

보험사 사례

산업 분야와 관계없이 데이터분석가로 일하는 과정은 대부분 비슷한 일련의 과정을 겪습니다.

1. 문제점 파악
2. 정보 수집 및 전문가 인터뷰
3. 가설 설정
4. 가설 검증을 위한 데이터 수집 및 분석
5. 검정과정에서 발견된 결과를 이용하여 해결책 개발
6. 해결책 테스트 및 해결책의 효과 측정

쉬운 이해를 위하여 보험회사 사례를 들어보겠습니다. 보험회사 영업회의 시간입니다. 영업 상무님이 코로나로 대면 채널이 어려워 텔레마케팅을 시도해 볼 계획인데, 워낙 반응율이 낮아(반응율 1%, 즉 100 명에게 전화하면 1명이 보험가입), 걱정이라고 하십니다. 그리고 신입 데이터사이언티스인 홍철에게 좋은 방법이 있겠냐고 물어봅니다. 홍철은 고민하다가

(1) 문제점파악: 전화로 보험을 잘 구매할 고객 군을 타겟팅해서 텔레마케팅을 하면 어떻겠냐고 대답합니다. 상무님은 그게 된다면 좋겠지만, 가능하겠냐고 반문하셨고. 이런 저런 이야기로 회의가 마무리 되었습니다. 홍철은 회의 후 자리로 돌아와 고민에 빠졌습니다. 신입으로 데이터사이언티스의 가치를 보여줄 좋은 기회인데, 어떻게 텔레마케팅에 반응율이 좋은 고객군을 찾아낼 수 있을까?

(2) 정보 수집 및 전문가 인터뷰: 일단 전화영업 담당자 인터뷰를 통해 관련 지식과 가설 설정에 도움이 될 만한 정보를 수집해 봅니다. 전화영업 담당자는 주로 인구통계에 의한 결과를 공유해 줍니다. 고령자고, 남성이 더 반응율이 좋다고 합니다. 아주 좋은 정보를 얻었습니다. 또 다른 담당자를 만났습니다. 이 분은 누구에게 전화를 하는 것보다는 어떤 텔레마케터가 전화를 하느냐가 더 중요하다고 합니다. 성과가 좋은 텔레마케터는 연령대에 상관없이 좋은 반응율을 보인다고 합니다. 또 좋은 정보를 얻었습니다. 텔레마케터 지인이 있어, 개인적으로 만나봤습니다. 이 분은 일단 전화를 받을 시간이 있는 사람에게 전화를 해야 한다고 합니다. 아무래도 블루칼라보다는 사무직이 전화받을 시간이 있는 것 같더라고 귀띔을 해 줍니다. 홍철은 소득에 관련해서도 물어봅니다. 하지만 전화를 받는 사람의 소득은 잘 모르겠다고 합니다.

(3) 가설 설정: 현재까지 정보를 바탕으로 몇 가지 가설을 세웁니다. 여기서 가설이란 “아마 이런 이유일 때문일 것이다” 하고 추정해보는 것입니다. 예를 들어 의사가 환자를 진단하는 절차도 비슷할 것입니다. 환자가 들어왔습니다. “아랫 배 많이 아픕니다” 라고 이야기를 합니다. 그러면, 의사는 (1) 상한 음식을 먹어서 장염이 발생했나? (2) 화장실을 자주 못가서 그렇나? (3) 아랫 배에 충격이 있었나? 등 여러가지 가설을 설정하고 환자와의 대화를 통하여 해답을 얻을 것입니다.

텔레마케터 담당자와 인터뷰를 통하여 새울 수 있는 가설은 아래와 같습니다. 좋은 가설은 업무 경험에서 나옵니다.

- 고령자일수록, 남성일 수록 반응율이 좋다.
- 반응율은 연령대와 상관없이 텔레마케터의 능력에 달려있다.
- 전화를 받을 시간적 여유가 있는 사람이 반응율이 좋다.
- 소득이 많을 수록 반응율이 좋다.

