

# 생명보험 고객 특성 기반 심사 기준 및 절차 개선을 통한 수익성 향상

---

A3

강연주 김예은 김준오 김지한 손혜원 하준범

# 목차

01 추진 배경

02 현상 및 개선 기회

03 분석 계획 및 결과

04 개선안 및 적용방안

05 프로젝트 후기

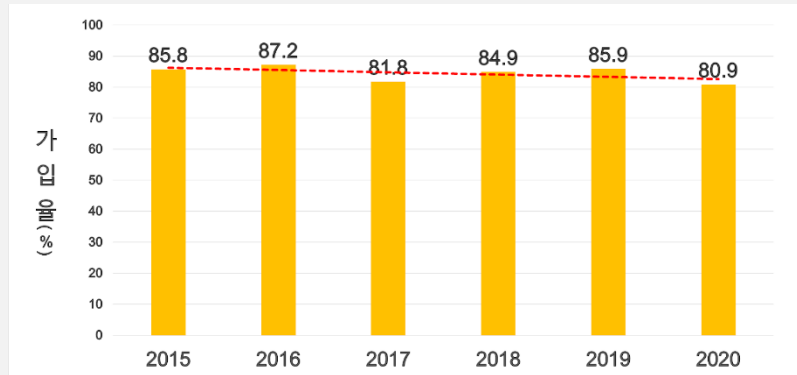
06 Web 시연

# 1. 추진 배경

## 포빅생명 수익성 향상 필요

### 생명보험 시장

<연도별 생명보험 가입률>



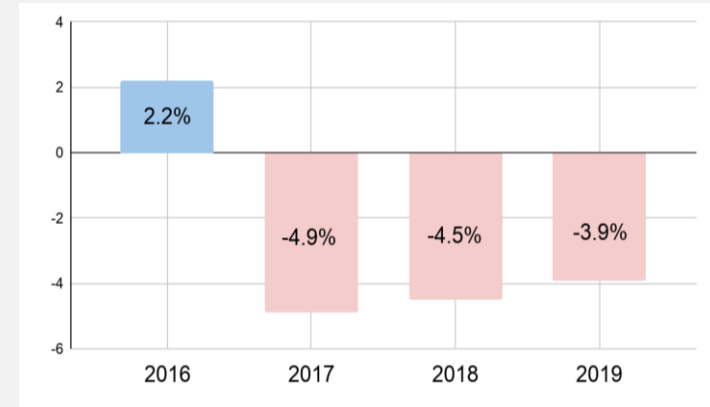
출처: 2021 보험산업 전망과 과제 (보험연구소)

- 가구당 보험 가입률 8-90%로 시장 포화 상태
- 2020년 보험 가입률: 80.9%

**“시장 포화로 신규 고객 확보 어려움”**

### 포빅생명

<포빅생명 수입보험료 증감률>



- 2016년 이후 매년 마이너스 성장률 기록
- 초회보험료\*는 2015년 ~ 2018년 3년 사이 40.5% 급감
- 보험 가입 거절 비율은 18.9%로 높은 편

**“수입보험료 감소 지속”**

\*초회보험료: 신규 가입자가 낸 첫 보험료, 통상적으로 보험사의 영업력을 측정하는 척도로 사용

## 2. 현황 및 개선 기회 수익성 하락 원인 및 해결 방안

<기존 고객>

보험료 과지급 발생



기존 고객 중 납입, 청구 대비  
과다지급 발생 (약 36%)



