

2ith POSCO ^^7

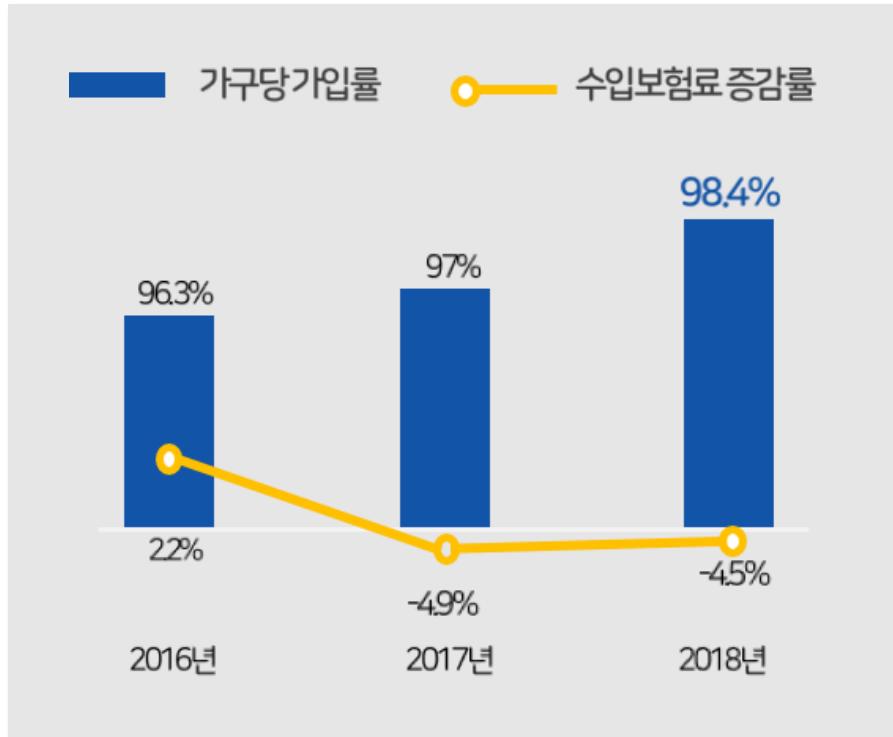
고객 특성 기반 가입심사 프로세스 개선 및
챗봇 보험 추천 서비스를 통한 수익률 증대 방안

