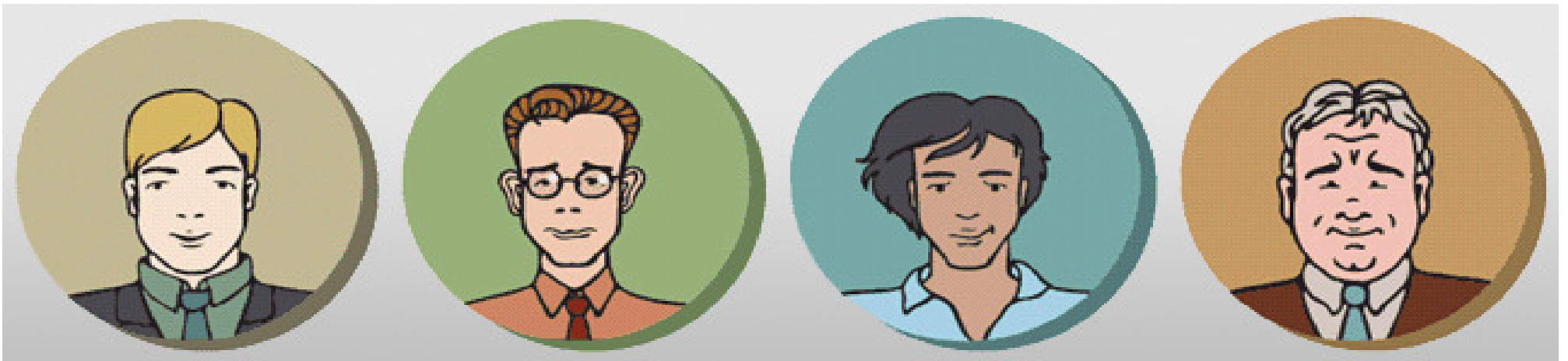


강현욱 선임
(주)데이타솔루션

보험 설계사 리크루팅 애널리틱스

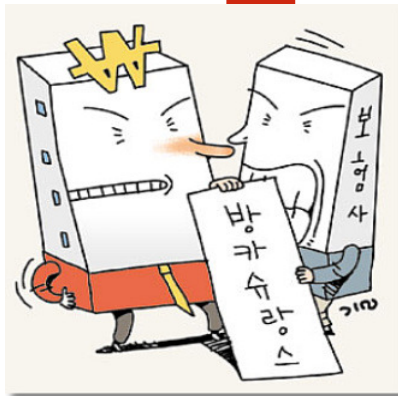


보험업계 경쟁환경



보험

대리점 및 설계사 간
고객 정보 미 공유,
고객 선점



전통적인 채널

보험

• 보험료 책정 기준: 나이, 성별, 건강(연령별)

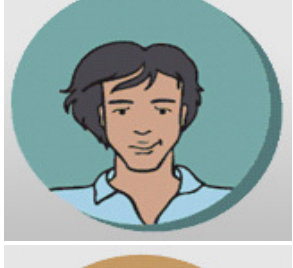
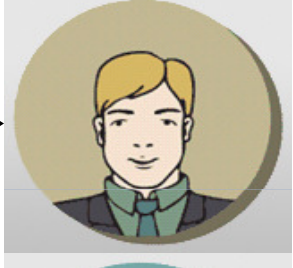
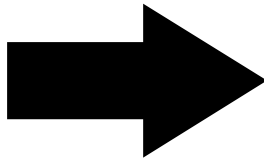
나이	보험료	나이	보험료
19세	10,000	20세	10,000
21세	10,000	22세	10,000
23세	10,000	24세	10,000
25세	10,000	26세	10,000
27세	10,000	28세	10,000
29세	10,000	30세	10,000
31세	10,000	32세	10,000
33세	10,000	34세	10,000
35세	10,000	36세	10,000
37세	10,000	38세	10,000
39세	10,000	40세	10,000
41세	10,000	42세	10,000
43세	10,000	44세	10,000
45세	10,000	46세	10,000
47세	10,000	48세	10,000
49세	10,000	50세	10,000
51세	10,000	52세	10,000
53세	10,000	54세	10,000
55세	10,000	56세	10,000
57세	10,000	58세	10,000
59세	10,000	60세	10,000
61세	10,000	62세	10,000
63세	10,000	64세	10,000
65세	10,000	66세	10,000
67세	10,000	68세	10,000
69세	10,000	70세	10,000
71세	10,000	72세	10,000
73세	10,000	74세	10,000
75세	10,000	76세	10,000
77세	10,000	78세	10,000
79세	10,000	80세	10,000
81세	10,000	82세	10,000
83세	10,000	84세	10,000
85세	10,000	86세	10,000
87세	10,000	88세	10,000
89세	10,000	90세	10,000

• 보험료 책정 기준: 나이, 성별, 건강(연령별)

