

1

데이터를 대하는 사용자들의 태도



데이터를 의사결정 과정에 반영하는 것

지난 25년 이상 미국에서 데이터베이스 마케팅 업계에 몸 담으면서 다양한 업종의 데이터와 분석에 관한 프로젝트들을 성공적으로 이끌어온 사람으로서 작금의 빅데이터(Big Data)란 말의 유행에 대해, 남들이 보기에는 좋아 보일지 모르겠지만 필자는 거기에 대해 걱정부터 앞선다. 과거에도 그랬듯이 과도한 유행은 과도한 투자를 부르게 마련이고, 과도한 투자는 실패나 실망과 질책으로 이어지는 데에 예외가 없다. 그 유행하는 것이 Toolset이건 개념이건 마찬가지이다.

멀리 볼 것도 없이 CRM이란 단어를 들어봐 보면 알 것이다. 사실 CRM(Customer Relationship Management)이란 말만 놓고 보면 나뭇게 하나도 없다. 고객의 정보를 수집해서 가공하고, 그 결과를 이용해 맞춤형 커뮤니케이션을 하며 고객관리의 틀을 만드는 것이 비즈니스를 하는 사람의 입장에서 어떻게 나쁜 일일 수가 있겠는가. 하지만 필자가 한국에 와서 데이터에 관한 강연을 하며 사례를 들다 보면 CRM 사춘쯤 되는 비슷한 개념만 언급을 해도 아예 손사래를 치는

분들을 만나는 경우가 비일비재하다. 자라보고 놀란 사람이 솥뚜껑 보고도 놀란다지만, 그런 식으로 데이터나 분석 자체에 알레르기 반응을 보이는 데에서 빅데이터를 논한다는 것은 어불성설이다. 하긴 'small data'도 제대로 만져보지 않던 기업들이 유행을 따라 'Big Data'를 하겠다고 뛰어드는 자체가 우스운 노릇이지만 말이다.

빅 데이터, CRM처럼
'Dirty Word' 될 가능성
경계해야

그렇다면 왜 CRM이 미국식 표현으로 소위 말하는 'Dirty Word'가 된 것일까?

첫 번째 이유는 앞서 말한대로 유행을 타다가 거기에 과도한 투자를 했기 때문이고, 두 번째 이유는 개념의 정립과 정확한 목표, 그리고 성공과 실패에 대한 기준도 없이 일을 시작했기 때문이며, 세 번째 이유는 Toolset만 믿고 기본기가 없는 상태에서 일을 추진했기 때문이다.

사실 그런 상황에서 일이 제대로 된다면 그게 더 이상한 일이다. CRM에 대한 투자는 엄청나게 했는데, 그런 투자를 해서 “성공했다”는 소리를 들으려면 그야말로 반응률과 매출이 두 배 이상 뛰어도 시원치 않을 기적적인 결과를 얻어야 한다. 그런데 데이터베이스 마케팅 분야에서 그런 기적을 바라보고 일을 추진하는 것은 좋게 봐줘서 무식한 짓이고 나쁘게 말하자면 미친 짓이다. 하지만 직함에 ‘CRM’자가 들어가 있는 사람들이 스스로 제대로 못했기 때문이라 말할 수는 없는 노릇이기에, 아예 CRM이란 개념 자체를 Dirty Word로 만들어 버린 것이다.

하지만 제대로 해보지도 않고 그 개념 자체를 나쁘다고 하는 것은 자신이 음치라고 음악이라는 것 자체가 쓰잘데없는 것이라 떠드는 것과 다를 바 없다. 음악을 아무리 좋아한다 해도 기본기도 없고 연습도 안 하는 사람이 비싼 기타만 구입했다고 수준급 기타리스트가 되는 것은 더 더구나 아닐 것이다.

문제는 작금의 빅 데이터도 그런 CRM의 전철을 밟을 가능성이 상당히 높아 보인다는 것이다. 그런 미국에서도 그렇지만 한국에선 더한 것 같다. 과거에 인터넷이 처음 등장했을 때 하도 매스컴에서 떠들어대니 인터넷 공부하겠다고 논발 팔아가지고 상경한 사람도 있었듯이 빅 데이터가 무슨 요술방망이이라도 되는 듯 떠드는 것은 바람직하지 않다.