보험 상품 및 고객 관리  
▶ 수익성 개선

<과거 가입 신청자(거절)>

신규 고객 유치 한계



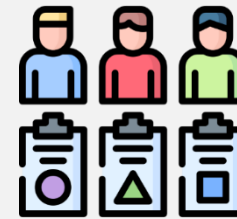
가입 거절 비율이 18.9%로  
신규 고객 유치 어려움 발생



거절 고객 재 유치  
▶ 매출 증대

<신규 가입 신청자>

가입 승인 모델의 오류



고객 위험도 판단의 어려움  
비적합 고객 유입 가능성 발생



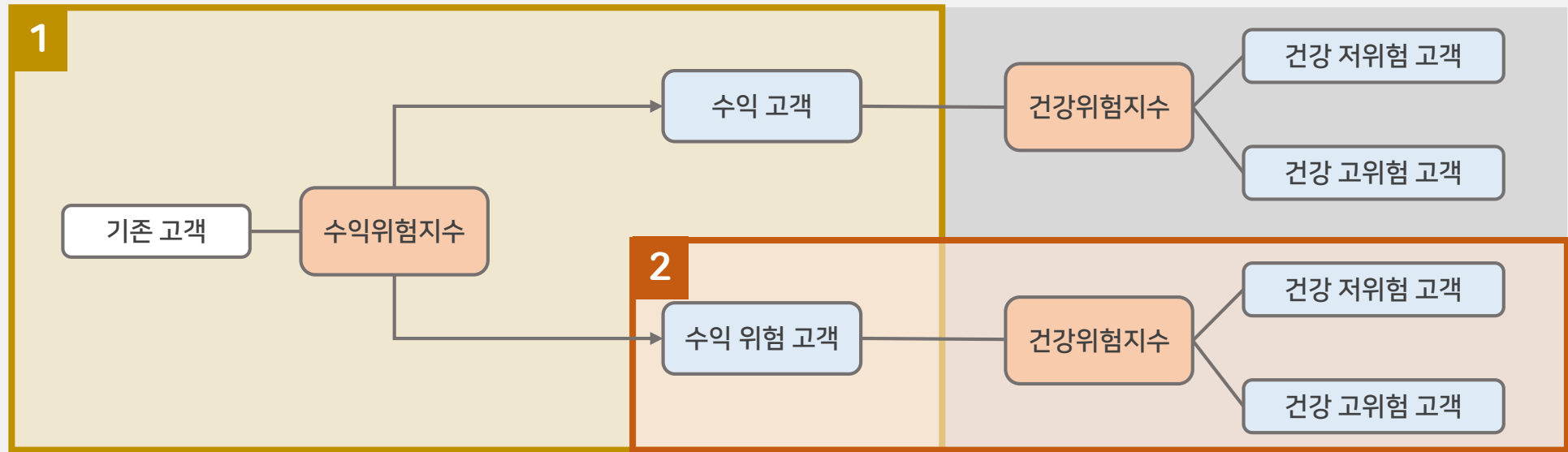
가입 승인 모델 개선  
▶ 리스크 감소 및 비용절감

생명보험심사 기준 및 절차 개선으로 수익률 10% 상승 목표

### 3. 분석 계획 및 결과    데이터 분석 목적 및 내용

목적	데이터	분석 방법	주요 내용
자사 보험 상품 및 고객 관리	<ul style="list-style-type: none"> <li>- 보험금 청구 및 지급 정보</li> <li>- 보험 가입 사전 검진 정보</li> <li>- 국민 건강 검진 정보</li> <li>- 보험 상품 정보</li> </ul>	막대그래프	<ul style="list-style-type: none"> <li>- 보험 상품별 수익 위험 고객 및 건강 위험 고객의 분포 확인</li> <li>- 기존 고객의 성별 및 연령대별 보험 상품 분포 확인</li> </ul>
		Box Plot	전체 데이터의 분포 및 이상치 확인
		Z 검정	혈액 검사 데이터에 따른 고객의 건강 위험도 차이 검정
		Bar Plot	국민 건강 검진 정보 데이터 중 흡연/음주 여부와 성별의 상관관계 분석
보험 가입 심사 시 탈락되었던 가입 신청자 중 신규 고객 유치	<ul style="list-style-type: none"> <li>- 보험 가입 사전 검진 정보</li> <li>- 보험금 청구 및 지급 정보</li> </ul>	Box Plot	가입 신청자들의 데이터의 분포 및 이상치 확인
		상관 분석	보험 가입 사전 검진 정보 중 위험도 측정에 중요한 영향을 미치는 변수 확인
		Z 검정	혈액 검사 데이터에 따른 고객의 건강 위험도 차이 검정