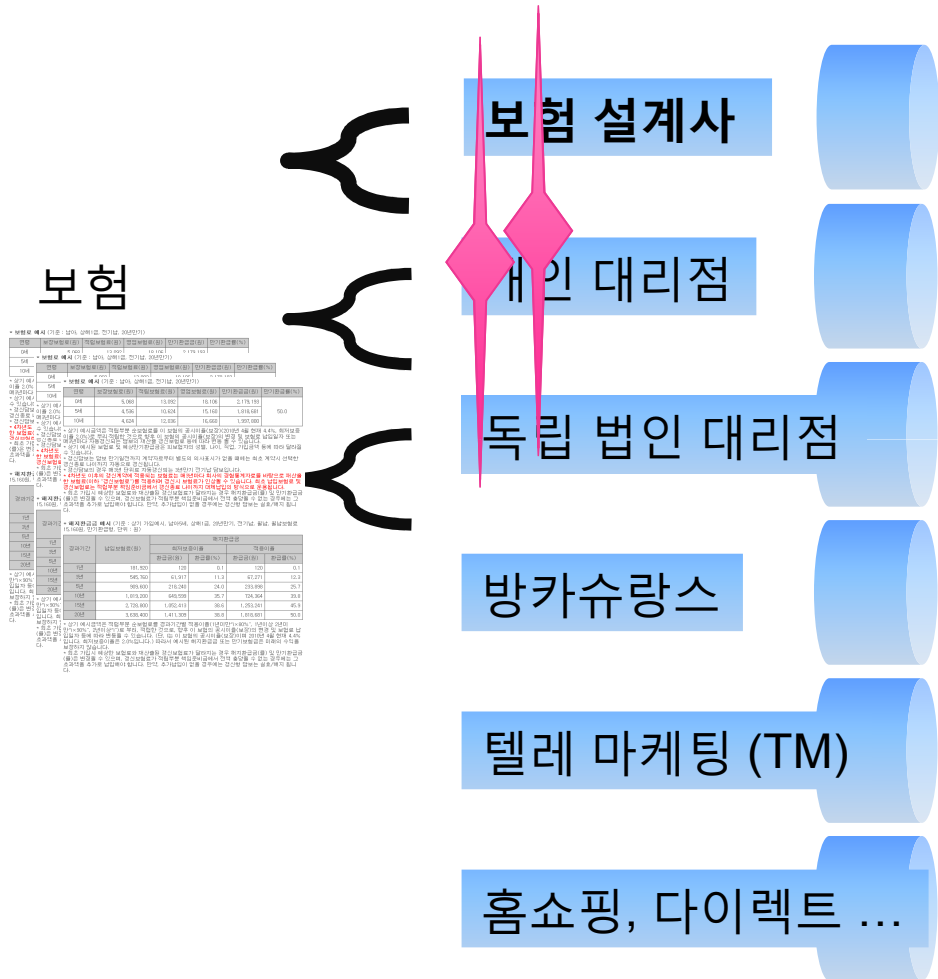
나이	보험료	나이	보험료
19세	10,000	20세	10,000
21세	10,000	22세	10,000
23세	10,000	24세	10,000
25세	10,000	26세	10,000
27세	10,000	28세	10,000
29세	10,000	30세	10,000
31세	10,000	32세	10,000
33세	10,000	34세	10,000
35세	10,000	36세	10,000
37세	10,000	38세	10,000
39세	10,000	40세	10,000
41세	10,000	42세	10,000
43세	10,000	44세	10,000
45세	10,000	46세	10,000
47세	10,000	48세	10,000
49세	10,000	50세	10,000
51세	10,000	52세	10,000
53세	10,000	54세	10,000
55세	10,000	56세	10,000
57세	10,000	58세	10,000
59세	10,000	60세	10,000
61세	10,000	62세	10,000
63세	10,000	64세	10,000
65세	10,000	66세	10,000
67세	10,000	68세	10,000
69세	10,000	70세	10,000
71세	10,000	72세	10,000
73세	10,000	74세 </tr	



보험 설계사



다양한 채널



보험

보험 설계사 (1인) 실적 (2012년 실적)

구분	보험 설계사 (1인) 실적 (2012년 실적)	보험 설계사 (1인) 실적 (2011년 실적)
매출액	1,191,161	1,191,161
인원	1,191,161	1,191,161

개인 대리점 (1인) 실적 (2012년 실적)

구분	개인 대리점 (1인) 실적 (2012년 실적)	개인 대리점 (1인) 실적 (2011년 실적)
매출액	1,191,161	1,191,161
인원	1,191,161	1,191,161

독립 법인 대리점 (1인) 실적 (2012년 실적)

구분	독립 법인 대리점 (1인) 실적 (2012년 실적)	독립 법인 대리점 (1인) 실적 (2011년 실적)
매출액	1,191,161	1,191,161
인원	1,191,161	1,191,161

방카슈랑스 (1인) 실적 (2012년 실적)

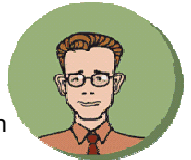
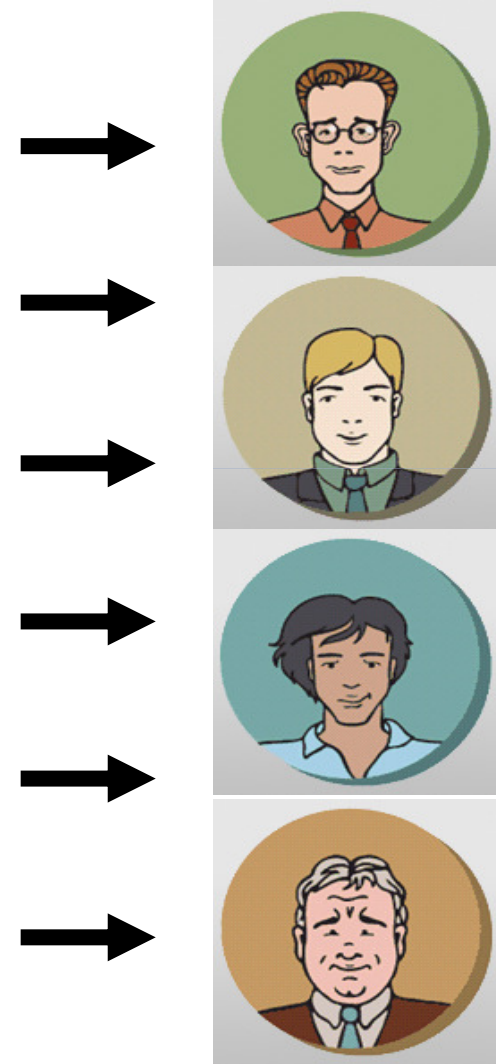
구분	방카슈랑스 (1인) 실적 (2012년 실적)	방카슈랑스 (1인) 실적 (2011년 실적)
매출액	1,191,161	1,191,161
인원	1,191,161	1,191,161

텔레 마케팅 (TM) (1인) 실적 (2012년 실적)

구분	텔레 마케팅 (TM) (1인) 실적 (2012년 실적)	텔레 마케팅 (TM) (1인) 실적 (2011년 실적)
매출액	1,191,161	1,191,161
인원	1,191,161	1,191,161

홈쇼핑, 다이렉트 ... (1인) 실적 (2012년 실적)

구분	홈쇼핑, 다이렉트 ... (1인) 실적 (2012년 실적)	홈쇼핑, 다이렉트 ... (1인) 실적 (2011년 실적)
매출액	1,191,161	1,191,161
인원	1,191,161	1,191,161

