

11

빅데이터가 아니라 스마트 데이터다 (Smart Data, Not Big Data)

‘왜 데이터를 다뤄야 하는가’에 대한 대답 반드시 있어야

데이터를 사업에 접목시키는 것을 오랫동안 직업으로 해온 사람으로서 필자는 이 빅데이터의 유행으로부터 안전하게 탈출하는 방법을 강구하고 있다. 돌아가는 것을 보고 있다면 과거의 모든 버블이 그러하듯이 이 빅데이터 버블도 곧 꺼지게 될 것이기 때문이다. 그 심판의 날은 2~3년 후가 될 수도 있겠지만, 그 날을 피할 수 없으리란 것을 직감하고 있다. 영화 스타워즈(Star Wars)에 등장하는 제다이 마스터(Jedi Master) 요다(Yoda)가 하는 말같이 들릴 수도 있겠지만, 모든 유

행은 과도한 투자를 불러오고, 과도한 투자는 반드시 실망을 가져오며, 그러한 실망은 질책으로 이어지기 마련이다. 과거에 CRM의 경우가 그랬듯이 몇 년 뒤에는 이 빅데이터에 대한 후회와 질책이 있을 것이며, 누군가가 거기에 대한 책임을 떠맡을 것이다.

강력한 주문
'빅데이터',
그러나 그 버블은...

그렇다면 그 어느 누가 왜 이 험난한 길에 남아있을 것인가?
그것은 나중에는 모르겠지만 지금 당장은 그 빅데이터란 말이
많은 기회를 열어주고 있기 때문이다.

얼마 전 2주 일정으로 한국을 방문했을 때 필자는 무려 9차례에 걸쳐 때로는 아주 큰 장소에서, 때로는 작은 규모의 모임에서 이 빅데이터에 대한 강연과 강의를 하고 또 그것을 주제로 한 토론에도 참여했다.

불과 몇 년 전까지만 해도 필자가 뭘 해서 먹고사는 사람인지 설명하기도 어려웠었던 것에 비하면, 지금은 이 빅데이터란 말이 그야말로 '열려라, 참깨' 정도로 강력한 주문이 돼 그간 닫혀 있던 문들을 열어주고 있다. 데이터를 다루는 사람들, 특히 자신을 데이터 사이언티스트라고 부르는 사람들은 그야말로 때를 만난듯 싶겠지만, 이런 현상은 앞으로 일을 제대로 해 좋은 결과를 가져와야만 오랫동안 유지될 수 있을 것이다.

앞으로 이 유행이 어떤 모습으로 지나가든 필자는 그 후에도 계속해서 데이터에 관한 전략을 수립하고, 망가진 데이터를 고치며, 데이터베이스를 디자인하고, 또 분석활동을 이끌어가며 살아갈 것이다. 다만 직책과 부서의 명칭이 달라질 수도 있겠다. 그간 수많은 유행어들이 나타났다 사라져갔지만, 이 데이터 산업은 그런 유행어나 값싼 단어들에 상관없이 수익과 이윤을 창출하는 기회를 만들어온 사람들에게 의해 유지돼온 것이다.

그 활동의 목적이 이익을 높이는 것이든 비용을 절감하는 것이든 결국 데이터를 다루는 일도 다 사업을 위한 것임은 분명하다. 유행어가 때로는 기회를 창출하기도 하지만, 중국에는 창출된 기회를 실제적 수익으로 이어가는 본질에 더 충실해야 하는 법이다. 그리고 그 본질이란 휘황찬란한 현수막이나 타이틀에서 나오는 게 아니다.

정말 가능하기 어려울 정도의 데이터를 보유하고 있는 구글이나 아마존 같은 회사가 자신들을 빅데이터 회사라고 부르고 있는 줄 아는가? 천만의 말씀이다. 그들은 그 데이터를 이용한 모든 활동이 User Experience, 즉 소비자의 경험을 위한 것이지 데이터베이스 크기나 플랫폼의 속도의 문제가 아니란 것을 잘 알고 있다.