		로지스틱 회귀분석 Decision Tree Random Forest Gradient Boosting	<ul style="list-style-type: none"> <li>- 고객 건강 위험도에 영향을 미치는 중요한 인자 확인</li> <li>- 평가 지표를 종합적으로 고려하여 분류 모델 정확도 확인 및 최종 모델 선정</li> <li>- 수익성에서 위험을 판단할 수 있는 지수 선정</li> <li>- 고객의 건강의 위험을 판단할 수 있는 지수 선정</li> <li>- 탈락되었던 고객 중 위험도가 낮은 고객 심층 분류</li> </ul>
보험 가입 심사 모델 개선	<ul style="list-style-type: none"> <li>- 보험 가입 사전 검진 정보</li> <li>- 국민 건강 검진 정보</li> </ul>	Box Plot	전체 데이터의 분포 및 이상치 확인
		로지스틱 회귀분석 Decision Tree Random Forest Gradient Boosting SVC KNN	<ul style="list-style-type: none"> <li>- 기존 고객을 성별로 분류 후, 심사 결과 예측 모델 평가 및 최종 모델 선정</li> <li>- 최종 모델을 활용하여 혈액 검사 데이터 중 중요한 인자 확인</li> <li>- 비중요 인자를 제거한 후, 기존 모델과 최종 모델의 정확도 비교 후 혈액 검사 간소화</li> </ul>

### 3. 분석 계획 및 결과 고객 세분화를 위한 목표변수 생성



#### 1. 수익위험지수

- 과지급 되고 있는 상품/고객 파악을 위해 목표변수 생성
- 기존 고객들을 대상으로 위험도 재 산정
- 높으면 '수익 위험고객', 낮으면 '수익 고객'으로 구분

$$\text{수익위험지수} = \frac{\text{지급 금액} - \text{청구 금액}}{\text{지급 금액}}$$

#### 2. 건강위험지수

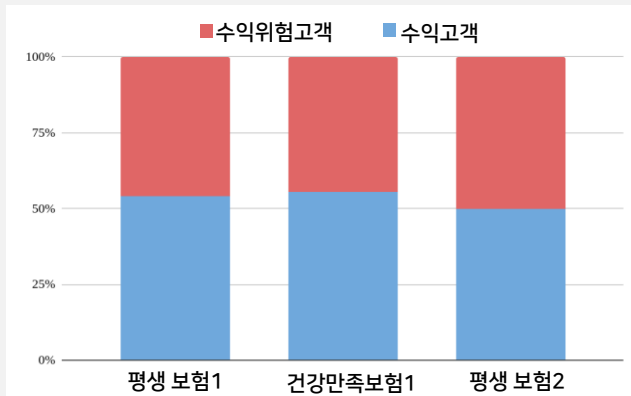
- 건강하지 않으면 보험 청구 건수가 증가하고, 그에 따라 보험료 과지급 가능성이 커진다고 판단
- 따라서 고객의 건강을 판단할 수 있는 새로운 지표 생성

$$\text{건강 위험 지수} = \frac{\sum_{n=i}^N \frac{X_i - \mu_i}{\sigma_i}}{N} \quad (\text{건강데이터 z-score})$$