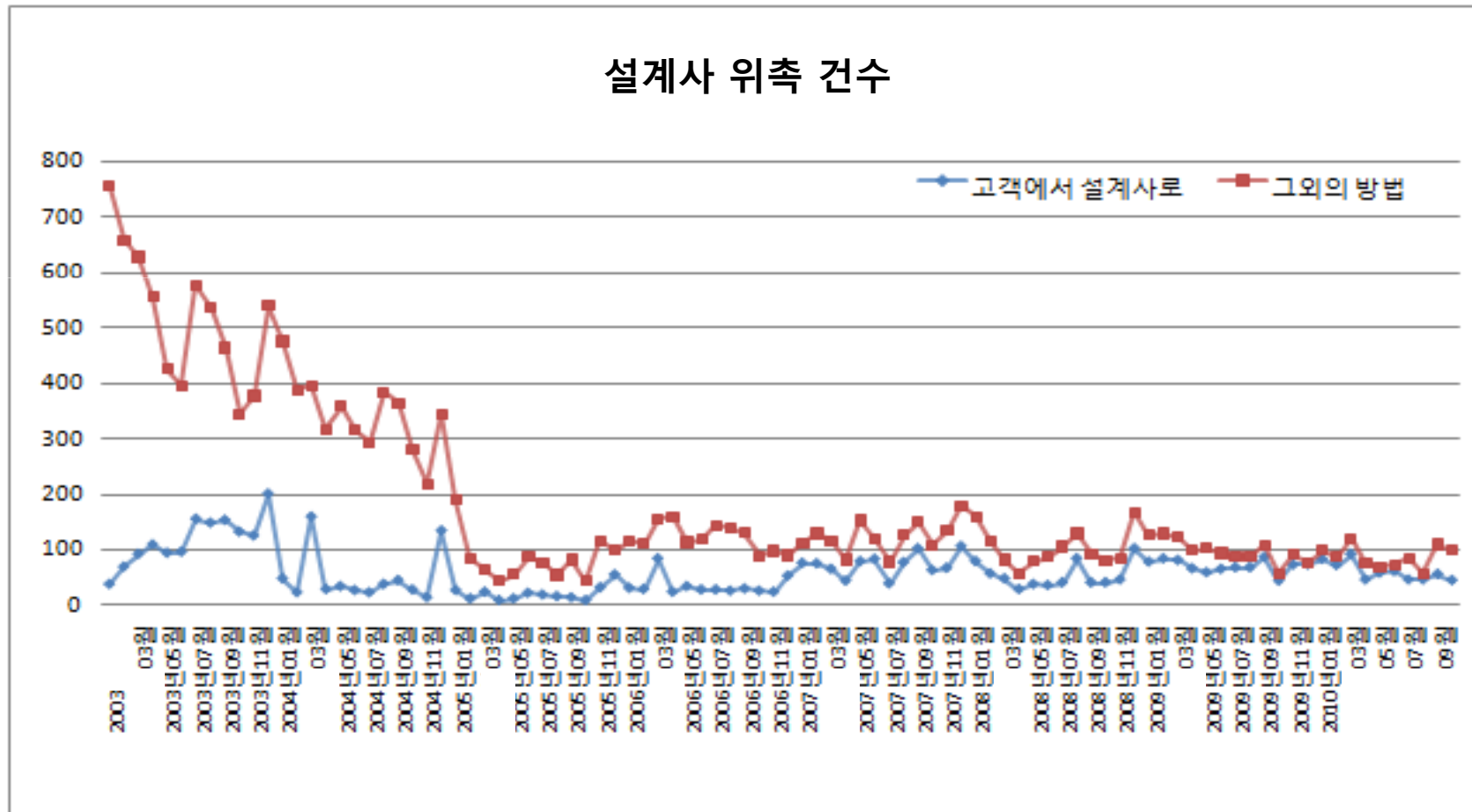


보험사 고민



문제점 발견

- ❖ 자본시장통합법 시행으로 보험 판매 채널이 증가하여 보험업계의 경쟁이 치열해짐에 따라 보험 설계사 (Advisor)의 위축 환경도 날로 어려워짐



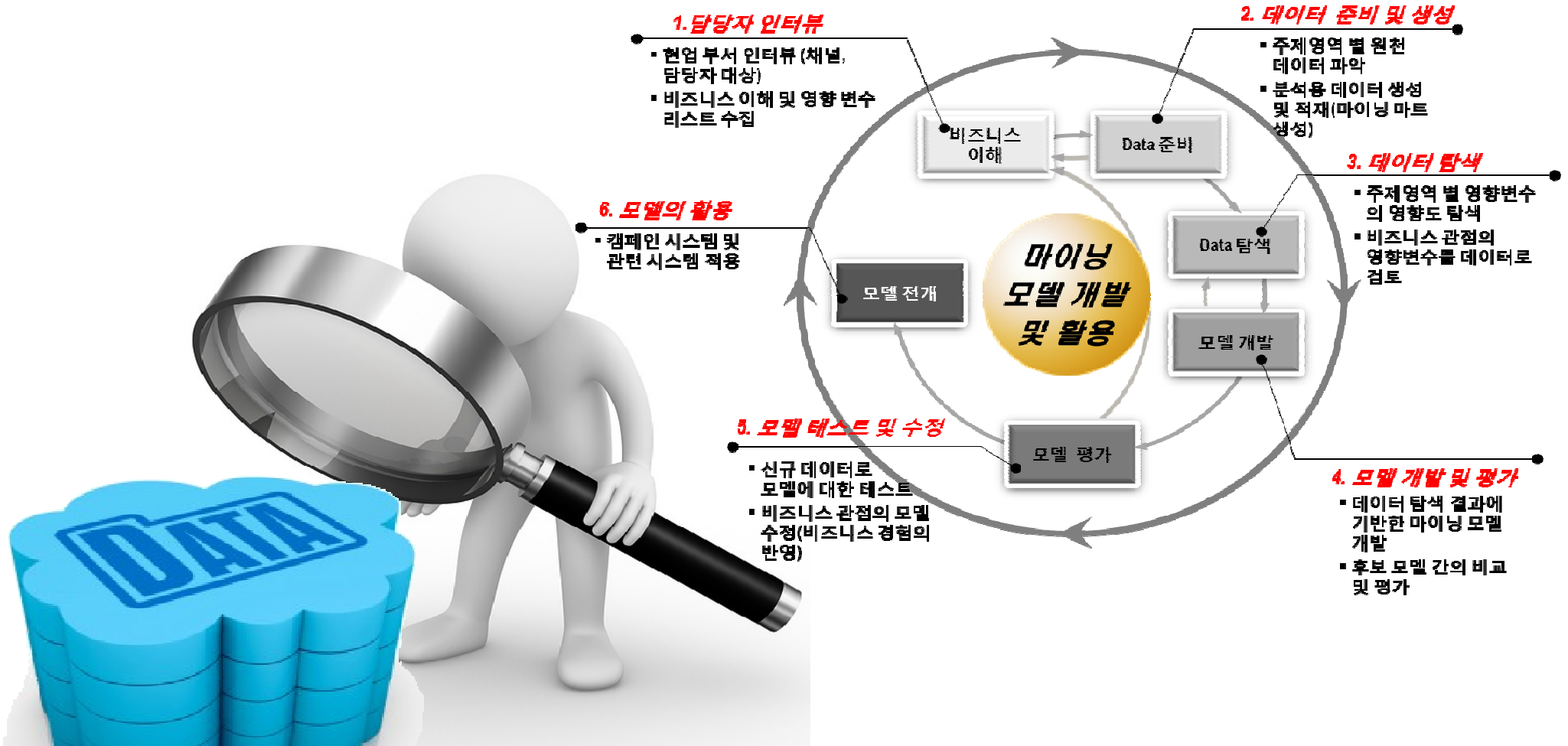
어떻게 접근할 것인가?