본질을 잘 모르고 남을 쫓아 하는 사람들이나 유행어나 유행에 목매는 법이다. 빅데이터, 클라우드 컴퓨팅, 혹은 붙여진 이름이 '쿨하다'고 여겨지는 도구나 프로그램 언어가 알아서 돈을 벌여주지 않는다. 그런 것들은 단지 단어일 뿐이다.

필자가 보기엔 애초에 그 '빅데이터(Big Data)'란 말 자체가 틀려먹었다. 왜냐하면 그 말은 큰 것이 항상 더 좋은 것이라는 인식을 주기 때문이다.

흔히 빅데이터를 규정짓는 3V, 즉 Volume(크기), Velocity(속도), 그리고 Variety(다양성)도 올바른 정의가 못 된다. 그 정의는 늘 데이터를 다뤄온 데이터 프로페셔널들에게는 과거의 데이터와 구분 짓는 의미로 해석될 수도 있겠지만, 의사결정자나 사용자들에게는 크고 빠

‘왜 데이터를 다뤄야 하는가’에 대한 대답 반드시 있어야

르고 다양한 데이터 자체가 목적인 것처럼 착각을 주게 되며, 크고 빠르고 다양한 데이터만 얻으면 모든 문제가 해결될 것이라는 잘못된 인식을 줄 수도 있다.

하지만 의사결정자들에게 데이터란 작을수록 좋은 것이다. 수학전문가가 아닌 사람들이 어느 세월에 어떤 능력으로 수천 가지 데이터 변수들을 분석해 제대로 된 대답을 얻어낼 것인가 말이다. 사업가나 의사결정자들, 혹은 사용자들은 그들의 질문에 대한 대답의 형태를 갖고 있는 작은 데이터가 필요한 것이다.

더욱이 IT세계에서 크고 빠른 것이란 전혀 새로운 것도 없다. 컴퓨터가 발명된 이래로 처리속도가 빨라지지 않았고, 저장용량이 더 커지지 않았던 해가 단 한 번이라도 있었는가?

이 빅데이터란 표현 자체가 소프트웨어나 툴셋을 파는 회사들이 마케팅용으로 만든 말이란 것을 간과해서는 안 된다.

우리는 그 말을 ‘이제는 아주 커다란 데이터도 처리하고 분석할 수 있다’고 새겨서 들어야지, ‘빅데이터가 모든 문제를 저절로 해결해줄 것이다’라고 여겨서는 곤란하다.

물론 툴셋을 파는 회사의 입장에서는 그런 식으로 그들의 제품을 홍보할 수밖에 없을 것이다. 만약에 기타를 제조 판매하는 회사가 제대로 된 기타 연주자가 되는 것이 얼마나 어려운지를 계속 강조한다면 그것이 그들의 사업에 도움이 될 성싶은가? 그냥 ‘이 기타는 당신의 기타 히어로가 사용하는 바로 그것입니다’라고 말하는 것이 훨씬 쉽고 간편한 판매방식인 것이다. 문제는 비싼 기타만 구입했다고 저절로 좋은 연주자가 되는 것이 아니고, 정말 어려운 일은 악기를 산 이후에 시작된다는 거다.

이 간단명료한 현실이 유독 데이터 산업에서만 간과되는 경우가 많다. 툴셋은 저절로 알아서 대답을 주지 않는다. 그런 것들은 일의 진행을 수월하게 할 수도 있겠지만, 결국은 의사결정자들이 논리적으로 질문을 창출하고, 또 분석을 통해 얻어진 대답들을 바탕으로 사업에 대한 결정도 내려야 한다. 그런 주어진 데이터에서 답을 얻어내는 활동도 마치 악기의 연주자가 끊임없이 연습을 하는 것처럼 트레이닝 과정을 필요로 한다.