물론 'Big Data'라는 말이 유행하면서 과거에 데이터를 다루지 않았던 회사들이 거기에 관심을 기울이기 시작한 경우를 미국에서 많이 볼 수 있긴 하다. 그러나 '데이터를 의사결정 과정에 반영하는 것' 자체는 새로운 개념도 아니다. 특히 마케팅의 경우 데이터에 관한 컨퍼런스를 한번 열면 촉구장 대역섯 개 정도의 자리를 차지하고 할 정도이며, 이미 산업별로 적용방법이 수십 년의 테스트와 실전경험을 거쳐 Best Practice까지 정립되어 있는 상황이다. 그러니 다루는 내용도 데이터의 저장능력과 처리 속도가 빨라지면서 그걸 다루는데 어떠한 새로운 방법을 적용할 것인가에 맞춰져 있지, 데이터를 써야 할지 말아야 할지는 논의의 주제가 되지 못한다. 그에 대해서는 이미 답이 나와있는 상태이기 때문이다.

위기 의식만으로 부족,
데이터 잘 다루는
기본기 필요

한가지 분명한 것은 앞으로 데이터를 잘 다루는 회사와
그렇지 못한 회사와의 차이가 더욱 커질 것이란 점이다.

그래서 미국에서도 데이터를 의사결정과정에 도입하지 않은 회사들 내에 엄청난 위기의

식이 자라나고 있다. 그것이 빅데이터란 말이 가져온 가장 큰 변화라고 할 수도 있겠다.

하지만 위기의식만 가지고 일을 추진 했다 가는 재미보기가 어렵다. 기본기가 제대로 되어있지 않은 축구팀이 후반 15분 남은 상황에서 위기의식만 가지고 골을 넣을 수 있는 게 아니듯이 말이다. 그렇다면 데이터를 잘 다루어 수익을 남기려면 어떤 기본기가 갖추어져 있어야 하는가에 대한 대답이 필요하겠다.

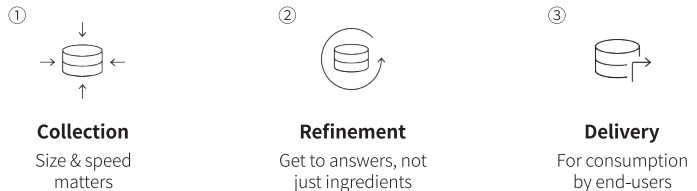
하지만 제대로 해보지도 않고 그 개념 자체를 나쁘다고 하는 것은 자신이 음치라고 음악이라는 것 자체가 쓰잘데없는 것이라 떠드는 것과 다를 바 없다. 음악을 아무리 좋아한다 해도 기본기도 없고 연습도 안 하는 사람이 비싼 기타만 구입했다고 수준급 기타리스트가 되는 것은 더더구나 아닐 것이다.

이 책의 주목적은 데이터에 관한 원론적인 이야기에서 한층 더 깊이 들어가, 어떻게 하면 많은 돈과 시간을 낭비하지 않고도 좋은 결과를 얻을 것이며, 어떻게 해야 전문가가 아닌 사람들도 쉽게 데이터를 의사 결정 과정에 사용할 수 있는지에 대해 구체적인 데이터 가공법, 통계에 기초한 미래 예측 모델을 만드는 데 있어서의 주의점과 최적환경 등에 대해 논하고자 하는 것이다.

우선 짚고 넘어가고 싶은 점은 데이터 방면에서 ‘고수’ 소리를 들으려면 ① 데이터의 수집(Collection), ② 가공(Refinement), 그리고 ③ 전달(Delivery), 이 세가지를 다 잘해야 한다는 것이다. 많은 빅 데이터에 관한 토론을 들어보면 불행하게도 어떻게 하면 많은 양의 데이터를 저장하고 그것을 빨리 꺼내 보는 것인가에 대한 기술적인 면에 치중되어 있는 것을 볼 수 있다.

하지만 그것은 가장 기초적인 첫 단계일 뿐이며, 의사 결정자들이 원하는 ‘대답’의 형태를 갖추려면 여러 가지 가공과 통계학을 이용한 모델들을 만들어야 하는 것이다. 그리고 세 번째 단계로 ‘그 사람이 이해하기 쉬운 형태의 대답’을 이용자들이 ‘선호하는 채널을 통해 쉽게’ 볼 수 있게까지 해주어야 비로소 “데이터를 제대로 사용하기 시작했다”라고 말할 수 있겠다. 여기서 굳이 ‘시작’이란 표현을 쓴 것은 이 과정 자체가 일회용이 아니고, 순환고리로 연결되어 의사 결정자의 행동들의 영향까지도 재수집하는 것이 일상화되어야 하기 때문이다.

데이터 다루기의 세 단계(수집, 가공, 전달)



큰 데이터를 저장하고
빨리 꺼내는 것보다
'가공'이 중요

이중에서도 데이터의 가공과정이 핵심이라고 할 수 있으며,
데이터의 용량이 커지고 의사결정 속도도 더 빨라야 하는
환경에서는 그야말로 필수적인 일이다.