딱 보이면 좋을 텐데...



데이터 분석 - 방법론(CRISP-DM)



데이터 분석 - 비즈니스의 이해 및 데이터 준비



**분석
목표**

효율적이고 체계화된 보험 설계사 리크루팅을
위해서 기존 고객에서 설계사로 전환가능성을
스코어링

데이터

2010년 8월 마감 기준 보유고객

**알고싶
은 것**

계약자(기존보유고객)에서 설계사로 전환한 고객은
어떤 특징을 갖고 있는가?



데이터 분석 - 데이터 탐색1

<전체 고객 중 설계사로 전환한 고객의 비율>

	Target	Control
비율	약 1%	약 99%

<보험 설계사 중 기존 계약자에서 위촉된 비율>

	기존 계약자	그 외
비율	약 3%	약 97%



데이터 분석 - 데이터 탐색2



Q.1

그들은 어떤 상품을 구매하는 고객인가 ?

Q.2

어떤 직업을 갖고 있었던 고객인가?

Q.3

몇 개의 상품을 유지하는 고객인가?

Q.4

보험금을 지급받은 경험이 있는가?

Q.5

보험을 가입하고 얼마 후에 설계사가 되는가?

Q.6

그들의 실적은 좋은가?

⋮



데이터 분석 - 데이터 탐색3

❖ 고객의 특성 및 인구 통계데이터, 구매 상품 및 계약 데이터, 거래 데이터, 담당 설계사 데이터 등 고객을 이해 할 수 있는 여러 관점의 약 100개의 변수를 대상으로 분석



Data1

고객 데이터 (고객 정보 및 인구통계학)

Ex) 연령, 성별, 학력, 직업, 소득 ...

Data2

구매 상품 및 계약 데이터

Ex) 구매 상품 수, 구매 상품 군, 유지 계약수 ...

Data3

거래 데이터

Ex) 월 환산 합계 보험료, 보험금 지급 여부 ...

Data4

담당 설계사 데이터

Ex) 담당 설계사 근무년 수, 성별, 학력 ...



데이터 분석 - 데이터를 통한 검증

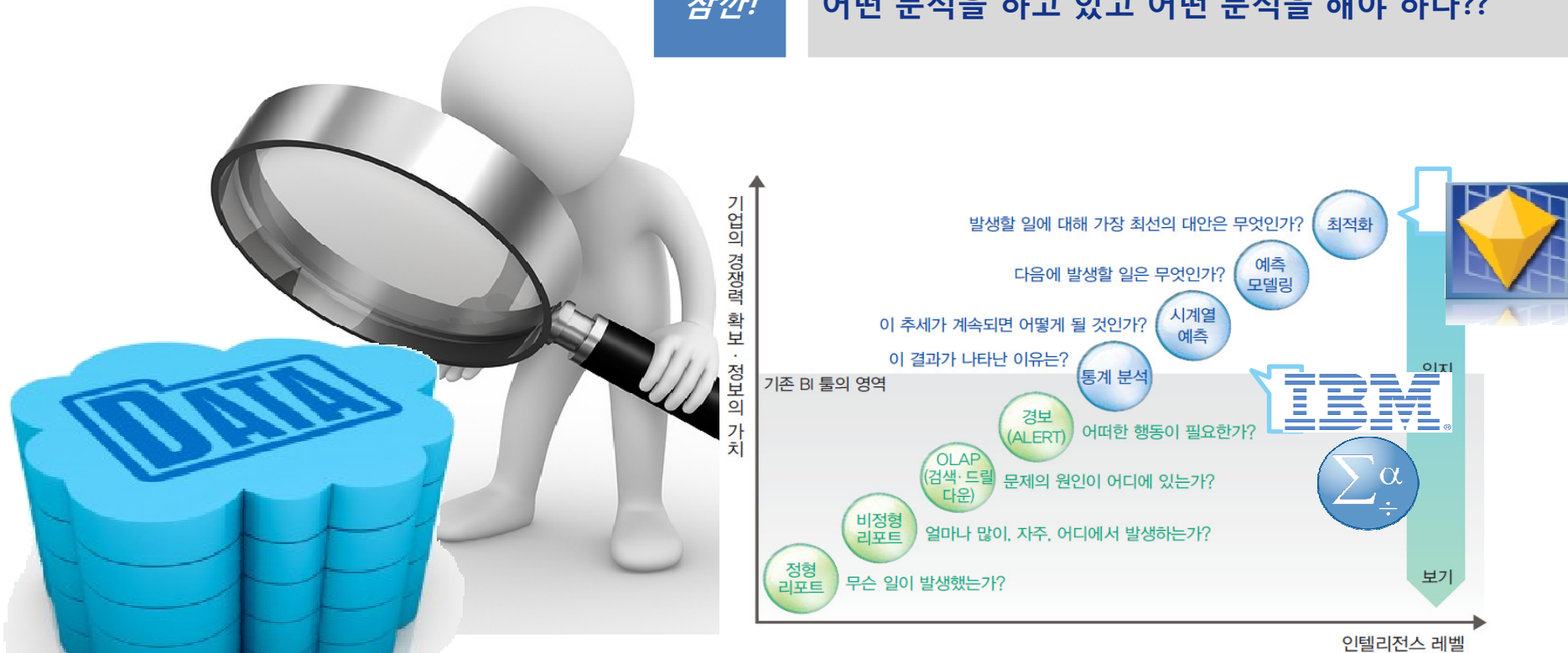


데이터 분석 - 데이터 분석의 깊이

- ❖ 대부분의 회사에서 데이터 분석을 사용하고 있으나 의미를 발견 하고 활용하는 것은 천차만별임
분석 수준은 어느 위치에 와있는지? 어디까지 끌어올릴 것인지? 확인 필요

잠깐!