A2 정현준 권진수 김형우 이상규 이상래 임우섭



신규고객 유입의 어려움으로 수익률이 감소하는 포빅생명

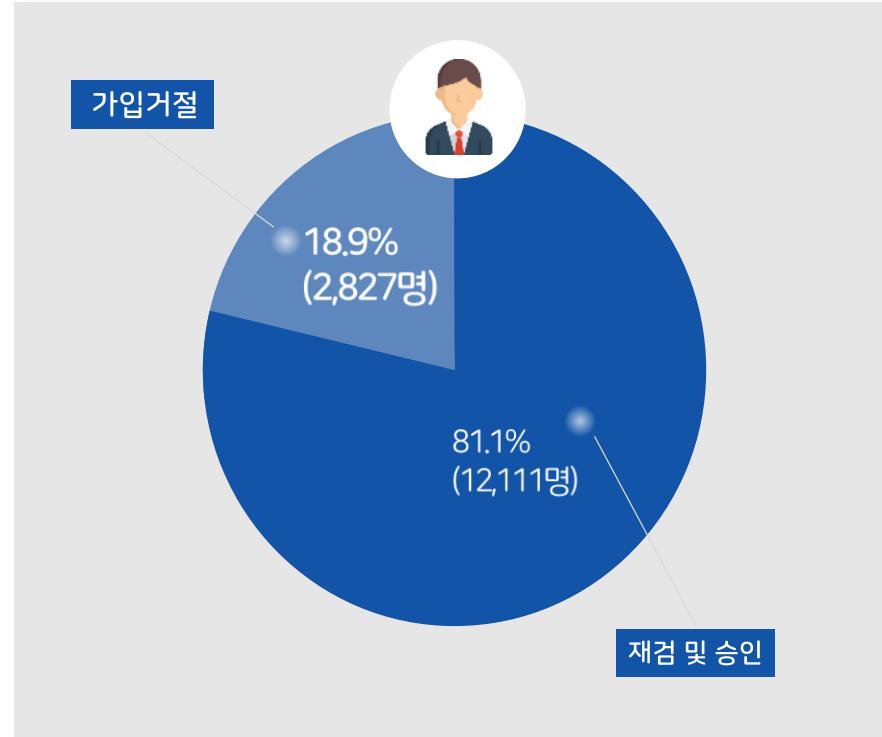
신규 고객으로부터 수익창출의 기회 감소



출처: 보험회사실적통계, (주)포빅생명

포화상태에 다다른 보험가입율
계속적인 감소율을 보이는 수입보험료 증감율

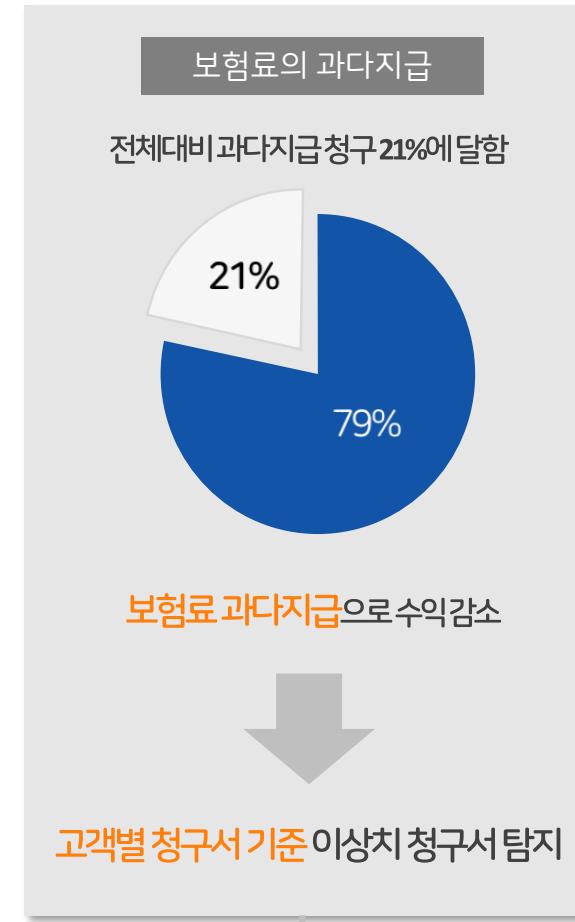
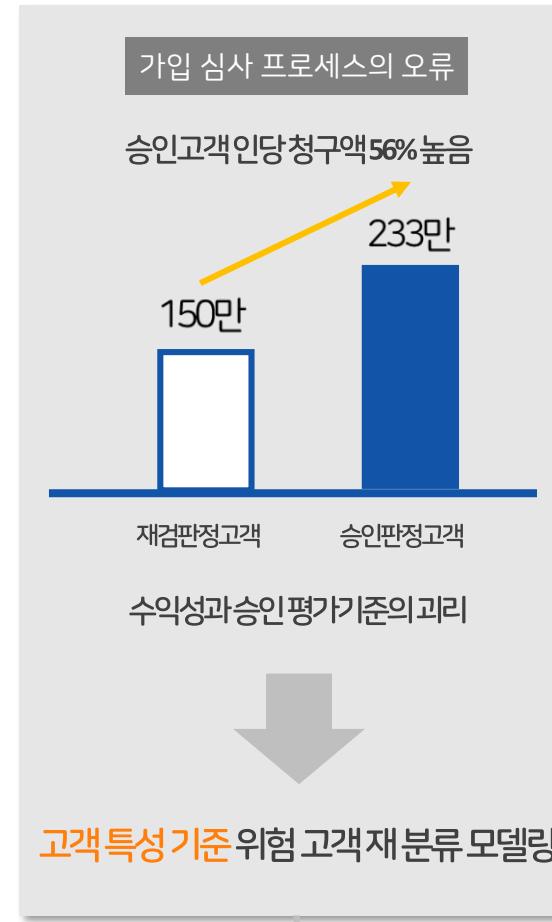
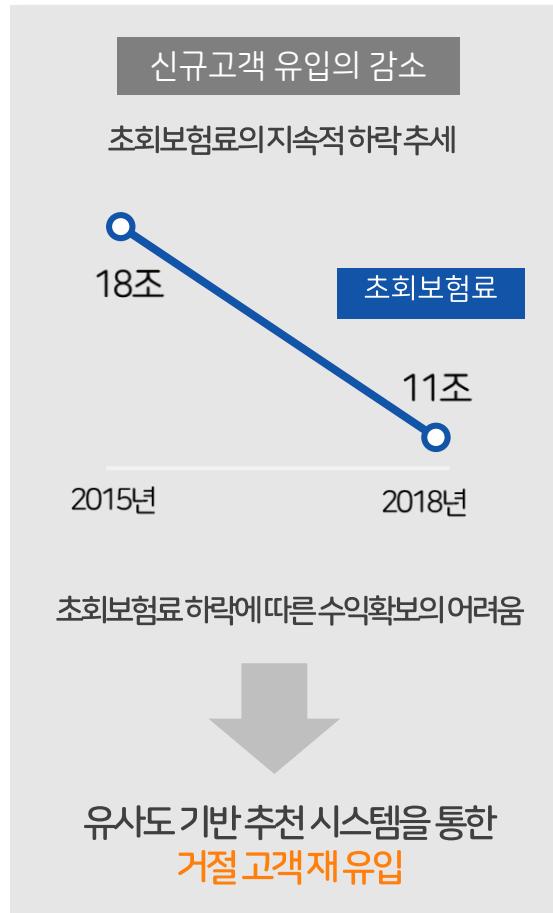
18.9%의 가입 거절로 신규고객의 유입이 어려운 포빅생명



출처: (주)포빅생명

전체 검사자 14,938명
12,111명 재검 및 승인
2,827명 가입 거절

개선기회 도출에 따라 수익성 향상을 위한 방향성 설정



수익성 향상

데이터 분석 및 데이터 간 연관성 파악을 통한 분석계획 설정

데이터 분석 계획 도출

데이터	데이터 개수	정제 후 데이터 개수	파생변수
보험가입 사전승인 검진 정보	14,938개	14,932개	FAT, BMI
보험 청구/지급 정보	49,449개	49,397개	Margin, Diff, Target
국민건강 검진결과	177,346개	175,614개	FAT,BMI

분석 목적	활용기법	분석 내용
변수별 분포 및 특성 파악	Bar Graph Histogram Boxplot Scatterplot	<ul style="list-style-type: none"> 결측치 정상 여부 판단 및 처리 의료 데이터 정상 범주 기반 이상치 처리 보험 상품 별 보장 질병 범위 확인
수익성 지표 수립 및 변수와의 연관성 파악	그룹별 기술 통계량 분석 연관규칙 분석	<ul style="list-style-type: none"> 고객위험도의 기준이 되는 수익성 지표 수립 가입 심사 거절 고객 재추천을 위한 수익성 우수 고객 군집화 질병 별, 고객 별, 보험 상품 별 수익률 분석
	그룹별 통계 분석	<ul style="list-style-type: none"> 검사결과 예측력 향상, 고객 위험도 예측을 위한 파생 변수 추가
수익성 지표별 고객 위험도 예측 모델 개발	Random Forest Gradient Boosting	<ul style="list-style-type: none"> 수익성 지표 기반 고객 위험도 분류 모델 생성 수익성 지표 기반 보험금 과다지급 분류 모델 생성
신규고객 유입을 위한 모델 개발	추천 시스템	<ul style="list-style-type: none"> 분류 결과 거절된 고객에 대한 상품 재 추천 할 수 있는 모델 생성

수익 예측을 위한 목표변수 설정

목표변수 설정

자사이익

*Margin(자사이익) = 누적납입보험료 - 보험금지급금액
고객별 Margin 계산을 통한 위험도 분석

자사이익

보험금 지급금액

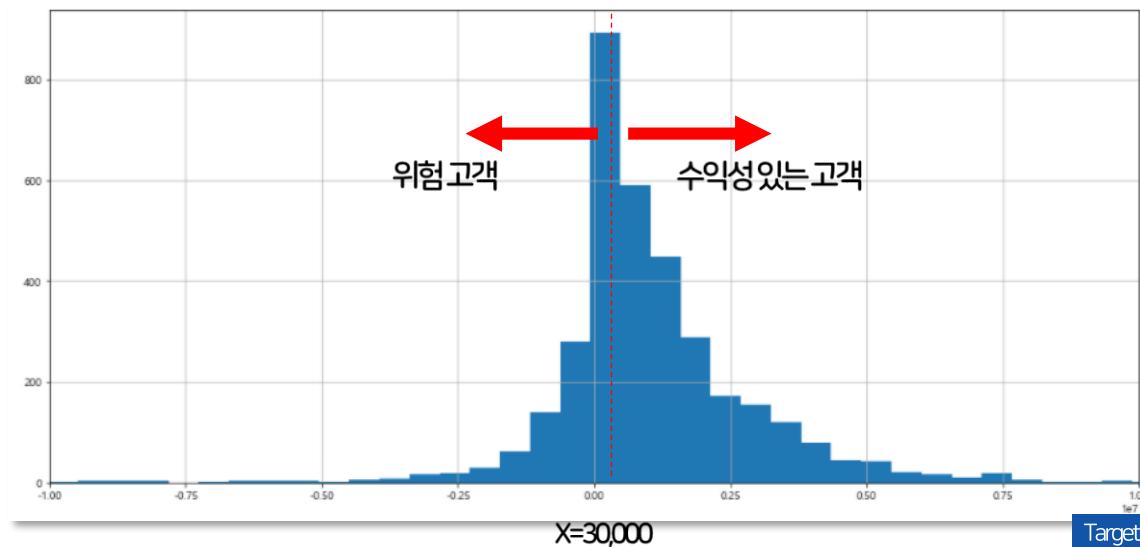
과지급금

*Diff(과지급금) = 보험금 지급금액 - 보험금 청구금액
고객별 Diff 계산을 통한 보험금 청구 이상징후 파악

과지급금

보험금 청구금액

* Target = $(\text{Margin}-\text{Diff}) / 2$ 의 지표를 활용하여 내부 위험도 판정 [프로세스 개선](#)