위 가설을 증명하기 위해서는 데이터를 수집해야 합니다. 하지만, 신규고객을 대상으로 테스트 마케팅을 하지 않는 이상, 위 정보를 얻을 수 는 없습니다. 가장 좋은 방법은 기존 고객을 대상으로 한 과거 캠페인 데이터를 수집하는 것입니다. 과거 기존 고객을 대상으로 한 캠페인 로그파일을 추출합니다. 기존 고객을 대상으로 교차판매 캠페인이므로 반응(신규 보험가입) 여부와 고객 프로파일이 존재합니다. (4) 가설 검증을 위한 데이터 수집 및 분석: 연령별, 성별로 반응율은 분석합니다. 각 텔레마케터 별 연령, 성별 분석도 합니다. 텔레마케터의 프로파일과 대상고객사의 프로파일도 비교합니다. 시간적인 여유가 있는 고객인지는 모르겠습니다. 하지만, 직업군으로 추정해볼 수 있습니다. 다행히 보험심사에 수집한 직업군 정보가 있습니다. 사무직이 현장직보다는 시

간적인 여유가 있을 거라고 생각합니다. 직업군별로 반응율을 분석합니다. 또 고객의 소득은 모르겠습니다. 하지만, 거주지의 특성으로 소득을 추정해 봅니다. 상식적으로 도곡동 타워팰리스 거주자가 중소도시 아파트 거주자보다는 고소득일 확율이 높습니다.

(5) 검정과정에서 발견된 결과를 이용하여 해결책 개발: 가설 검정 분석에서 여러가지 분석결과를 얻었습니다. 연령별, 성별 평균 반응율, 직업군별 평균 반응율, 소득별 평균 반응율이 알게 되었습니다. 반응율에 유의미한 변수(피쳐) 등을 알아내었고, 반응율이 높은 고객군을 만들어보기로 하였습니다. 분류할 변수가 많아 고객군을 추출하기가 좀 어렵습니다. 이를 해결하기 위해 통계 스코어링 모델을 만들기로 합니다. 반응은 예/아니오는 이진 분류이므로 로지스틱회귀모델을 만들어서 고객 스코어링을 합니다. 그리고 스코어가 높은 순으로 마케팅 대상고객을 선정합니다.

(6) 해결책 테스트 및 해결책의 효과 측정: 로지스틱 회귀모델이 얼마나 효과있는 지 알기 위해서 약 1천명의 고객은 랜덤하게 추출하고, 1천명은 모델 스코어에 의해 추출합니다. 그리고 테스트 텔레마케팅을 하고 반응율의 차이를 비교합니다. 랜덤하게 뽑힌 대상은 이전과 동일하게 1%의 반응율을 보였고, 모델을 통하여 뽑힌 대상은 2% 반응율을 보였습니다. 즉 2천명을 랜덤하게 뽑았으면 20명의 신규고객을 얻었을 것이고, 회귀 모델로 2천명을 뽑았으면 40명의 고객이 생겼을 것입니다. 이 번 캠페인에서는 천명씩 테스트 했으므로 30명(10명 + 20명) 의 고객이 생겼습니다. 즉 모델로 10명의 고객을 더 획득하였고, 한 고객이 가져오는 현금흐름의 현재가치가 100 만원이 라면, 이번 테스트 마케팅에서 보여준 모델의 가치는 1천만원 됩니다.

참고로 데이터분석가 해결책을 만드는 방법은 맥킨지 컨설팅이 해결책을 제시하는 방법과 유사합니다. 맥킨지 컨설팅이 고객의 문제를 해결하기 위해서는 중요하게 생각하는 3 요소는 다음과 같습니다. 첫번째, 선입견없이 아무것도 모른다고 생각하고 문제에 접근할 것(zero based), 두 번째, 생각을 MECE(Mutually Exclusive Collectively Exhaustive) 하게 구조화할 것. 세번째, 가설기반으로 분석하고 해결책을 만들 것(Hypothesis driven). 이 중에 세번째가 빠르게 해결책을 찾는 핵심입니다. 데이터 마이닝이라는 접근법도 있습니다. 가능한 모든데이터를 한 곳에 집중시켜 분석함으로써 알 수 없었던 새로운 통찰을 알아내는 방법인데요. 대표적인 예가 월마트의 기저기와 맥주 에피소드(별도 에피소드 설명)입니다. 하지만, 이 방법은 시간이 오래걸리고, 유의미한 결과를 얻지 못하는 경우도 종종 있습니다. 사례 3 번에서 간단하게 소개하도록 하겠습니다.