어떤 분석을 하고 있고 어떤 분석을 해야 하나??



모델 개발 - 모델링 기법(1)

- ❖ 의사결정나무 모델(Decision Tree Model)은 Target을 몇 개의 소집단으로 분류하는 기법으로 교호작용을 반영할 수 있는 예측 기법

의사결정나무 모델(Decision Tree Model)

- 의사결정나무는 의사결정규칙을 도표화하여 **관심대상이 되는 집단을 몇 개의 소집단으로 분류**하거나 예측을 수행하는 분석방법임.
- If A, then B. Else C 형식의 규칙으로 표현

의사결정나무 모델의 특징

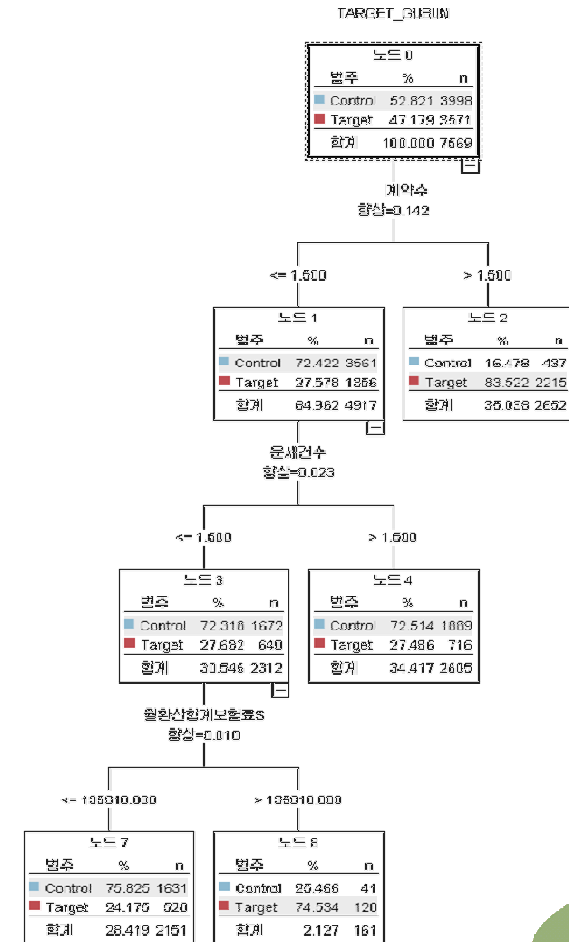
- **해석의 용이성** : 나무구조에 의해 모델이 표현되므로 사용자가 이해하기 쉬움. 그러나 나무의 깊이가 깊어져 조건이 많아지면 가독성이 떨어지는 문제점이 있음.
- **교호효과**의 해석 : 유용한 입력변수나 교호효과를 자동으로 찾아내는 알고리즘.
- **비모수적 모델** : 선형성, 정규성 또는 등분산성 등의 가정을 필요로 하지 않는 비모수적 방법이며, 연속형 변수의 이상치에 민감하지 않음.

의사결정나무 방법론 (C&RT 알고리즘)

- 이지분리를 하는 의사결정나무 모델 알고리즘으로 부모 노드에서 자식 노드로 분리를 할 때 항상 2개 미만 분리를 함.
- 분류가 비교적 간단하여 해석이 편하다는 장점을 가지고 있음.
- Gini 지수를 이용하여 데이터의 불순도를 측정하여 분류를 수행함. (Gini 지수가 작을 수록 Good!)

$$Gini\ 지수 = 1 - \sum_{l=1}^k p_l^2$$

(p_l : 범주의 비율, k : 범주의 개수)



모델 개발 - 모델링 기법(2)

- ❖ 평점표 모델(Score Card Model)은 각 변수의 구간화 후 구간의 점수의 가중치를 부여하는 방식으로 비즈니스 룰을 반영하기 용이한 예측 기법

평점표 모델(Score Card Model)

- 각 변수의 구간화 후 구간의 점수의 가중치를 부여하는 방식으로 개발되는 예측 기법
- 각 단일 변수의 구간별 점수를 합산하는 방법

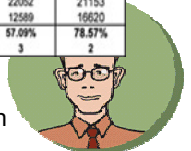
평점표 모델의 특징

- 해석의 용이성 : 평가 항목에 대한 점수와 가중치를 통해 최종 점수에 대한 해석 용이함.
- 비즈니스 룰 반영 용이 : 업무에 사용되던 규칙 및 중요 변수에 대한 가중치를 반영하기 용이함.
- 교호효과 반영이 어려움 : 각 변수간의 상호작용을 반영하기 어려움.

평점표 방법론

- 변수 선택 및 가중치 정의
데이터 탐색을 통해서 분포 차이비율과 비즈니스 정의에 의한 변수 선택 및 가중치 정의
- 변수의 구간(항목)화 및 점수화
 1. 각 변수 항목별 구성비에 대한 Control 대비 Target 비율 산출
→ Target/Control 비율산출
 2. 1번의 최대값이 100%가 되도록 보정
→ 관측값/최대값
 3. 각 변수의 가중치를 사용하여 항목별 점수로 최종 점수 산출
- 최종 점수의 표준화