자사이익 및 과지급금 기준 고객 위험도 분류

분류 결과 및 분포 확인

분류 결과		예측
실제	0	1
0	819	23
1	213	10

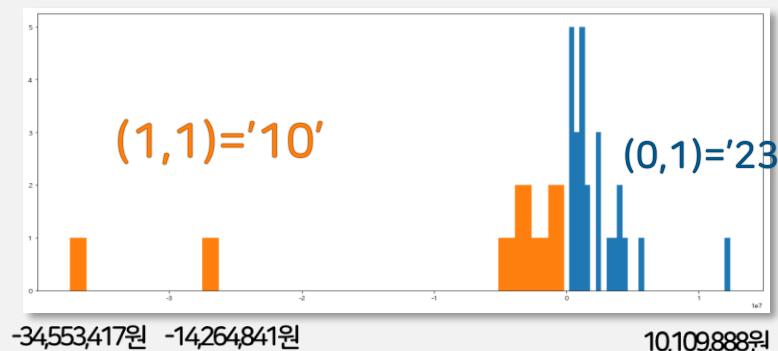
목표변수 : Target

(자사이익-과지급금)

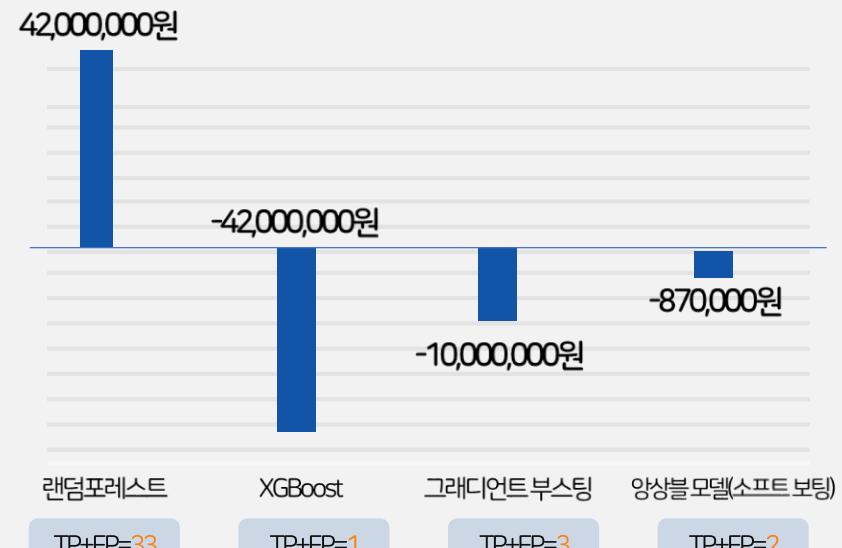
보험금청구데이터 및 가입검진정보 사용
일반검진데이터(9종)

혈액검사데이터(3종)을 이용한 분류 진행
가입거절고객에 대한 추천시스템 서비스 제공

<분류 데이터의 마진분포>



개선된 프로세스의 기대 수익



편향성이 심하고 편차가 큰 데이터 세트으로 인한 오류

가장 성능이 좋았던 **랜덤 포레스트 모델**을 통한 고객 재분류 프로세스를 진행하여
약 **42,000,000원**(총 수익 **21억 중 2%**)의 지출 감소를 통한 영업이익 상승 기대

과지급금 기준 청구 위험도 분류

분류 결과 및 분포 확인

분류 결과		예측
실제	0	1
0	2762	122
1	386	100

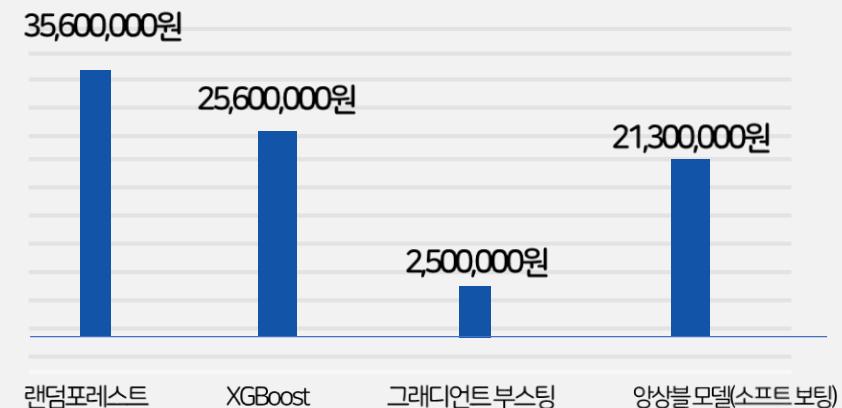
목표변수 : Diff(과지급금)

과지급금(보험사지급액-고객청구액)
기준의과다지급을분류
청구된고객의일반검진데이터9종및
혈액검사데이터3종활용

<분류데이터의과지급금 분포>



개선된 프로세스의 기대 수익



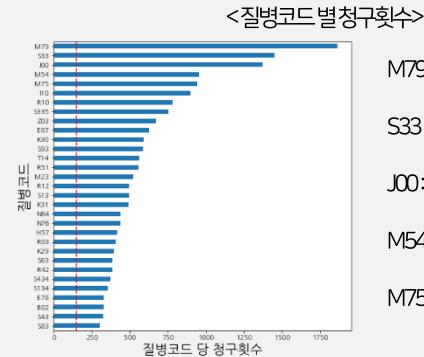
가장 성능이 좋았던 **랜덤 포레스트 모델**을 통한 청구서 과다지급여부 확인 절차를 진행하여
약 **35,600,000원** (총 과지급금 3.5억 중 10%)의 지출 감소를 통한 영업이익 상승 기대

심사기준 변경에 따른 재유입 고객 관리 방안

보험상품 추천시스템 시안

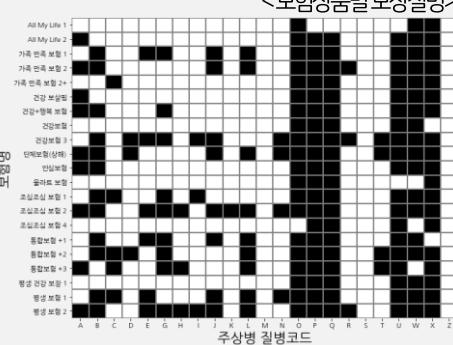
The image shows a mobile application interface for messaging. At the top, there is a profile picture of a man with glasses and a beard, followed by the text "포빅생명" and "당신 곁의 든든한 친구". To the right of the text is a grey star icon. Below this, a blue rectangular box contains the identifier "C134240". The main conversation area consists of two green message bubbles. The first bubble contains the text "C134240님의 이력으로 추천합니다. 질병 코드를 입력해주세요." The second bubble contains the text "입력 시간, Today me". Below this, another blue rectangular box contains the identifier "S33". Further down, another green message bubble contains the text "포빅생명 챗봇" followed by "입력 시간, Today". The final green message bubble at the bottom contains the text "고객님과 유사한 고객님은 다음과 같은 상품을 기임했습니다. [울리트 보험·'건강보험·'조심 조심 보험 4']". At the very bottom right, there is a large, bold, white text box containing the text "<챗봇추천예상사례>".