### 3. 분석 계획 및 결과 ① 기존고객 분석을 통한 보험상품 및 고객 관리



#### 1. 보험상품 수익성 분석



**수익 위험 고객 36%(1,278명)**

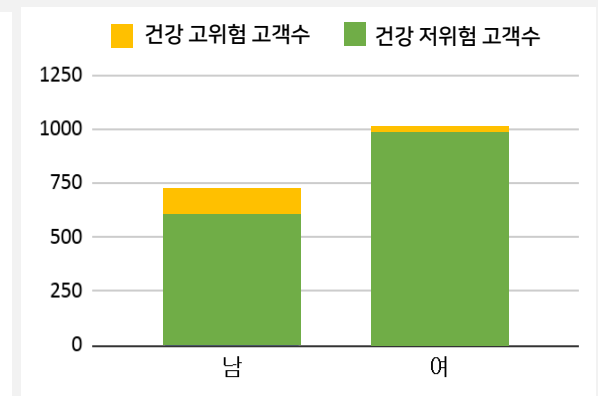
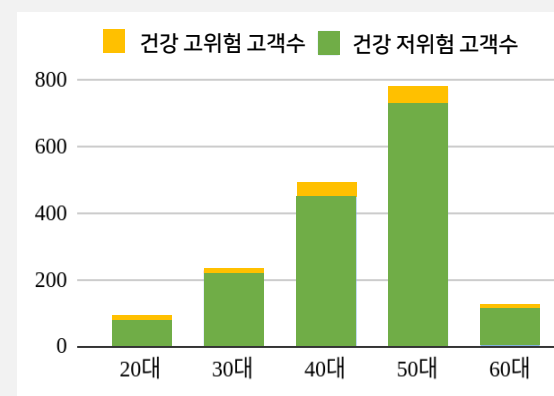
건강만족보험1(45%)

평생보험1(46%)

평생보험2(50%)

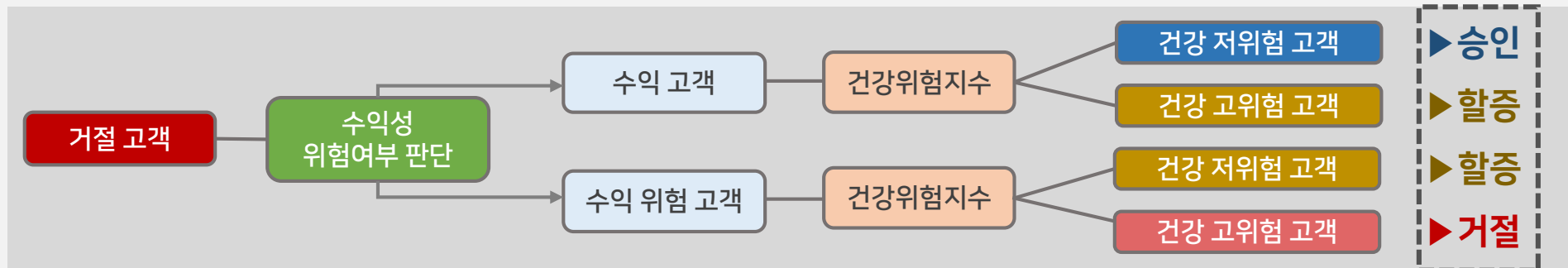
- 결과: 위 3가지 보험상품에 수익 위험고객 집중 분포 (낮은 수익성)
- 특정 보험상품에 대한 지급액 산정기준 강화 필요