Calling Module	1A	1B	2A	2B	3A	3B
Service Provider	Carrier A	Carrier B	Carrier C	Carrier D	Carrier E	Carrier F
Technology	GSM	CDMA	GSM	CDMA	GSM	IDEN
Frequency	850	1900	1900	1900	850	800
Phone Type	Nokia 6230	Nokia 6230	Samsung A680	Nokia 6230	Motorola i205	Kyocera KK5
COMPOSITE SCORE (+/- 7.52%)	70.00%	31.68%	39.24%	25.77%	46.13%	61.28%
COMPOSITE SCORE RANKING	1	5	4	6	3	2
Call Attempts	349	349	335	356	339	368
Successful Initiations	348	327	327	313	338	326
% Successful Initiations	99.71%	93.70%	97.61%	87.92%	99.71%	88.59%
Home Samples	21727	4874	12043	18173	20253	21153
% Home Samples	98.52%	23.07%	54.60%	87.82%	91.81%	100.00%
Analog Initiations	0	0	0	0	0	0
% Analog Initiations	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Completed Calls	343	305	316	296	329	313
% Completed Calls	98.56%	93.27%	96.64%	94.57%	97.34%	96.01%
Dropped Calls	5	22	11	17	9	13
% Dropped Calls	1.44%	6.73%	3.36%	5.43%	2.66%	3.99%
RETAINABILITY (+/- 0.87%)	98.54%	93.27%	96.64%	94.57%	97.34%	96.01%
RETAINABILITY RANKING	1	6	3	5	2	4
Failed Initiations	1	22	8	43	1	42
% Failed Initiations	0.29%	6.30%	2.39%	12.08%	0.29%	11.41%
Reorder	0	0	0	0	0	0
% Reorder	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Fast Busy	0	0	0	0	0	0
% Fast Busy	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
No Service	1	11	4	33	0	35
% No Service	0.29%	3.15%	1.19%	9.27%	0.00%	9.51%
Others	0	0	0	0	0	0
% Others	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Failed Access	0	11	4	10	1	7
% Failed Access	0.00%	3.15%	1.19%	2.81%	0.29%	1.90%
ACCESSIBILITY (+/- 2.44%)	99.71%	93.70%	97.61%	87.92%	99.71%	88.59%
ACCESSIBILITY RANKING	1	4	3	6	2	5
MOS Measured By Mobile End (Downlink)						
Mean Downlink MOS Score	3.62	3.59	3.64	3.43	3.57	3.57
Number of Samples	1626	1500	1499	1427	1547	1554
Number of Samples >= 3.25	1555	1470	1419	1350	1435	1465
% of Samples >= 3.25	95.63%	98.00%	94.66%	94.60%	92.70%	94.27%
MOS Measured By Fixed End (Uplink)						
Mean Uplink MOS Score	3.44	3.14	3.63	3.22	3.32	3.42
Number of Samples	1625	1493	1517	1439	1596	1552
Number of Samples >= 3.25	1302	926	1423	994	1182	1383
% of Samples >= 3.25	80.12%	62.02%	93.80%	69.08%	74.06%	89.11%
Combined MOS						
Mean Combined MOS Score	3.53	3.36	3.64	3.33	3.44	3.49
Number of Samples	3251	2993	3016	2866	3143	3106
Number of Samples >= 3.25	2857	2396	2842	2344	2617	2848
% of Samples >= 3.25	87.88%	80.05%	94.23%	81.79%	83.26%	91.69%
VOICE QUALITY (+/- 2.61%)	87.88%	80.05%	94.23%	81.79%	83.26%	91.69%
VOICE QUALITY RANKING	3	6	1	5	4	2
Receive Signal Strength						
Mean RSSI (dBm)	-72.28	-86.65	-85.96	-88.17	-83.83	-74.09
Number of Samples	22053	19999	21990	19446	22052	21153
Number of Samples > -85	17872	9056	9709	7368	12589	16620
SIGNAL STRENGTH (+/- 8.56%)	81.04%	45.28%	44.15%	37.89%	57.09%	78.57%
SIGNAL STRENGTH RANKING	1	4	5	6	3	2



모델 개발 - 모델링 기법(3)

- ❖ 혼합 모델(Hybrid Model)은 2개 이상의 예측모델을 혼합하여 사용하는 기법으로 각 모델의 단점을 상호 보완하여 예측력을 향상시킬 수 있음

혼합 모델(Hybrid Model)

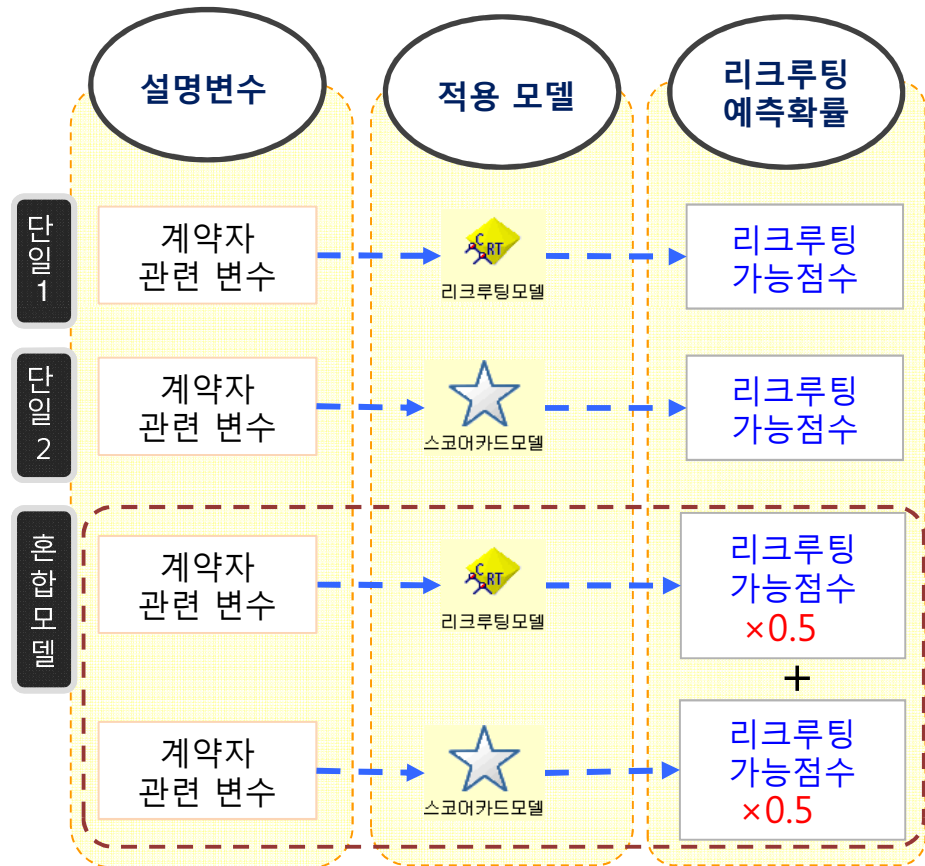
- 더 정확한 예측을 위해 두 개 이상의 다른 모델링 기법을 혼합하여 사용하는 기법
- 의사결정 나무와 평점표를 혼합하여 사용

혼합 모델의 특징

- 상호 보완적 : 상이한 모델링 기법을 사용 혼합하기 때문에 각 모델의 단점을 상호 보완할 수 있음
- 향상된 예측력: 다수의 모델로부터 예측값을 결합시킴으로써 전반적인 정확도를 올릴 수 있음.