1 빈발질병코드 선택지 제공



- M79:근골격계질환
- S33:요추관련질환
- J00:급성비인두염(감기)
- M54:경추등등관련통증
- M75:아크과관련질병

2 해당 질병을 포함한 고수익 유사상품 추천



3 기존고객 과의 유사도 산정 및 추천 상품 도출

	추천상품명	유사도
1	All My Life 2	0.015
2	율라트 보험	0.015
3	율라트 보험	0.015
...
15	율라트 보험	0.014

유사한상위 15명중 60%가 울라트보험가입

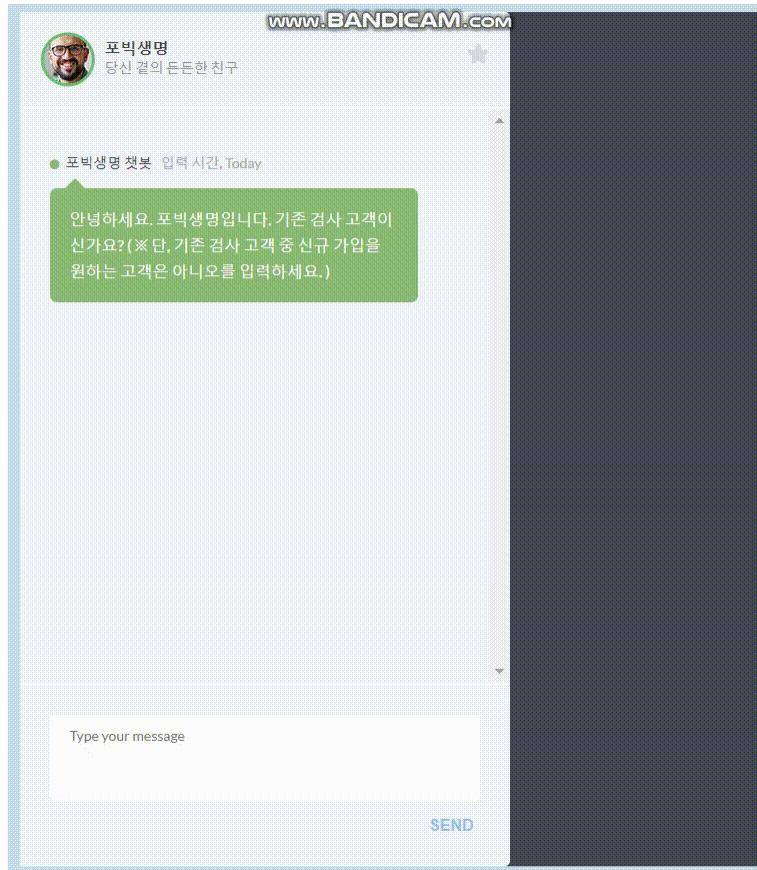
4 각상품 시나리오별 수익성 도출 및 우선 추천

	추천상품명	기대수익
1	울라트 보험	4,200,000
2	건강보험	1,900,000
3	건강 보살핌	630,000

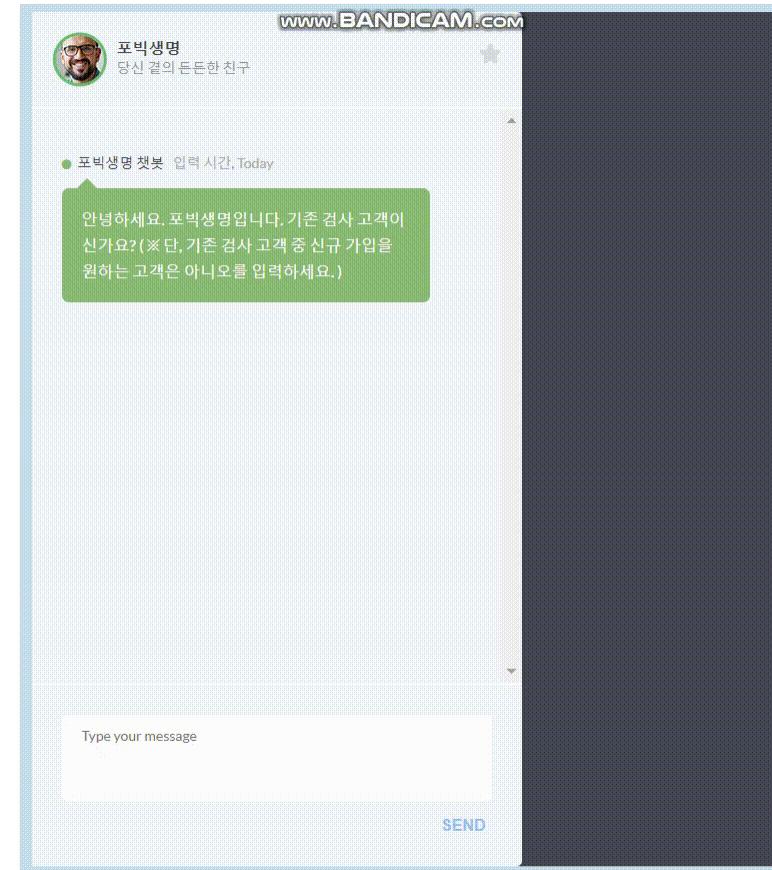
보험가입시기대수익이가장높은울라트보험우선추천

재유입 고객의 선택 질병을 기반으로 유사 고객의 고수익 가입 상품을 추천

신규고객



기존고객



최종 모델 성능 평가

챗봇 보험상품 추천 서비스 기대값

1 기준 고객 데이터 기반 추천 결과 도출

	고객	가입상품	추천 상품
1	C112358	All My Life 1	[건강보험, 조심조심 보험 4, 올라트 보험]
2	C117222	올라트 보험	[건강보험, 조심조심 보험 4, 올라트 보험]
3	C127104	가족 만족 보험 2+	[건강보험, 올라트 보험, 평생 건강 보장 1]
...			...
891	C117283	조심조심 보험 4	[조심조심 보험 4, 올라트 보험, 건강보험]

가입상품이 추천상품에 포함될 경우 1, 아닐 경우 0으로 산정

2 추천 시스템 예측 정확도 산정

	고객	예측 성공 여부
1	C112358	0
2	C117222	1
...		...
891	C117283	1

예측 정확도 =
예측 성공 여부 값의 평균값

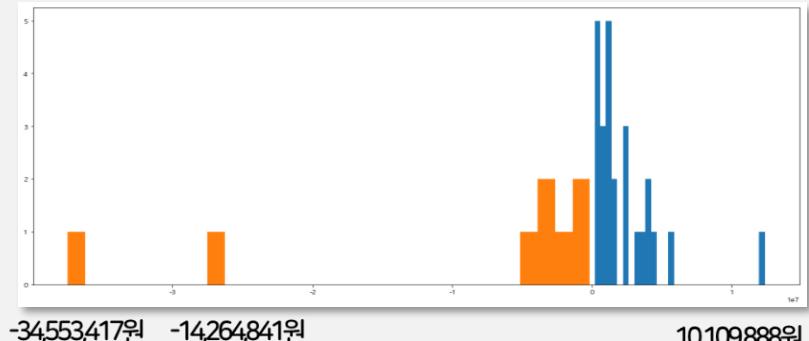


평균값 = 0.534231

예측 정확도는 0.534%로 판단됨

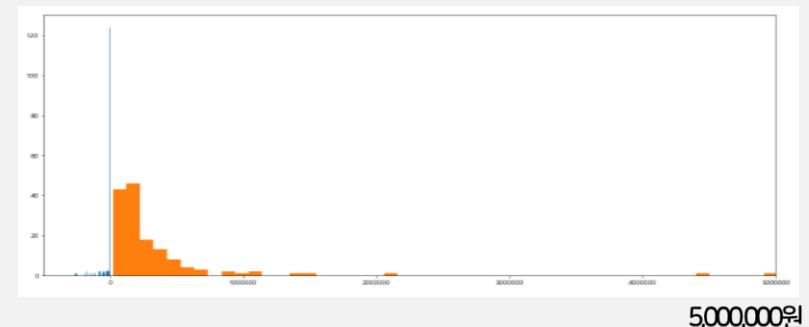
분류 데이터 모델별 기대값

<마진 기준 고객 분류 모델>



약 42,000,000원 (총 수익 21억 중 2%)