비즈니스를 하는 사람들은 빅데이터라는 열차에 올라타기 전에 ‘왜 데이터를 다뤄야 하는가’, ‘그것으로 무슨 대답을 얻기를 원하는가’에 대한 대답을 반드시 해야 한다. 거기에 대한 답이 없다면 그 유행에 편승할 이유도 없다. 남들이 다들 올라간다고 그들을 쫓아갈 필요는 없는 것이다. 물론 그것이 아주 크고 재미있는 파티로 보이겠지만, 왜 거기에 가야 하는지부터 정해야 한다는 말이다.

게다가 비즈니스에 관한 질문을 제대로 하다보면 작은 데이터(Small Data)를 가지고도 충분

히 답을 낼 수 있는 경우가 많고, 그 빅데이터란 것이 늘 필요한 것은 아니란 것도 보이게 된다. 오히려 그런 경우에는 빅데이터는 지름길이 아니라 아주 많이 돌아가는 길인 것이다.

하지만 주어진 온갖 데이터를 전부 다 사용하려고 하고, 툴셋을 바탕으로 한 프로세스를 중심으로 질문을 하다보면 그러한 일이 자주 생긴다. 그래서 빅데이터라는 자체가 비즈니스에 관한 명제이지 IT나 데이터의 문제가 아니라고 하는 것이다.

Smart Data, Not Big Data

그렇다면 어떻게 이런 유행어의 족쇄에서 벗어나 본질로 돌아갈 것인가?

흔히 사람들은 자신들이 인식하고 있는 것이 곧 진실이라고 믿기 때문에 그 유행어를 대체할 다른 단어를 제시할 필요가 있다. 필자의 의견은 그 ‘빅데이터(Big Data)’는 ‘스마트 데이터(Smart Data)’가 돼야 한다는 것이다.

구분되지도 않고 정리되지도 않은 멍청한 데이터는 산더미처럼 많이 쌓여봤자 아무짝에도 쓸모없다. 아무런 레이블도 붙어있지 않은 박스들이 1943년부터 면지에 묻혀가며 커다란 창고를 가득 메우고 있다고 상상해보면, 그 창고의 크기만 갖고 자량을 할 일인가? 그 안에 설사 인디어나 존스가 수집해온 보물이 있다 해도, 아무도 그게 어디 있는지도 모르고, 설사 찾아낸다 해도 그걸 갖고 무엇을 할지에 대해 아무 생각도 없다면 그게 다 무슨 소용이겠는가?

그렇다면 어떻게 해야 데이터가 더 스마트해질 것인가? 스마트 데이터란 ‘질문에 대한 작은 대답들’이다. 하나의 일기예보를 내놓기 위해 수천 가지 변수를 고려했을 수도 있지만, 사용된 데이터의 크기와 상관없이 ‘내일 오후에 소나기가 내릴 확률이 70%’라는 한마디가 스마트 데이터인 것이다.

그 대답을 만들기 위해 쓰인 모든 변수들의 숫자와 크기만을 강조하는 것은 어느 누가 벽에 걸어놓을 그림을 사러갔는데 가게 주인이 그림 대신 물감과 붓을 주며 ‘당신이 이 재료로 직접 그림을 그려 벽에 붙이시라’고 말하는 것과 비슷한 형국이다. 한심하다고 들릴 수도 있는 이런 비유가 불행히도 빅데이터의 현실이다.

게다가 데이터 분석을 하다보면 그 재료들도 변변치 않은 경우가 허다하다. 데이터 분야의 컨설팅 일을 하다보면 정리되지도 않고 구분되지도 않은 정말 저저분한 데이터베이스들을 많이 마주치게 된다. 혹자는 그런 상황이 필자의 사업에 도움이 되겠다고 생각할 수도 있겠지만, 단적으로 말해 그런 혼란스러운 경우는 그 누구에게도 유익하지 않다.