#### 2. 고객 특성별 수익성 분석



- 결과: 4-50대 남성에 건강 고위험 고객 집중 분포
- 4-50대 남성에 대한 심사기준 강화 필요

### 3. 분석 계획 및 결과 ② 거절 고객 중 저위험 고객 분석



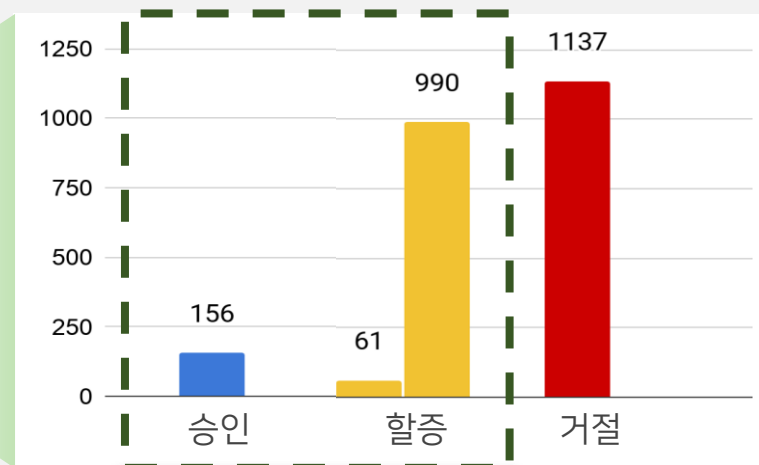
#### 수익성 위험 여부 판단

- 거절 고객은 보험금 청구 및 지급 데이터가 부재  
→ **거절된 고객의 수익성 판단이 불가능**



**“수익 고객, 수익 위험 고객 분류 모델 개발”**

승인	건강 저위험 고객	156
할증	건강 고위험 고객	61
	건강 저위험 고객	990
거절	건강 고위험 고객	1137



**“신규고객 확보”**



### 3. 분석 계획 및 결과 ③ 혈액 검사 간소화(신규가입 승인 모델 개선)

#### 성별로 적합 모델 도출

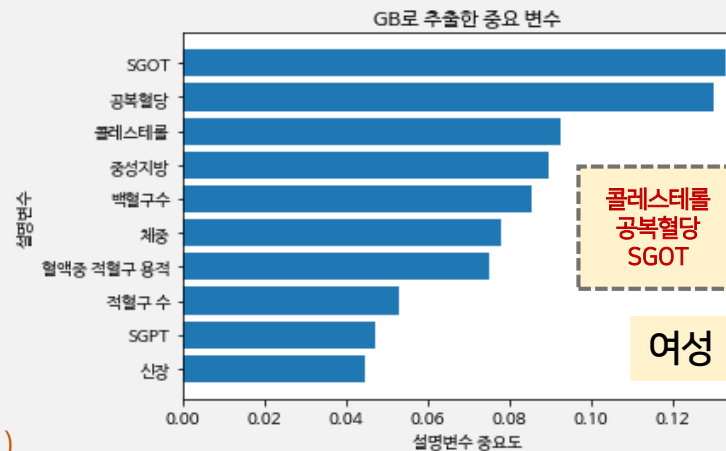
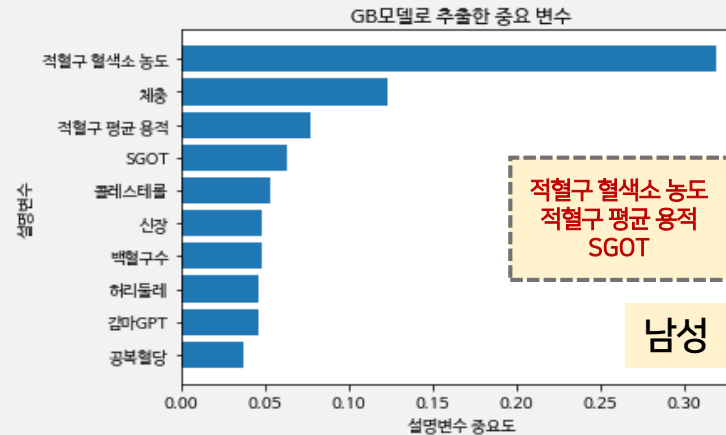
##### 남성

모델	Train accuracy	Test accuracy
Decision Tree	0.78	0.77
Random Forest	0.77	0.76
Gradient Boosting	0.78	0.77
KNN	0.80	0.75

##### 여성

모델	Train accuracy	Test accuracy
Decision Tree	0.85	0.81
Random Forest	0.89	0.84
Gradient Boosting	0.90	0.84
KNN	0.85	0.83