<과자급금 기준 청구서 기준 분류 모델 분류 모델> 분포

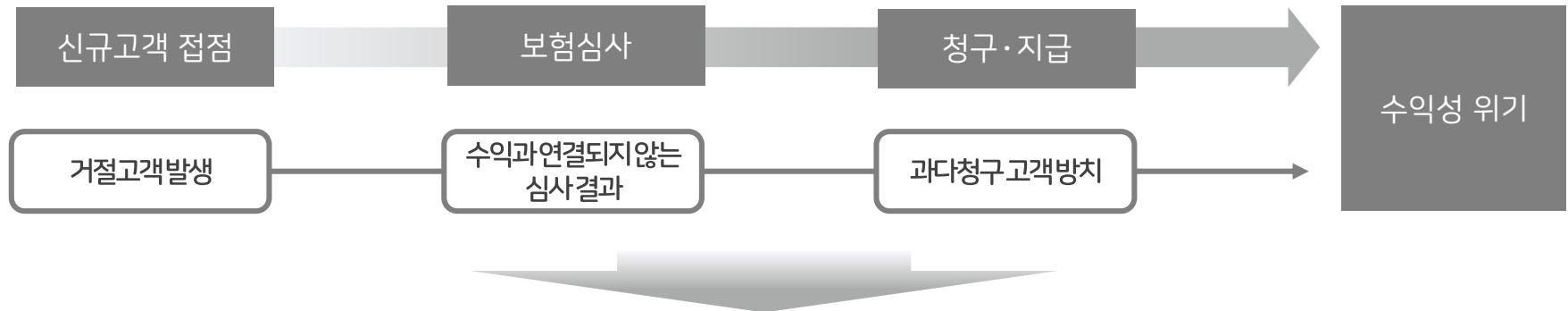


약 35,600,000원 (총 과자급금 3.5억 중 10%)

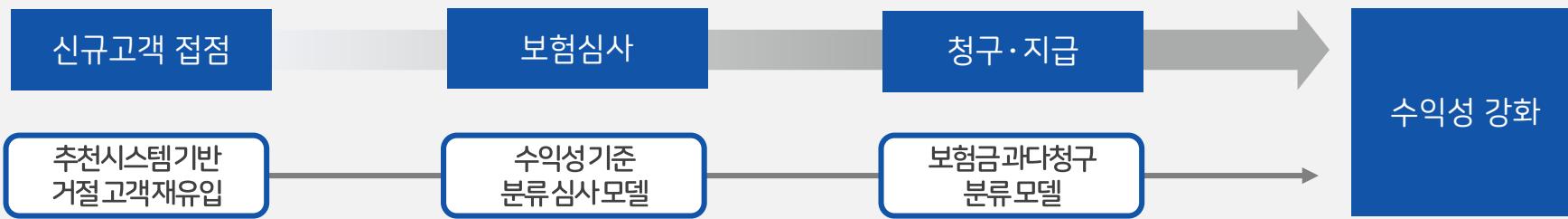
정확도 높은 추천 시스템과 분류 모델의 수익 창출력을 기반으로 포빅생명의 수익 모델 확보

포빅생명의 영업 프로세스별 개선기회 기반 잠재수익 확보

AS-IS 프로세스



TO-BE 프로세스



오분류된 거절 고객 재유입 및 기존 고객 수익성 강화를 통해 포빅생명의 각 프로세스별 잠재 수익 확보

정현준

극한의 상황까지 스스로를 몰아붙이면서 내가 어디까지 할 수 있을지 시험해볼 수 있었다. 덕분에 내가 부족한 부분도 느끼게 됐고 실력이 좋은 팀원들에게 많이 배우며 실력을 키울 수 있었다. 빅데이터 프로젝트를 했던 기억을 떠올리며 이번 프로젝트에 적용해서 한결 수월했는데 다음에 다른 프로젝트를 할 때 이번 프로젝트의 경험이 큰 자산이 될 것이라는 느낌이 들었다.

권진수

개발 경험이 부족해 걱정됐지만, 모르는 부분은 어떻게 해서라도 알려주고 함께 해결해나가는 방법을 찾아내는 팀원들에게 정말 많이 배웠다. 그리고 머릿 속에서 산재 돼 있던 분석 이론들을 실제 활용해보는 과정에서 성장했음을 느꼈다. 열정 넘치는 팀원들과 함께했기에 가능했다고 생각한다.

이상규

한 집에 모여 긴 테이블에 모여 앉아 채팅이 아닌 말로 주고 받으며 프로젝트를 했던 경험은 잊을 수 없을 것 같다. 인턴 생활을 했던 때가 떠오르기도 했고, 이야기를 나누며 개발을 했던 시간들이 마치 스타트업의 실무 환경을 보는 것 같았다. 개인적으로 잘하고 싶은 마음이 컸기 때문에 결과가 잘 나오지 않을 땐 화가 나기도 했지만 팀원들이 있어서 버틸 수 있었다.

이상래

다양한 대외활동과 장교 복무로 어떤 일어든 다 자신 있다고 생각했는데 역시나 익숙하지 않은 컴퓨터 언어로 데이터를 직접 다루는 과정이 쉽지만은 않았다. 전역하고 용사로 재입대한 기분이었다. 몸은 많이 피곤했지만 개그를 잘 받아줬던 팀원들과 함께 프로젝트를 진행할 수 있어서 피곤함을 잊으며 개발을 경험할 수 있어서 좋았다.

임우섭

개발을 배운지 1년, 그래도 어느 정도 개발이 익숙해졌다고 생각했는데 개발의 세계는 넓고도 깊었다. 단순히 개발을 잘 한다고 해서 프로젝트에서 관심은 결과물을 내는 것이 아니라는 것을 배웠다. 다양한 도메인에서 각각의 IT 프로젝트를 수행함에 있어서 도메인에 대한 이해도 사전에 충분히 숙지되어 있어야 함을 느꼈다. 다음에 빅데이터 프로젝트를 하게 된다면 이런 부분을 기억하고 있으면 좋을 것 같다.

김형우

온라인 교육으로 전환되면서 과연 잘 할 수 있을까라는 생각이 들었다. 팀원들과 이야기 해서 서울로 상경해 합숙하면서 프로젝트를 수행한 덕분에 교육장에서 직접 받는 수업에 준하는 효율을 낼 수 있었던 것 같다. 팀원들이다들 열정적이라서 스스로에게도 좋은 자극이 되었다.