분석을 전문으로 하는 사람들은 데이터를 제대로 분석해 의미를 끄집어내는 일을 하길 원하는 것이지, 방치돼있다고 한 데이터베이스를 살려놓고 허구한 날 남이 망쳐놓은 데이터나

데이터의 ‘잡음’을 없애야 작고 의미 있게 만들 수 있어

고치는 일을 바라는 게 아니다.

진정으로 유익한 데이터는 작고, 간결하고, 오류 없이 깨끗하며, 분류도 제대로 돼있는 법이다. 빅데이터란 분석가나 분석의 뒷배경까지 들여다보고 싶은 마니아들에게나 흥미로운 것이지만 일반 의사결정자들이 그런 것까지 관심을 둘 필요는 없다.

필자는 ‘빅데이터란 작아져야만 한다(Big Data must get smaller)’란 주장을 오랫동안 해왔고, 이제는 그것이 데이터 산업 내에서 영향력 있는 움직임이 될 때까지 반복하고자 한다. 이 책을 통해 반복해서 언급했듯이 빅데이터 운동은 주어진 많은 데이터에서 잡음을 제거하고 분석을 통해 의사결정자들이 원하는 대답을 제공하는 것이 돼야만 한다.

데이터에는 정말로 많은 잡음이 있는 법이고, 그것을 잘라내 버리는 것은 데이터를 작고 의미 있게 만드는 첫 걸음이다.

문제는 그 ‘잡음’이나 ‘소음’의 정의가 고정적인 것이 아니라는 점이다. 필자가 학창시절 즐겨듣던 록 음악은 우리 부모님 세대에겐 분명히 소음이었고, 그와 마찬가지로 우리 아이들이 듣는 음악의 일부는 필자에게 소음으로 들릴 수도 있다.

데이터베이스를 만들다 보면 많은 정보가 쌓이게 되는데, 일례로 ‘제품의 색깔’이란 변수는 재고관리를 위한 데이터베이스에는 분명히 필수적이지만, 의류제품을 파는 것이 주목적인 마케팅 용도를 놓고 보면 그것은 잡음일 수도 있고 아닐 수도 있다. 의류판매를 위한 데이터 변수 중에는 스타일, 브랜드, 가격대, 타겟 구매자의 성별 등이 훨씬 더 유익할 수 있고, 색깔은 그저 있으나마나 한 정보일 경우가 많다. 다시 말해 어떤 여성이 빨간 구두를 샀다는 것은 그 고객이 계속 빨간색 제품만 살 것이 아니라면 그다지 중요한 정보가 아니란 뜻이다.

그렇다면 어떤 식으로 무엇이 소음인지 아닌지를 결정할 것인가? 일단 분석의 목적을 분명히 하고, 그 목적을 접근 가능한 변수들을 이용해 수학적으로 표현하며, 그 다음에 수학적 과정이 각 변수의 중요성을 정하도록 하는 것이다. 그런 통계적 방식으로 확신을 갖고 데이터를 줄여나가는 것이며, 그 과정에서 아무리 사람의 눈에 중요하게 보이는 변수라도 목적에 부합하지 않으면 버려지게 된다. 그리고 목적이 달라지면 필요변수도 달라지는 것은 당연하다.

그 수학적인 길을 계속 따라가다 보면 두 번째 명제인 ‘의사결정자들이 원하는 대답을 제공하는’ 단계에 다다르게 된다. 그 작은 대답들이란 예/아니오, 확률, 혹은 어떠한 종류의 점수가 될 수 있겠다. 이전에 언급한 일기예보에서의 예에서 질문은 ‘특정일에 비가 올 확률’인 것이고, 그에 대한 대답은 ‘70%’가 되는 것이다.