#### GB 모델로 중요 변수 추출



#### 단계별 혈액 검사 모델

##### 남성

```

judge      0
age        0
height     0
weight     0
waist      0
bt_chol    0
bt_crea    3020
bt_gluc    0
bt_hb      0
bt_hbbsa   0
bt_hct     3421
bt_mch     3421
bt_mchc    3421
bt_mvc     3421
bt_plat    3421
bt_rbc     3421
bt_wbc     3421
bt_rgpt    2
bt_sgpt    2
bt_sgpt    0
bt_trig    0
bp_judge   0
pulse_count_judge 0
dtype: int64
  
```

신장질환

혈구질환

지질질환

2단계

3단계

1단계

##### 여성

```

judge      0
age        0
height     0
weight     0
waist      1
bt_chol    0
bt_crea    3010
bt_gluc    0
bt_hb      439
bt_hbbsa   0
bt_hct     439
bt_mch     439
bt_mchc    439
bt_mvc     439
bt_plat    439
bt_rbc     439
bt_wbc     439
bt_rgpt    4767
bt_sgpt    4512
bt_sgpt    0
bt_trig    0
bp_judge   0
pulse_count_judge 0
dtype: int64
  
```

신장질환

혈구질환

간질환

지질질환

2단계

1단계

3단계

Train accuracy : 0.79

Test accuracy : 0.77

Train accuracy : 0.85

Test accuracy : 0.84

▶ 결측치 개수로 혈액 검사의 단계성 파악

중요 변수 추출



혈액 검사 단계

정확도를 유지하며 저감된 비용의 혈액 검사

보험 가입 사전 검진 정보 데이터로 학습 ~ 승인/재검(0), 거절(1)

▶ Gradient Boosting을 예측 모델로 채택

## 4. 개선안 및 적용방안 ① 보험상품 및 고객 건강 관리를 통한 수익성 개선

보험상품 기준 개선을 통한 수익성 개선



예시

건강만족보험1  
평생보험1  
평생보험2



수익성  
15% 상승

보험상품별 지급액 산정 기준 강화

▶ 보험상품 수익성 개선

고객 건강 관리를 통한 수익성 개선



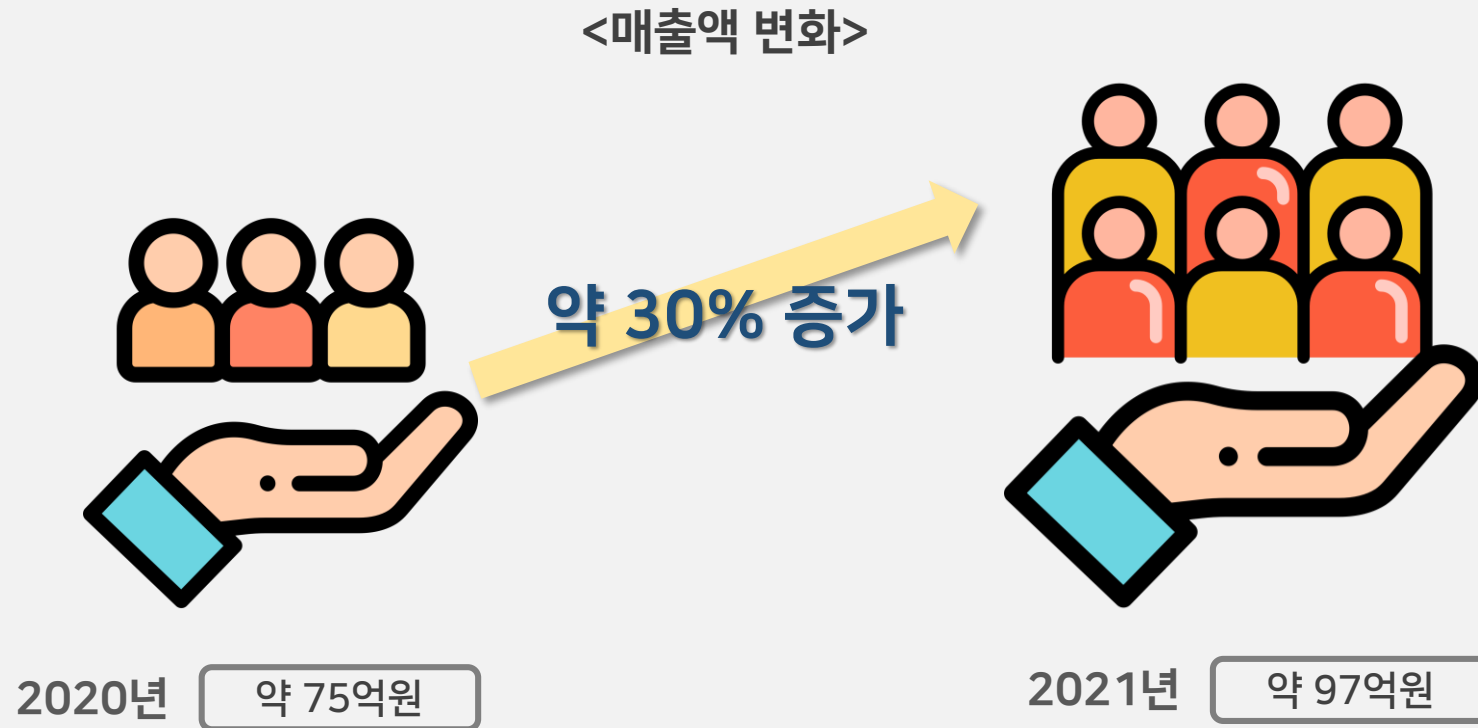
- 걷기 목표 달성 시 마일리지 지급
- 금연 목표 달성 시 마일리지 지급
- 꾸준한 건강검진 유도
- 맞춤형 보험상품 추천 및 보험료 할인 혜택