Q&A

포빅생명의 수익성에 영향을 미치는 잠재원인 분석 및 중요도 선정

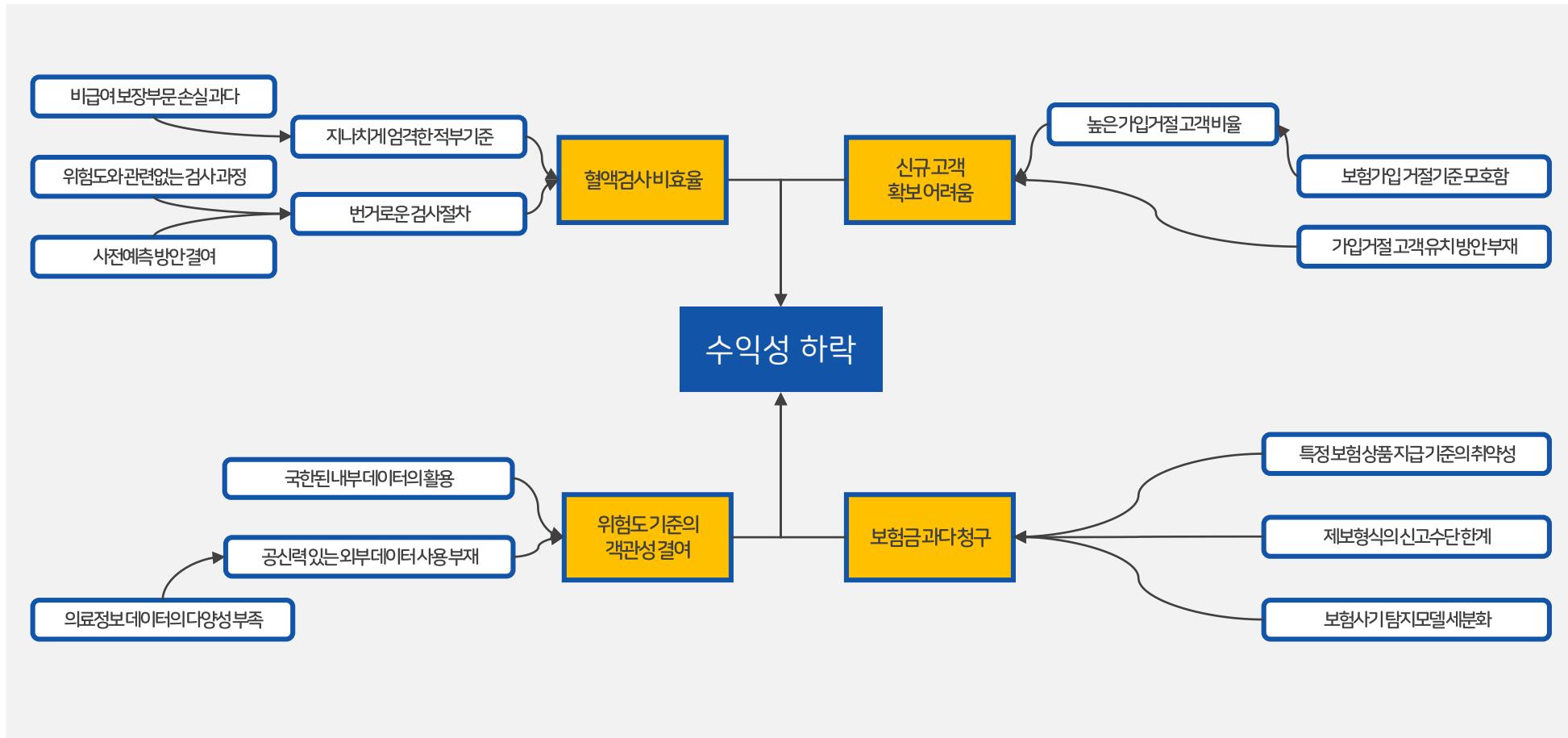
현재 생명보험사 수익창출 프로세스



생명보험사의 보험 영업 **프로세스**를 기반으로 현 문제 상황을 야기하는 Business Task 파악

포빅생명의 수익성에 영향을 미치는 잠재원인 분석 및 중요도 선정

특성 요인도



특성 요인도를 통해 각 개선기회에 대한 9가지 핵심요인을 선정하였음

결측치 및 이상치 탐색 프로세스 확립

이상치 처리 기준 확립



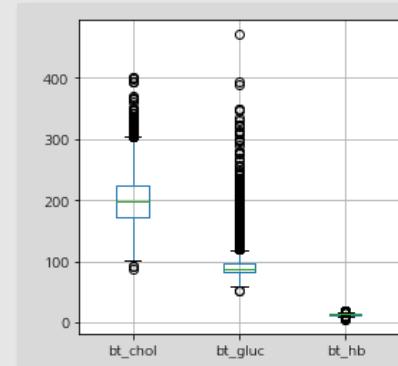
1 의료데이터 정상치

	특성요인	정상수치
1	bt_chol	200~240
2	bt_crea	0.5~1.4
3	bt_gluc	70~126
4	bt_hb	12~18(G/DL)
5	bt_hct	남:40~54 여:38~47
...	...	
16	bt_trig	150~500

2 요약통계량

	height	weight
1	Count	14938
2	Mean	48.702
3	std	10.782
4	Min	15.000
5	25%	42.000
...	...	
8	Max	70.000