통계적 모델링은 결코 쉽거나 간단한 것이 아니지만, 그것은 데이터를 더 스마트하게 만드는 데 있어 필수적인 과정이다. 또한 모델링은 복잡하고 많은 데이터를 간결하게 만들고 빈 곳도 채워주는 가장 효과적인 방법이기도 하다. (제3장: ‘빅데이터의 핵심은 분석, 분석의 중심은 모델링’ 참조)

대부분의 사용자들은 수학이나 통계에 관한 학위를 갖고 있지 않겠지만, 다들 ‘회사 야유회 날 70%의 비가 올 확률’과 같은 정보를 어떻게 이용해야 할지 잘 알고 있을 것이다. 일부는 그러한 대답이 단정적인 ‘예/아니오’의 형태를 갖고 있지 않다고 불평을 할지도 모르지만, 대다수는 그런 확률이나 점수를 제공하는 것이 가공되지 않은 데이터를 사용자에게 그냥 떠넘기는 것보다 훨씬 인간적이라는 점에는 이의를 제기하지 않을 것이다.

영업을 하는 사람들도 수학전문가가 아닌 경우가 대부분일 테지만, 그들도 어떤 대상을 상대로 상품을 판매하고자 할 때 그런 노력이 성공할 확률이 얼마인지 미리 알 수 있다면 그런 정보를 마다할 사람은 하나도 없을 것이다. 그런 세일즈 리드(sales lead)에 달린 확률이나 점수는 결과에 대한 확답은 아닐지라도 그것을 바탕으로 어느 대상을 우선적으로 접근해야 할지를 정할 수 있기 때문이다.

그래서 빅데이터를 운운하는 사람들이나 자신을 데이터 사이언티스트라고 부르는 사람들은 데이터를 더욱 ‘인간적’으로 만들어야 하는 의무가 있는 것이다. 데이터의 크기나 내세우고 플랫폼이나 툴셋에 대해서만 관심을 두며, 정작 사용자들에게는 사업하는데 별 의미도 없고 가공되지도 않은 데이터나 잔뜩 진열해놓는 것은 다 의미 없는 일이다. 데이터를 단순하고 쉽게 만드는 것이 목적이야하지 일을 더 복잡하게 만들지 말라는 것이다.

어떤 이들은 데이터나 절차를 복잡하게 만드는 것이 자신들의 철방통을 지키는 길이라고 여길지 모르지만, 그건 천만의 말씀이다. 그런 태도야말로 이 빅데이터 운동을 망가뜨리는 첫째 요인이다. 우리는 이미 충분히 복잡한 세상에 살고 있으며, 일을 더 복잡하게 만드는 사람이 필요한 게 아니다. 이에 대해 앞으로 이 책의 후반부에서 ‘훌륭한 데이터 사이언티스트가 되는 법’이란 주제를 보다 상세히 다룰 것이다.

사용자들의 바람직한 태도

동시에 의사결정자들이나 사용자들도 데이터에 대한 태도를 바꿀 필요가 있다.

01 목적을 분명히 할 것

이 글의 일관된 주제는 빅데이터는 비즈니스에 관한 것이지 IT나 데이터가 주가 돼서는 안 된다는 것이다. 하지만 많은 사업가들과 의사결정자들 중에는 ‘난 그런 거 몰라’라는 태도로 아예 데이터에 관한 일들에 손도 대기 싫어하는 사람들이 많다.

경영자가 데이터 관련 프로젝트에 투자는 허락할지 몰라도, 그 데이터베이스의 목적

무엇인지 명확하게 제시하지도 않고 그저 장래에 어떤 천재적인 분석가가 홀연히 나타나 그간 정리도 제대로 하지 않고 쌓아놓기만 한 데이터에서 무슨 대단한 의미를 찾아내주기를 기다리고 있는 경우가 흔히 있다.

하지만 그런 구원군은 제대로 원하는 바를 제시하지도 않으면 결코 저절로 나타나지 않는다. 만약에 데이터로 무슨 문제를 해결하고 싶든지도 정해져있지 않다면 데이터 관련 프로젝트를 시작도 하지 않는 것이 낫다. 그런 식으로 일을 시작하면 목적지도 없이 하염 없이 떠돌게 되면서 시간과 예산만 낭비하게 된다.

