이 사용될 것으로 보이지만,



이 방법이 모든 종류의 데이터와 분석에 적합한 것은 아니다.

때로는 그래픽보다 결과물을 통해 이해를 도울 수도 있다. 예를 들어, 여러 대기업에서 시장조사 연구자와 전략가로 활동하는 Vince Barabba는 시장조사 결과를 잘 전달하는 방법에 대해 연구할 때 창의적인 사고를 적용하였다. 한 자동차 회사에서 시장조사 부서를 이끌었을 당시 그는 임원들이 3D 모형으로 자동차의 시장 잠재성을 검토하는 데 익숙하다는 것을 알게 되었다. 그는 중요한 시장조사 결과를 3D 모형으로 만들어 임원들이 이를 직접 보고 만질 수 있도록 하였다. 시장수요의 “급등”을 보면서 만지는 경험이야말로 그들에게 새로운 의미를 부여하게 되었다.

Intercontinental Hotels Group(IHG)에는 여러 애널리틱스 그룹이 있다. David Schmitt는 재무조직 소속의 전략기획 그룹을 이끌고 있다. Schmitt의 그룹에서는 IHG의 실적에 대해 보고서 형식으로 이야기를 한다. 그들은 “데이터를 가지고 스토리를 만드는 것”, 분석 결과에 관심을 유도하고 그 결과에 따른 행동을 끌어내기 위한 도구의 활용에 집중하고 있다. 그들은 이를 위해 청중에 따라 구사할 수 있는 다수의 테크닉을 보유하고 있다. 그 중 하나는 “뮤직 비디오”를 사용하는 것이다. 5분 길이의 뮤직비디오는 이미지, 오디오, 비디오를 이용하여 도출한 결과 뒤에 숨어있는 광범위한 개념을 모두 다룬다. 그리고 나서 개념 뒤에 숨어 있는 의미를 효과적으로 전달하기 위해 구두 프리젠테이션으로 논거를 제시한다.

최근 Schmitt의 그룹은 여름철 수요를 예측하는 비디오를 제작하였다. “Summer Road Trip”이라고 불리는 이 비디오에는 길을 따라 내려가는 자동차가 등장한다. 비디오에서 자동차는 “전방에 높은 수요가 예상됨”이라는 표지판과 시장 통계가 적힌 광고판을 지나치게 된다. 이 비디오의 목표는 다가올 여름에 실적의 주요 동력이 무엇일지, 그리고 이를 어떻게 각 지역과 연계할지에 대해 직원들이 생각하도록 만드는 것이었다. Schmitt 말에 따르면 “데이터가 핵심이 아니다: 수치가 중요한 것이 아니다—중요한 것은 아이디어이다.” 기본 아이디어가 전달되었다면 Schmitt는 데이터를 더욱 심도있게 다루기 위해서는 전통적인 프리젠테이션 기법을 사용할 것이다. 그러나 그는 우선 비디오를 통해 청중들이 마음의 준비가 되어 있기를 원하였다.

게임은 분석 결과와 모델을 전달하는 또 다른 방법이다. 게임을 통해 복잡한 관계에 있는 변수들이 어떻게 상호작용하는지를 전달할 수 있다. 예를 들어, 맥주 회사의 유통과정을 시뮬레이션한 “Beer Game”은 MIT가 1960년대에 개발했는데, 수천 개의 기업과 대학이 “채찍효과(Bullwhip Effect)”와 같은 공급망 관리 모델과 원칙을 교육하는 데 사용하였다. 기업들은 특정 목표를 전달하기 위해 자체적으로 게임을 개발하기 시작하였다. Schneider National이라는 화물운송회사는 화물차와 트레일러를 배차하는 데 분석적인 사고가 얼마나 중요한지 설명하기 위해 시뮬레이션 게임을 개발하였다.

게임의 목표는 주어진 매출액에서 변동비를 줄이고 운전사의 유휴시간을 감소시키는 것이었다. 화물을 받을 것인지, 또는 화물차를 빈 채로 이동시켜야 할지에 대한 결정은 플레이어가 내리는데, 이 때 의사결정을 돕는 도구를 이용한다. Schneider는 게임을 이용하여 직원들이 의사결정을 돕는 분석적인 도구들의 가치를 이해하도록 하고, 비즈니스의 역학관계를 설명하고, 직원들이 스스로를 “주문접수자”가 아닌 “이익창출자”로 생각하도록 하였다. Schneider의 일부 고객도 이 게임을 활용하였다.

기업들은 현대 기술을 이용하여 의사결정자들이 데이터와 직접 상호작용할 수 있도록 할 수 있다. 예를 들어 Deloitte Consulting LLP는 공항 운영에 관한 질문과 보고를 할 수 있는 iPad 기반의 시스템을 만들었다. 이 시스템은 Google Maps를 이용하여 특정 항공편이 어느 공항으로 가는지 지도 상에 보여준다. 비행기의 색(나쁘면 빨간색, 좋으면 초록색)은 특정 공항의 실적이 좋은지 나쁜지를 나타낸다. 지도 상에서 특정 공항의 상자를 터치하면 그 공항의 재정 및 운영 데이터를 볼 수 있다. 다양한 버튼을 터치하여 직원, 고객서비스 수준, 재정상태, 운영, 문제 영역 등을 확인할 수 있다. 이 “앱”은 오늘날의 인터랙티브하고 사용자 친화적인 기술로 구현할 수 있는 것 중의 한 가지 사례라 할 수 있다.