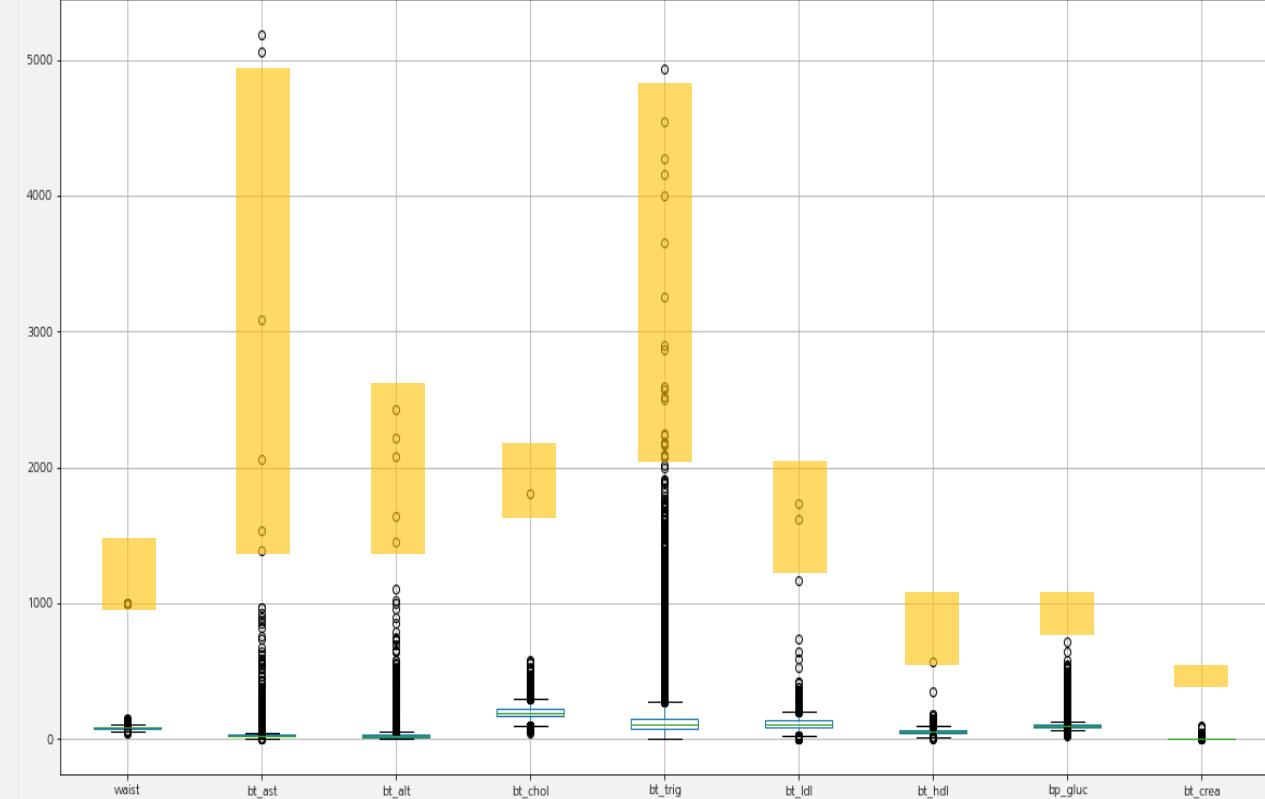
3 데이터 시각적 분포



데이터 셋에 대한 세 가지 기준에 따른 이상치 및 결측치 처리 프로세스를 확립하여 EDA를 위한 데이터를 구성함

건강관리공단 정보_ 이상치 제거 PROCESS

이상치 제거 Box plot



PROCESS

```

df_3 = df_3[df_3['waist']!=999.0]
df_3 = df_3[df_3['bt_ast']<1000]
df_3 = df_3[df_3['bt_alt']<1000]
df_3 = df_3[df_3['bt_chol']<1000]
df_3 = df_3[df_3['bt_trig']<2000]
df_3 = df_3[df_3['bt_ldl']<1000]
df_3 = df_3[df_3['bt_hdl']<300]
df_3 = df_3[df_3['bp_gluc']<500]
df_3 = df_3[df_3['bt_crea']<20]
df_3 = df_3[df_3['eyesight_left']!=9.9]
df_3 = df_3[df_3['eyesight_right']!=9.9]
df_3.reset_index(drop=True,inplace=True)

```

waist: 몸통둘레

ast, alt(간수치):
40~200경도, 200~400중등도, 400이상중증

Rgt(간수치): 남자 8~61, 여자 5~36

chol: 200미만 정상, 200~240 경계, 240위험

trig(중성지방): 150미만 정상, 150~199경계,
200~499심질환위험증가, 500이상매우증가

gluc(식전혈당): 70~99정상,
100~125포도당장애, 126이상당뇨병

ldh(혈색소): 24~30pg정상범주

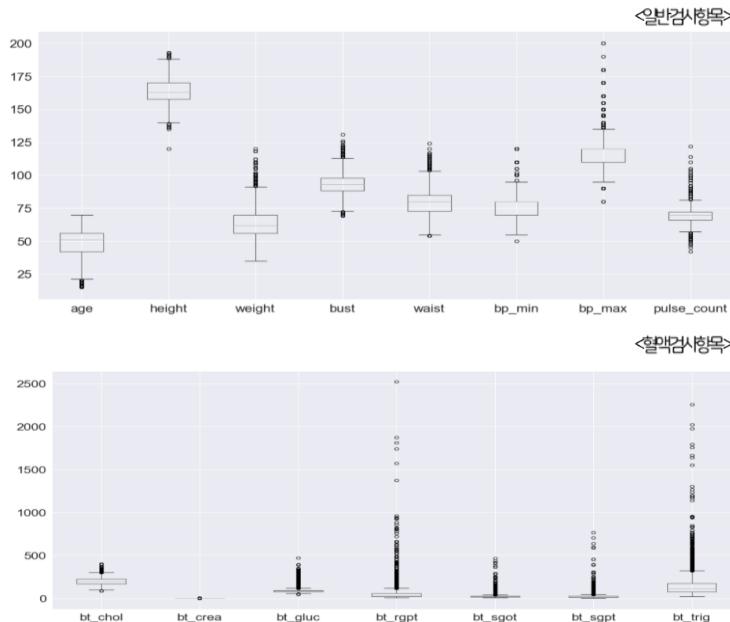
crea(혈청): 0.5~1.4정상범위

eyesight: 시력

대한진단검사의학회의 정상수치에 따라 이상치를 판별

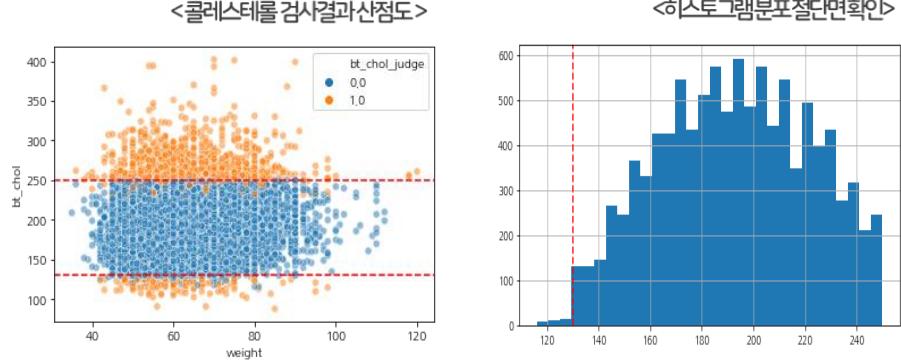
보험가입 사전 승인 정보_데이터 결측치 처리

검사항목의 데이터분포



대부분 연속적인 분포로 별도의 이상치를 확인할 수 없었기에
이상치에 대한 처리는 진행하지 않음

혈액 검사항목의 데이터오류확인



산점도 분포 기반 검사 결과 데이터의 오류 발견

히스토그램 분포 절단면 기반 정상·비정상 기준 재정립
검사 결과 데이터의 기준 확립 및 오류값 정정

보험가입 사전 승인 정보_파생변수 생성

Pre-Review.csv

체중, 성별, 신장, 나이 데이터로 파생변수 생성

체질량지수와 질병 간의 높은 상관관계

- 대한비만학회지: 제 10 권 제 2 호 2001 -

Table 5. Comparison of the Prevalence Rates of Normal, Overweight, Obesity groups

	Normal*	Overweight**	Obesity***	Number (%)	p values†
Hyperlipidemia	153 (22.9)	191 (28.6)	324 (48.5)	0.000	
Respiratory dis.	103 (13.6)	50 (9.7)	59 (8.5)	0.005	
Anemia	71 (9.3)	32 (6.2)	33 (4.8)	0.002	
Liver dis.	60 (7.9)	75 (14.6)	187 (26.9)	0.000	
DM	45 (5.9)	52 (10.1)	121 (17.4)	0.000	
Abnormal EKG	112 (14.7)	89 (17.3)	127 (18.3)	0.172	
Hypertension	66 (8.7)	85 (16.5)	180 (25.9)	0.000	
GI dis.	307 (40.4)	241 (46.9)	318 (45.8)	0.038	
Hyperuricemia	57 (7.5)	85 (16.5)	153 (22.0)	0.000	
Renal dis.	165 (21.7)	74 (14.4)	114 (16.4)	0.002	
Morbidity Index	1.24±1.16	1.68±1.24	2.06±1.40	0.000	