02 데이터를 심각하게 받아들이 것

꼭 과학자들만 과학적인 사고방식을 필요로 하는 것이 아니다. 컴퓨터가 쏟아내는 많은 정보를 여과 없이 다 받아들이는 태도에도 문제가 있지만(제10장: ‘데이터베이스라고 다 같은 것이 아니다’에서 언급했듯이 데이터에도 많은 오류가 있을 수 있다), 반면 많은 사람들은 아직도 자신의 직관에 우선적으로 의지하며 데이터를 무시하는 경향이 있다.

만약 데이터에 기초한 고객의 성향이 의사결정자들이 갖고 있는 선입견과 전혀 다르게 비치더라도 그런 분석결과를 간단히 무시해서는 안 된다. 오히려 그러한 결과를 새로운 광맥을 찾은 것처럼 소중히 받아들여야 하는 것이다. 사람들의 직관은 분명 빅데이터보다도 더 과대평가돼있다.

03 논리적일 것

비논리적인 질문에는 답이 없다. 그리고 아직은 사람의 마음을 읽을 수 있는 컴퓨터나 톨셋이 존재하지 않는다. 만약에 공상 과학 영화에서와 같이 그런 기계가 존재하더라도 질문을 논리적으로 해야 답이 나오는 법이다.

의사결정자들이 프로그램을 짜는 수준까지 가라는 말이 아니라, 논리적 표현이 어떤 모습 을 띠고 있는지, 그리고 프로그래머들이 기계와 어떻게 소통을 하는지에 대해 기본적인 지식은 갖춰야 한다는 뜻이다. 이 세상은 더 이상 컴퓨터 전문가와 비전문가로 단순 구분이 되는 곳이 아니며, 정보가 넘쳐나는 시대에 기본적인 논리적 소양은 누구나 갖춰야 할 자질이다.

04 작은 성공들을 추구할 것

대규모 빅데이터 프로젝트에 투자하기 이전에 많은 작은 성공사례를 만들어보는 것이 바람직하다. 작은 프로젝트를 다루더라도 그 일을 제대로 마치기 위해서는 모든 과정을 한 번씩은 거쳐야 하는 법이다. 정보의 전체적 흐름을 파악하는 것은 매 단계만큼이나 중요한 것인데, 그것은 많은 오류들이 단계들의 사이에서 일어나기 때문이다.

아폴로 프로그램 전에 제미니 프로그램이 있었듯이, 무작정 달을 향해 가기 전에 대기권 바깥에서 도킹을 제대로 하는 법부터 마스터해야 하는 것이다. 과도한 투자는 IT가 논의를 주도할 때 발생하는 경우가 많은데, 비즈니스를 다루는 사람들이 나서서 초기 단계에서는 중요한 과정이라도 과감히 아웃소싱해 정보와 일의 흐름부터 제대로 배우고 겪어 보는 것이 바람직하다.

05 소비자 중심의 관점을 가질 것

이 빅데이터의 유행이 바꾼 것들은 무엇인가?

그 어느 고객도 마케터가 관심을 갖는 한 채널에 갇혀있지 않다. 하지만 아직도 많은 비즈니스들은 채널, 상품, 브랜드, 심지어는 자신이 속한 부서를 기준으로 소비자를 파악하려 든다. 단지 이메일 수신을 ‘아직’ 거부하지 않았다는 이유만으로 그 아무도 소위 말하는 온라인 전용의 사람(online person)이 아닌 것이며, 그 누구도 일차원적이지 않다.

고로 의사결정자나 마케터들은 그러한 채널, 상품, 브랜드, 혹은 부서 중심의 관점에서부터 탈출해야만 한다. 그런 태도를 고치지 않으면 소비자 중심으로 잘 구성된 데이터베이스를 갖고도 비효율적인 마케팅을 하게 된다.