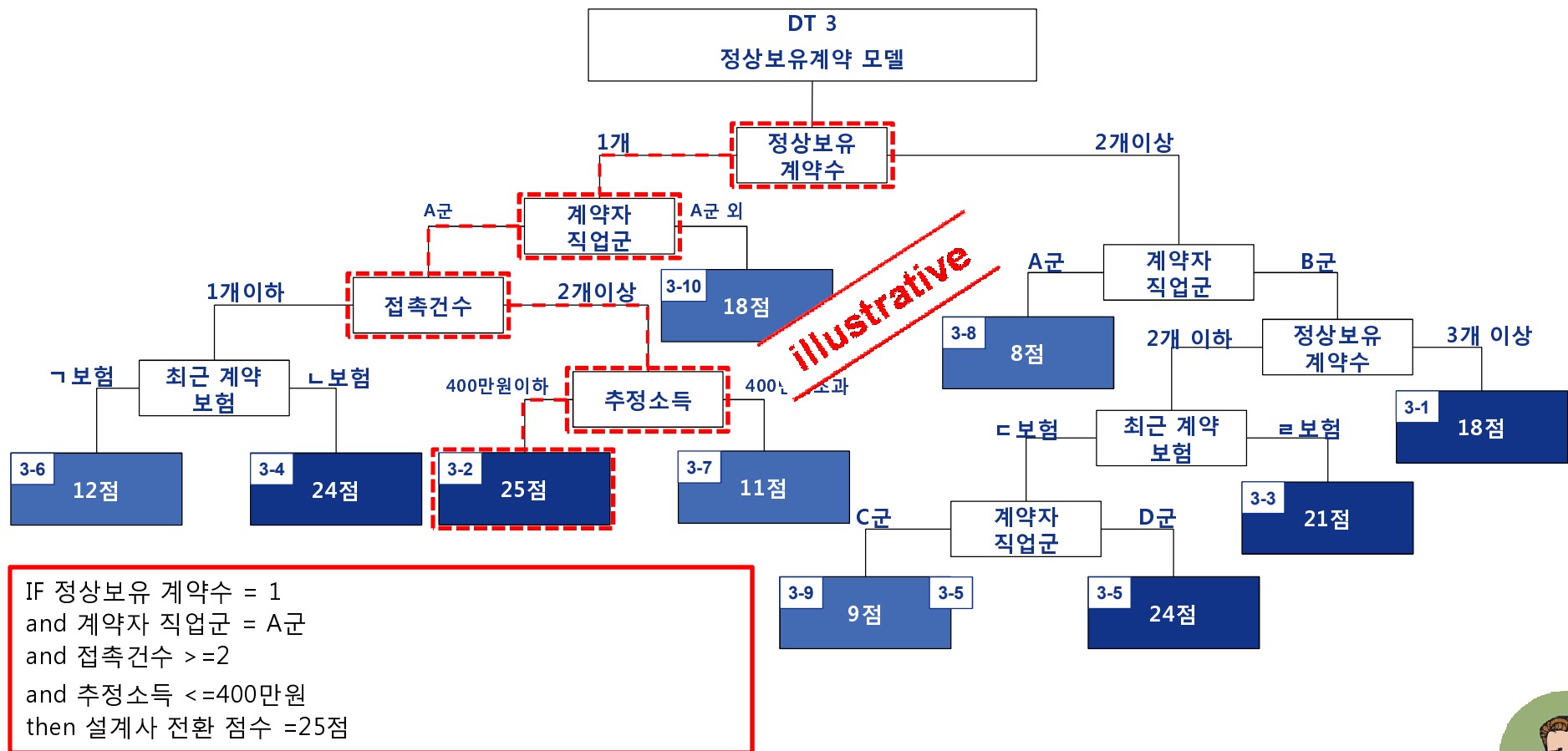
혼합 모델의 방법론

- 연속형 목표변수에서는 점수의 평균을 이용함.
- 범주형 목표변수일 경우 voting을 사용하여 예측값이 얼마나 많이 나왔는가를 비교선택
- Confidence와 Propensity 값을 기초로 가중치를 주거나, Winning Value를 사용함.



모델 개발 - 의사결정나무(결과)

❖ 의사결정나무 분류를 통해 각 고객의 보험설계사 전환 가능 점수를 스코어링 함

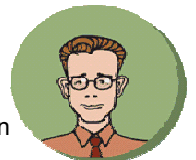


모델 개발 - 평점표모델(결과)

❖ 통계적인 기법을 사용하여 영향력이 큰 변수를 선택하고 구간화하여 점수화 함

변수	항목(구간)	Target분포	Control분포	항목(구간)별 가중치	표준화	최종점수
월환산 합계보험료 (최고점:5)	10만원이하	31.2%	31.7%	0.99	0.69	3.46
	11~20만원	23.7%	30.1%	0.79	0.55	2.76
	21~30만원	14.3%	14.9%	0.96	0.68	3.38
	31~40만원	7.4%	5.6%	1.33	0.93	4.67
	41~50만원	6.0%	5.6%	1.08	0.76	3.80
	50만원 이상	17.3%	12.2%	1.42	1.00	5.00
총 계약수 (최고점:5)	1개	37.4%	57.9%	0.65	0.15	0.74
	2개	6.1%	10.9%	0.30	0.07	0.34
	3개	2.7%	1.9%	0.30	0.07	0.35
	4개	17.3%	4.6%	3.77	0.86	4.33
	5개 이상	36.6%	8.3%	4.39	1.00	5.00
정상보유계약수 (최고점:5)	1개	51.1%	72.4%	0.71	0.25	1.27
	2개	24.5%	17.1%	1.43	0.52	2.58
	3개	11.4%	5.6%	2.04	0.73	3.67
	4개	5.9%	2.3%	2.53	0.91	4.56
	5개 이상	7.1%	2.5%	2.77	1.00	5.00