*: Normal weight: BMI < 23 kg/m²**: Overweight: 23 kg/m² ≤ BMI < 25 kg/m²***: Obesity: BMI ≥ 25 kg/m²

† p values are from Kruskal Wallis test, chi-square test

심경원, 이상화, 이정수(2001). 체질량지수와 질병이환의 관련성
대한비만학회지BMI와 고지혈증, 고혈압, 고노산혈증 등의 질병과
회귀분석 결과 유의한 p-value를 확인할 수 있었음

Table 5. Regression of body fat percentage (BF%) as dependent variable, and body mass index (BMI), age and sex as independent variables in the populations of adults (age ≥ 16 years)

Group n	Regression coefficients						R ²	
	BMI	Sex*	Age	Intercept	SEE			
	Mean	se	Mean	se	Mean	BF% CV%		
A 356	+1.61 0.10	—	—	-12.1 2.4	7.2 29	0.40		
	+1.63 0.07	-9.8 0.6	—	-8.9 1.8	5.4 22	0.67		
	+1.16 0.06	-10.7 0.5	+0.23 0.01	-5.0 1.4	4.2 17	0.80		
B 393	+1.45 0.09	—	—	-8.1 2.3	7.0 27	0.36		
	+1.63 0.07	-10.0 0.5	—	-8.3 1.7	5.1 20	0.66		
	+1.26 0.06	-10.9 0.4	+0.22 0.01	-5.6 1.3	4.0 16	0.80		
A + B 747	+1.53 0.07	—	—	-10.1 1.6	7.1 28	0.38		
	+1.63 0.05	-9.9 0.4	—	-8.6 1.2	5.2 21	0.67		
	+1.20 0.04	-10.8 0.3	+0.23 0.01	-5.4 1.0	4.1 16	0.79		

SE: standard error of estimate; CV%: coefficient of variation; R²: explained variance.

* Sex: males = 1, females = 0.

PAUL DEURENBERG, JAN A. WESTSTRATE, JAAP C. SEIDELL(1991).
Body mass index as a measure of body fatness: age- and sex-specific regression formulas.
British Journal of Nutrition16세 이상의 747명의 데이터로 예측지행시
설명력(R²)가 0.79로 나타남

BMI 및 체지방률 변수추가

BMI

$$\text{BMI} = \text{WEIGHT}/(\text{HEIGHT}^{**2})(\text{kg}/\text{m}^{**2})$$

체지방률

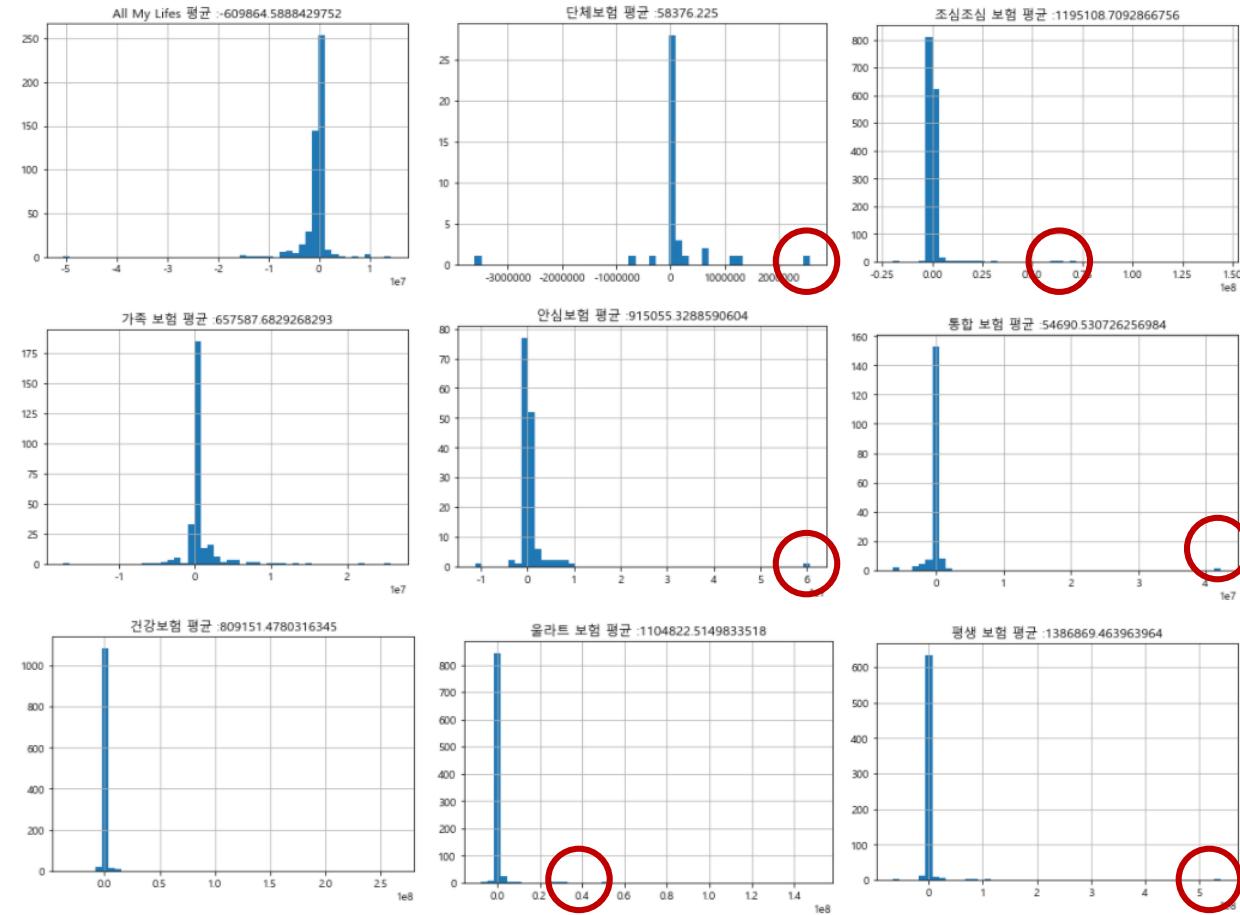
$$\text{FAT} = 1.2 * \text{BMI} + 0.23 * \text{AGE} - 10.8 * \text{GENDER} - 5.4$$



	customer_id	bmi	fat
0	C112346	23.140496	27.668595
1	C112350	23.671254	28.075504
2	C112355	24.158818	28.890582
3	C112356	23.507805	28.109366
4	C112358	26.491508	31.689810

보험 청구 지급 정보_ 상품군별 과지급 금액 분포

보험상품군별 과지급금액(Diff) 분포 확인



21개의 보험상품을 동일한 이름의 상품으로 묶어 9개의 그룹으로 분류 'All My Life' 상품군을 제외한 모든 상품에서 양의 과지급금액(Diff) 확인

이상치로 파악되는 데이터의 지나친 과다보험금 청구로 인한 회사 부담금의 증가가 확인되었음