포빅 마일리지 제도 도입

▶ 소비자과 자사 효용

동시 증대

## 4. 개선안 및 적용방안 ② 거절고객 중 신규고객 확보를 통한 매출증대

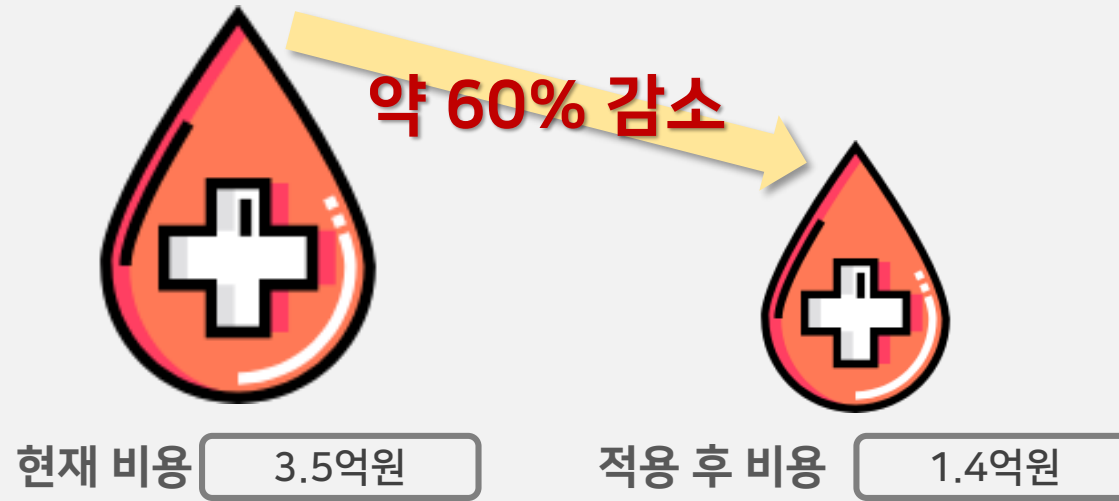


승인된 고객에게는 기존 보험료 산정  
저위험 고객에게 보험료 10% 할증



**매출액 21억 증가**  
**2020년 대비 30% 상승 기대**

## 4. 개선안 및 적용방안 ③ 혈액검사 간소화를 통한 리스크 감소 및 비용절감



- 성별에 따라 다른 기준으로 구성된 간소화 된 혈액검사
  - ▶ 보험가입 심사 및 웹 가입 시 적용 예정

3.5억원 혈액검사 비용을 1.4억원까지 절감 가능  
▶ 약 60%의 혈액검사 비용 절감 효과

## 5. 프로젝트 후기

### 강연주



프로젝트를 진행하며, 실무 환경에서 데이터 분석을 하는 방법을 익힐 수 있었습니다. 필요한 도메인 지식을 익히고, 정제되지 않은 데이터를 처리하고, 적합 모델을 찾는 과정을 통해 데이터 분석 역량을 함양할 수 있었습니다. 조원들과 충분히 토의하고 방향성을 조정하여 문제를 해결하며 즐거움을 느낄 수 있었습니다. 밤늦게까지 함께한 조원들과, 프로젝트 진행에 도움을 주신 교수님들께 감사의 말씀 전하고 싶습니다.

### 김지한



보험을 주제로 프로젝트를 진행하며, 도메인 지식은 분석을 위해 중요한 것을 알 수 있었습니다. 주제 선정, 계획, 분석 등 데이터 분석의 과정을 경험하면서 실무에서의 데이터 분석을 명확하게 알 수 있었습니다. 데이터 전처리를 진행하면서 분석에 있어 전처리의 중요성을 체감했습니다. 조원들과 의견을 공유하는 과정을 거치며, 분석의 내용이 한 단계 발전하는 것을 보며, 성장함을 느낄 수 있었습니다. 마지막으로 함께 고생해준 조원들과 지도해 주신 교수님들께 감사의 말씀드립니다.