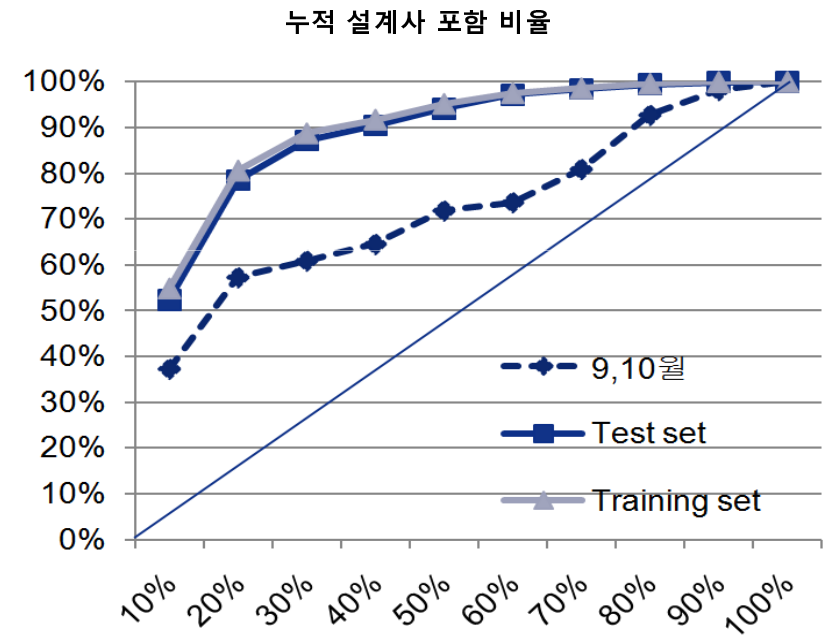
illustrative



모델 평가

❖ 예측 모델을 통해 선발된 상위 30%를 대상으로 리크루팅을 진행 할 경우 전체 고객에서 설계사로 전환하는 설계사의 약 61%를 포함함

등급	보유고객 수	Target	Target의 누적건수	설계사 포함비율	누적설계사 포함비율	lift	누적lift
10%	127,671	41	41	0.032%	37.27%	3.70	3.70
20%	205,027	22	63	0.011%	57.27%	1.24	2.18
30%	54,212	4	67	0.007%	60.91%	0.85	2.00
40%	120,352	4	71	0.003%	64.55%	0.38	1.61
50%	138,484	8	79	0.006%	71.82%	0.67	1.41
60%	126,882	2	81	0.002%	73.64%	0.18	1.21
70%	120,717	8	89	0.007%	80.91%	0.76	1.15
80%	134,556	13	102	0.010%	92.73%	1.11	1.14
90%	119,465	6	108	0.005%	98.18%	0.58	1.08
100%	119,971	2	110	0.002%	100%	0.19	1.00
전체	1,267,337	110		0.009%		1.00	



Summary

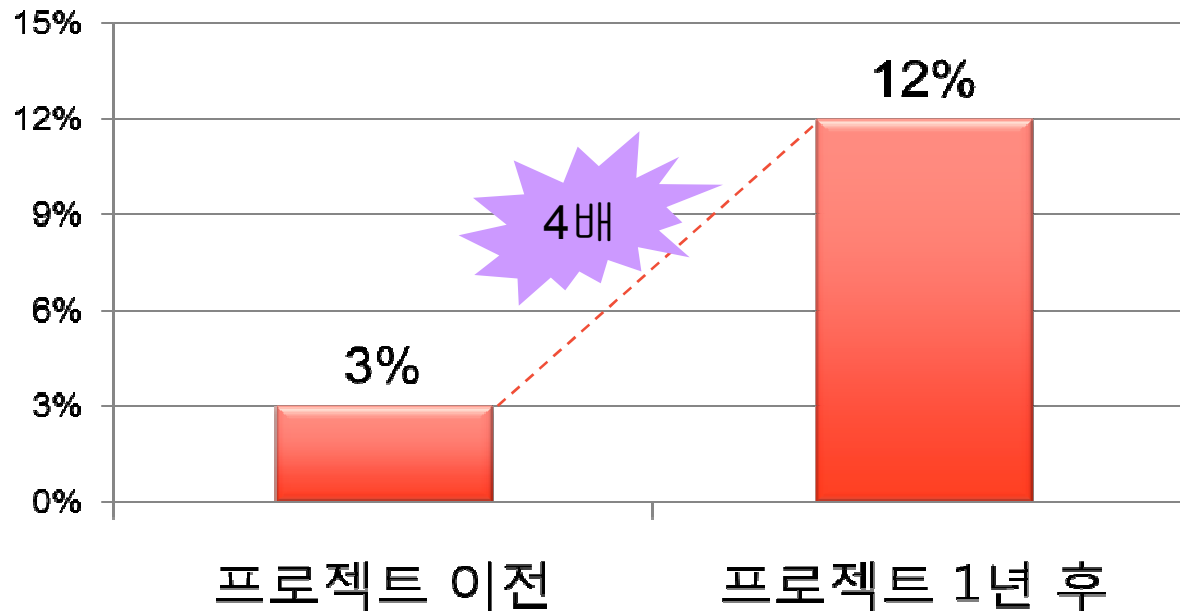
- ❖ 모델 스코어 상위 30% 대상의 리크루팅 성공률(Advisor전환률)은 전체 성공률 보다 2배 높음.
- ❖ 스코어 상위 30%에 Advisor로 전환한 고객이 전체의 60.91%



모델 전개 - 1년간 운영 결과

- ❖ 예측 모델을 사용해 1년간 기존 고객에서 설계사로 전환한 고객의 비율을 평가해본 결과 프로젝트 이전보다 4배 향상된 결과를 얻음

고객에서 설계사로 전환한 고객 비율



사 족



첨언1	애널리틱스와 유관부서와 협업이 중요
첨언2	잘 만드는 것도 중요하지만 지속적 운영이 중요
첨언3	애널리틱스의 벽을 낮출 수 있는 환경
첨언4	비용 대비 효과도 고려



감사합니다!