예를 들자면 이메일을 위해 조직된 부서에서는 이메일 채널만을 중심으로, 특정 상품을 위해 존재하는 부서에서는 모든 대상을 그 한 가지 상품만을 중심으로 보는 경우가 많은데, 그래서 회사 전체의 마케팅이 효율적일 수가 없게 된다. 데이터가 많이 모일수록 소비자를 감동시킬 수 있는 기회는 늘어나는 법인데, 그런 기회들을 마케터들의 협소한 사고방식에 소비자를 가둬둠으로써 날려버리면 곤란하다.

빅데이터는 마케터나 의사결정자들의 나쁜 버릇을 더 견고하기 위해 존재하는 것이 아니다. 데이터가 더 커지고 빨라지면서 마케터들도 더 빠른 속도로 소비자들을 더 짜증나게 만들 수 있게 된 것이다.

그렇다면 이 빅데이터의 유행이 바꾼 것은 무엇인가? 일단 그것은 사람들이 데이터를 대하는 태도를 많이 바꾸었다.

일부는 아주 많은 양의 정보도 겁내지 않고 보게 됐고, 실제로 데이터를 의사결정과정에도 도입하는 조직들도 늘어나고 있다. 많은 이들은 우리가 늘 많은 정보에 둘러싸여 있으며, 그것은 마케팅의 경우뿐 아니라 정치, 미디어, 안보, 의료, 그리고 치안에도 적용될 수 있다는 것을 알게 됐다.

반면에 이러한 정보의 홍수를 우려 깊게 보는 사람들도 많아졌다. 특히 개인정보가 허락 없이 사용되는 것을 걱정하는 사람들이 증가하고 있다. 데이터를 늘 만지던 사람들은 클라우드 컴퓨팅과 오픈소스 환경을 즐기고 있는 반면, 다른 많은 소비자들은 이러한 데이터에 관한 활동들에 의심어린 눈초리를 보내고 있으며, 아예 그 클라우드라는 구름 속에 자신의 그 어떠한 정보도 넣기를 거부하는 사람들도 있다. 심지어는 불편을 감수하면서까지 행동양식을 바꿔 그 어떤 채널로도 자신의 정보와 행동이 추적되는 것을 의도적으로 피하는 사람들도 있는 것이다.

하지만 모든 것이 나빠진 것만은 아니다. 데이터의 양이 실제로 급격히 커졌고, 과거에는 손도 댈 수 없었던 모바일이나 SNS 데이터 같은 다양한 정보도 널리 사용되기 시작했다. 과거

의 포커스 그룹(focus group) 등의 표본 집단은 트위터를 팔로우하는 사람들의 크기만큼이나 커진 것이다.

고객들의 거래활동을 기록하는 POS(Point of Service) 데이터의 수집율도 꾸준히 증가하고 있으며, 그러한 매장에서의 기록들은 데이터 분석이나 사용 시 온라인과 오프라인의 차이를 무색하게 만들 것이다.

그리고 그런 새롭고 많은 양의 데이터를 잘 다루는 사람들의 숫자도 많아졌으며, 그 중 일부는 익숙한 앞선 나머지 가끔은 고객들을 질리게 만드는 사용자들도 있었지만 고객에게 감동을 주는 새로운 방법을 창출해내는 분석가들도 다수 등장했다.

더욱이 많은 다른 종류의 데이터가 함께 사용되기 시작된 것은 고무적인 일이며, 그러한 접목하는 활동들은 앞으로 분석가들의 예측력을 더 발전하게 해 곧 빅데이터 운동의 중심이 될 것이다.

완벽한 데이터베이스는 여전히 이루기 어려운 꿈이지만, 분석가들은 정보의 부재(missing data)에 대해 조금은 걱정을 덜게 됐다. 오픈소스 환경에서 일하는 개발자들은 새로운 플랫폼과 기구들을 이용해 아주 빠른 속도로 새로운 데이터 패키지를 만들 수 있게 됐으며, 과거에 Direct Marketing이란 이름으로 활동하던 초창기 분석가들이 6개월 이상 소요하던 일을 단 몇 시간 내에 할 수 있게 됐고, 가까운 미래에는 그런 일을 단 몇 초 만에 해치울 수도 있게 될 것이다.