BEYOND THE REPORT

프리젠테이션이나 보고서가 분석 프로젝트의 유일한 결과물은 아니다. 분석가들이 가치를 창출하는 결과물을 도출한다면 더할 나위 없이 좋을 것이다. 예를 들어, 오늘날에는 자동화된 의사결정 환경 안에 애널리틱스를 포함시키는 기업들이 증가하고 있다.¹⁴ 보험, 금융, 소비자 지향 가격 결정 환경(항공사나 호텔 등)에서는 분석 시스템에 기반한 자동화된 의사결정을 흔하게 볼 수 있다. 이러한 환경에서는 선택의 여지 없이(선택의 여지가 아주 조금 있을지도 모른다. 사람들은 때로 예외적인 사례를 검토하기도 한다) 애널리틱스가 사용될 것이다. 계량 분석가나 중요한 의사결정을 해야 하는 기업 대표라면, 이러한 시스템을 개발하고 실행하는 것이 보고서를 만드는 것보다 훨씬 더 효과적일 뿐만 아니라 번덕스러운 분석 결과를 전달하는 수고도 생략할 수 있다.

온라인 정보산업의 기업들은 페타바이트급 정보가 포함된 “빅데이터”를 보유하고 있다. 엄청난 양의 새로운 정보가 급속도로 밀려 들어오기 때문에 모든 정보를 이해하는 것은 어렵다. 이러한 환경에서, 이러한 기업에서 일하는 “데이터 과학자” (데이터 관리와 프로그래밍 기술이 뛰어난 계량분석가)들은 보통 상품개발 조직에 소속되어 있다.¹⁵

그들의 목표는 보고나 프리젠테이션을 하는 것이 아니라 상품 원형과 새로운 상품의 기능을 개발하는 것이다. 예를 들어 비즈니스 네트워크 웹사이트인 LinkedIn의 데이터과학 그룹은

상품 조직 산하에 있는데, 소셜네트워크와 일자리 간의 관계에 기반하여 새로운 제품 특성과 기능을 개발하였다. 그 예로 People You May Know, Talent Match, Jobs You May Be Interested In, InMaps Visual Network Displays, Groups You Might Like를 들 수 있다. 이 중 일부(특히 People You May Know)는 LinkedIn의 고객층 성장과 유지에 엄청난 영향을 끼치고 있다.

목표가 의사결정자가 사용하는 접근법을 바꾸는 것이든 상품이나 과정을 개선하는 것이든, 성공하기 위해서는 커뮤니케이션이 중요하다. 분석의 초기 단계부터 분석가는 결과를 어떻게 전달할 것인지를 고민해야 한다. 훌륭한 분석결과 전달자는 분석이 끝날 때까지 기다리지 않고 전체 과정을 이해관계자와 소통하는 수단으로 이용한다. **DR**

Thomas H. Davenport is a visiting professor at Harvard Business School, professor at Babson College, and a senior advisor to Deloitte Analytics, Deloitte Touche Tohmatsu Limited.

Endnotes

1. I am grateful to Jinho Kim for these historical examples.
2. I.B. Cohen, *The Triumph of Numbers: How Counting Shaped Modern Life* (W.W. Norton, 2006), chapter 9.
3. Wikipedia, "Florence Nightingale," <http://en.wikipedia.org/wiki/Florence_Nightingale>; P. Nuttall, "The Passionate Statistician," *Nursing Times* 28 (1983), pp. 25–27.
4. Gregor Mendel, *Experiments in Plant Hybridization*, <<http://www.mendelweb.org/Mendel.html>>; Wikipedia, "Gregor Mendel," http://en.wikipedia.org/wiki/Gregor_mendel>; "Johann Gregor Mendel: Why his discoveries were ignored for 35 years," <<http://www.weloennig.de/mendel02.htm>>.
5. Jiří Sekerák, "Gregor Mendel and the scientific milieu of his discovery," <http://www.2iceshs.cyfronet.pl/2ICESHS_Proceedings/Chapter_10/R-2_Sekerak.pdf>.
6. Wikipedia, "Credit score (United States)," <[http://en.wikipedia.org/wiki/Credit_score_\(United_States\)](http://en.wikipedia.org/wiki/Credit_score_(United_States))>; Malgorzata Wozniacka and Snigdha Sen, "Credit Scores: What You Should Know About Your Own," *PBS Frontline*, November 2004, <<http://www.pbs.org/wgbh/pages/frontline/shows/credit/more/scores.html>>.
7. FICO, <<http://www.fico.com/en/Company/Pages/history.aspx>>.
8. J.M. Gottman, "Predicting the Longitudinal Course of Marriages," *Journal of Marital and Family Therapy* 17, no.1 (1991): pp. 3–7.
9. Edge, "The Mathematics of Love: A Talk with John Gottman," April 14, 2004, <http://www.edge.org/3rd_culture/gottman05/gottman05_index.html>.
10. Xiao-Li Meng, Abstract for "Statistical Education and Educating Statisticians: Producing Wine Connoisseurs and Master Winemakers," seminar presented at the University of Minnesota, October 28, 2011, <<http://catalystsumn.blogspot.com/2011/11/statistics-education-seminar-presented.html>>.
11. David Schmitt, "Tell a Story," June 27, 2012, <http://www.allanalytics.com/author.asp?section_id=2092&doc_id=246428>; The Pixar Touch, "Pixar story rules (one version)," May 15, 2011, <<http://www.pixartouchbook.com/blog/2011/5/15/pixar-story-rules-one-version.html>>.
12. SAS Visual Analytics 5.1 User's Guide, "Working with Automatic Charts," <<http://support.sas.com/documentation/cdl/en/vaug/65384/HTML/default/viewer.htm#n1xa25dv4fyz6n1etsfkbz75ai0.htm>>.
13. "Hans Rosling: Stats that reshape your worldview," filmed February 2006. TED video, 19:53; <http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html>, posted June 2006.
14. James Taylor, *Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics* (IBM Press, 2011).
15. Thomas H. Davenport and D.J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review*, October 2012.