현재와 같은 빅데이터 시대에는 데이터의 규모들이 너무 빨리 커져서 사람들이 그것을 제대로 이해하고 사용할 수 없는 것이 문제이지 단지 데이터를 쌓아 놓을 데가 없어서 걱정할 일은 아닌 것이다. 버릴 것은 버리고 고칠 것은 고치며, 빈 곳이 있으면 메워주고, 확답을 주기가 어려우면 확률로 표현하는 것, 이 모든 것이 다 가공과정에 속한다.

그리고 그것을 주도하는 것은 경험을 가진 사람이어야 한다. 그렇지 않고 처음부터 Toolset에 만 의지하여 모든 것을 자동화하려 한다면 낭패로 이어지기 십상이다. 기계가 어찌 무엇을 어떻게 버릴지까지 정해줄 것이며, 어떤 데이터가 프라이버시에 관한 문제를 일으킬지 판단할 것이며, 사람의 '의도'를 수학적으로 '표현'까지 해 줄 것인가 말이다.

사실 그냥 모아 놓은 데이터(Raw Data)와 가공된 대답(Refined Answer)의 차이는 매우 크다. 마케팅의 예를 들자면 구매기록, 사용한 상품, 구매한 채널, 설문조사에 대한 대답, 그리고 대상자들의 인구적 데이터(Demographic Data)가 가공되지 않은 데이터의 예가 되겠다.

하지만 그런 단편적인 데이터는 '도움'은 될지 몰라도 '대답'까지 주지는 못한다. 비즈니스 분야에 따라 다르겠지만 마케터들이 알고 싶어하는 것들은 “누가 무슨 물건에 관심이 있는가”, “어떤 고객이 VIP가 될 것인가”, “누구에게 credit을 줄 것인가”, “가격이 지역과 채널에 따라 최적화 되어 있는가”, “어디에 매장을 열어야 수익이 극대화 될 것인가” 등이지만, 단지 나열된 데이터가 아닌 것이다.

툴셋은 툴셋일 뿐,
경험을 가진 '사람'이
주도해야

그렇기 때문에 이 모든 일의 시작은
“대체 데이터에서 어떤 대답을 원하는가”이어야 한다.

그냥 데이터베이스를 만들어 놓고 사용자가 알아서 의미를 찾아내길 바라는 것은, 마치 설계도도 없이 건물을 짓기 시작 하거나 영화감독 없이 영화를 찍기 시작하는 것과 다를 바가 없다. 그런 프로젝트에 참여한 사람들이 아무리 유능하고 열심히 일을 한다 해도 결과가 좋을 수는 없는 것이다. 게다가 앞서 언급했듯이 Toolset에 관한 맹신은, 사용자의 의도까지 알아서 일을 해주는 컴퓨터가 미래에 등장하기 전까지는, 돈을 낭비하고 일을 그르치는 지름길이다. 물론 Toolset은 반드시 필요하지만 하둡(Hadoop)이건 SAS건 R이건, 그런 것들은 망치나 대패이지 설계자가 아니란 뜻이다.

비즈니스를 하는 이들이 일반적으로 데이터베이스에서 원하는 것(목적)을 몇 가지 적어보자면

- Order fulfillment - 고객이 주문한 오더를 이행하기
- Inventory management and accounting - 재고관리와 회계
- Contact management for sales - 영업을 위한 고객관리
- Dashboard and report generation -대시보드나 리포트를 통한 지표보고
- Queries and selections - 특정 질문에 대한 대답과 간추린 정보
- Campaign management - 마케팅 캠페인 관리
- Response analysis - 캠페인의 결과분석
- Trend analysis - 동향분석
- Predictive modeling and scoring - 예측적 모델링과 그 모델의 알고리즘을 적용하기 등이 있다.

그리고 이런 다양한 목적들을 분야별로 구분하자면

- Customer Acquisition vs. Customer Relationship Management - 새 고객창출과 기존 고객관리
- B-to-B vs. B-to-C - 비즈니스 대상과 일반인 대상
- Domestic vs. International - 국내용과 해외용
- Marketing channels – mass media, online, email, direct marketing etc. - 매스컴, 웹/온라인, 이메일, 디렉트메일 등 마케팅 채널

등으로 나누어 볼 수 있겠다.

결론적으로 말하자면 이런 방대한 목적과 채널들에 대한 대답을 한꺼번에 다 줄 수 있는 데이터베이스는 불가능하지는 않지만, 그런 식의 프로젝트는 공사비를 수십 배로 불리는 함정이란 것이다. 자동차 설계에 비유하자면 ‘스포츠킴만큼 빠른 트랙’을 처음부터 만들겠다고 덤비는 격이다.