그러나 이 모든 것의 끝이 좋을 수도 있고 나쁠 수도 있다. 데이터를 다루는 사람들이 많은 소비자들의 짜증을 부추기지 않고 투자자들의 돈을 낭비하지 않으면서 일을 제대로만 한다면, 우리는 미래를 예측하는 능력을 더욱 키워나가며 데이터를 이용해 의사결정을 더 쉽고 편리하게 만들 수 있는 좋은 위치에 서있다고 할 수 있다.

반면, 정보관리도 제대로 못하거나 정보를 남용해 고객들을 화나게 하고 정작 의사결정자들이 원하는 대답도 내놓지 못해 사업에 직접적으로 긍정적인 영향도 미치지 못한다면, 이 빅데이터는 많은 사업이나 날리게 만든 또 다른 유행어의 하나로만 기억될 것이다.

빅데이터는 남들이 산다고 따라 사는 무슨 가구 같은 것이 아니다. 데이터를 사용하는 것은 장기적으로 의사결정 방식 자체를 바꾸는 행위이다. 그것은 무슨 유행을 좇아 잠깐 시도하다가 그만두는 게 아니다.

특히 마케팅에서의 정보의 사용은 마케터가 시도한 방식들에 대한 구체적 결과가 연속적으

빅데이터란 없다 (There is No Big Data)

로 분석할 더 많은 데이터를 창출해내게 되는 순환고리와 같은 것이다. 그런 순환적 정보의 흐름이 더 훌륭한 효과를 장기적으로 유지하게 하는 유일한 길이기도 하다.

그리고 이러한 모든 데이터 관련 활동들은 ‘Big’이란 단어와 별 상관이 없다.

목적에 따라 제대로만 다룬다면 Small Data도 얼마든지 유용하며, 실상 많은 회사들의 지난 2~3년 동안의 거래기록을 전부 모아 봐도 용량이 좀 넉넉한 스마트폰에 들어갈 정도밖에 안 되는 경우도 많을 것이다.

이는 데이터의 크기가 문제가 아니라 데이터를 갖고 무엇을 할 것이냐는 것이고, 그런 목적은 우선적으로 비즈니스의 관점에서 수립돼야 한다. 빅데이터란 몇몇 IT나 데이터 관련 종사자들의 작은 서클 내에서나 알아주는 ‘쿨한’ 테크놀로지에 국한되는 것이 아니며, 그런 사람들 전용의 놀이터는 더더구나 아니다.

필자는 얼마 전 브라질 상파울루에서 열린 ‘QIBRAS’라는 컨퍼런스에 연사로 참석했는데, 그 커다란 모임의 주제가 ‘Big’이 아니라 ‘Data Quality’, 즉 ‘데이터 품질’이라는 것을 알게 돼 크게 고무됐다. 즉 세상 다른 곳에는 데이터를 ‘Big’이란 관점과 잣대만 가지고 접근하지 않는 사람들도 있다는 말이다.

그리고 둘러보자면 데이터를 그야말로 도사급으로 다루는 사람들치고 빅데이터란 말을 쓰는 사람은 찾아보기도 어렵다. 그들은 그저 사업에 도움이 되니까 크고 작은 데이터를 열심히 다룰 뿐이다. 거기에는 유행어나 커다란 현수막도 필요 없는 게, 당연히 늘 하고 있는 일에 특별한 명칭을 부여하는 것 자체가 쓸데없는 일이기 때문이다.

더욱이 마스터급 데이터 전문가들에게 빅데이터란 없다. 만약에 구글이 난데없이 자신들을 ‘빅데이터 회사’라고 부르기 시작한다면 그것은 전혀 ‘쿨하게’ 들리지 않을 것이며, 오히려 그런 말이 그들이 할 수 있는 일의 범위를 규정짓는 걸림돌로 작용하게 될 것이다. 그것이 바로 우리 모두가 심각하게 생각해봐야 할 점이다.