“어떤 대답을 원하는가”
목적 명확히

또 다른 종류의 오류를 보자면 어떤 한가지의 목적으로 만들어진 데이터베이스를 다른 목적으로 사용하는 것이다.

예를 들자면 회계와 재고관리를 주 목적으로 하는 데이터베이스를 마케팅 용도로 사용하면 불편한 점이 한두 가지가 아니게 된다. 이렇게 되면서 통계하는 사람들이나 분석가들이 대부분의 시간을 데이터를 고치고 가공하면서 보내는 경우를 흔히 보게 된다.

다시 자동차의 예를 들자면, 그것은 마치 100명을 운송하는데 2인용 스포츠카를 이용하는 격이다. 한 사람씩 옆에 태우고 백 번을 왔다갔다하는 것이 불가능 하지않겠지만 그건 누가 봐도 미친 짓이다. 하지만 데이터베이스 마케팅이 수십 년 이상 자리 잡아 온 미국조차도 데이터베이스에 관한 한 이런 일이 비일비재하게 벌어진다. 차에 대한 기본 개념도 없는 사람이 자동차를 디자인 할 수 없듯이 데이터베이스의 디자인과 이용도 마찬가지다.

이 책을 통해 반복적으로 나올 주제이겠지만, **목적이 분명한 데이터베이스를 만드는 것이야말로 성공의 지름길이다.** 그리고 그 '성공'에 대한 정의도 막연해서는 안 된다. 만약 효율적인 마케팅이 목적이라면 어느 정도의 반응률이 나오면 성공이라 할 수 있는지, 어느 정도의 비용 절감이 목표인지도 현실성 있게 정의하고 일을 시작해야 한다.

또한 마케팅을 주목적으로 하는 데이터베이스라고 다 비슷한 게 아니다. 수익모델, 판매방식, 판매자와 구매자가 선호하는 채널 등에 따라 구조 자체가 완전히 달라질 수 있기 때문에 주의해야 한다. 특히 미래지향적인 분석(Predictive Analytics)이 목적에 포함되어 있다면, 일반적인 Toolset들이 제공하는 것 이상의 가공이 필요하다는 것을 염두에 두고 일을 시작해야 한다.

요술 방망이는 없다

앞서 말했듯이 CRM 자체를 실패라고 부르는 경우를 보자면, 뭐가 실패인지에 대한 정의도 없는 경우가 많다.

단지 거기에 대한 투자를 많이 했으니 무슨 수를 쓰더라도 본전을 뽑아야 한다는 것인데, 그것은 욕심이지 목표가 아니다. 게다가 특히 데이터베이스 마케팅이란 수많은 테스트와 시행착오를 겪으며 거기서 배운 점을 다음 캠페인에 써먹는다 라는 태도로 차근차근 성공을 이끌어 나가는 것이어야 한다. 어떤 의사결정 과정을 데이터를 사용하는 쪽으로 바꿨다고 당장 대박이 나길 바라는 것 또한 욕심이지 목표가 될 수 없다. 그것은 마치 야구 감독이 9회말에 지고 있는 상태에서 대기하고 있는 모든 타자가 안타나 홈런을 쳐서 경기에 이기길 기대하고 손을 놓고 있는 것과 다를 바가 없는 것이다. 거듭 말하지만 Big Data건 small data건 그 어떤 Toolset이건, 다들 도구일 따름이지 요술 방망이는 아니다.

이 책의 목적은 데이터를 다루고 사용함에 있어서 구체적으로 어떠한 방법이 성공으로 이끄는 길이 되고, 더 중요하게는 과거의 사례로 보아 어떤 길을 반드시 피해야 하는지에 대한 경험을 나누기 위함이다. 거기에는 데이터베이스의 구조, 데이터를 제대로 가공하는 방법, 특히 통계에 기초한 분석(Analytics)의 변천사와 빅데이터 시대에서의 새로운 역할, 그런 환경을 만드는 방법과 유념해야 할 점 등이 다루어질 것이다.

원래 이 데이터에 관한 일들은 거시적인 목표와 미시적인 방법들이 마치 영화감독이 그의 비전을 화면으로 옮길 때 큰 그림과 작은 그림들을 동시에 봐야 하듯이 함께 어우러져야 하는 것이다. 많은 독자와 사용자들이 데이터 관련 작업에 큰 예산을 써가며 시행착오를 하지 않길 바라는 마음으로 글을 이어 나가겠다.