### 김예은



보험이라는 생소한 주제의 프로젝트를 진행하면서 어려움도 많았지만 그만큼 많은 것을 배울 수 있었습니다. 조원들과 서로의 의견을 공유하며 문제를 해결해가는 과정을 통해 프로젝트의 방향성을 찾아갈 수 있었고 의미 있는 결과를 도출해 낼 수 있었습니다. 데이터 전처리부터 개선안 도출까지의 과정에서 목표 변수를 설정하고 논리를 쌓아가는 등 여러 어려움이 있었지만 함께 밤새워 고생한 조원들 덕분에 무사히 프로젝트를 완성할 수 있었습니다. 고생한 조원들에게 고맙다는 말 전하고 싶습니다!

### 손혜원



프로젝트를 진행하며 시작점에서 막막함을 느꼈지만 다양한 논문을 찾아보고 조원들과 열린 토론을 하며 방향을 찾고 차근차근 진행하여 성과를 내며 뿌듯함을 느꼈습니다. 조원 각자의 장점이 조화롭게 작용하여 열정과 노력을 바탕으로 프로젝트를 마무리할 수 있었고 많은 시행착오와 난관을 겪었지만 의미 있는 밑거름이었다고 생각합니다. 다음 데이터 분석 프로젝트에서는 이번 경험을 바탕으로 데이터를 가지고 놀 수 있는 실력으로 향상시키고 싶습니다. 함께 고생한 조원들과 교수님들께 감사드립니다.

### 김준오



실무 환경에서 활용하는 데이터를 직접 전처리를 진행하고 분석하며 결론을 도출해나가는 과정이 뿌듯했습니다. 특히 언더라이터를 위한 웹페이지를 개발하는 업무를 수행하여 수많은 오류를 마주했지만 고객의 정보를 입력하면 심사 결과가 도출되는 결과물 덕분에 뜻깊은 시간이었습니다. 일주일 동안 함께 고생해 준 조원들과 4주간 교육에 힘써주신 교수님들께 진심으로 감사드립니다.

### 하준범



평소에 관심 있던 보험을 주제로 빅데이터 분석을 진행할 수 있어서 재미있었습니다. 그동안 학습하면서 다뤄온 정제된 데이터가 아닌 로우 데이터를 다루면서 처음 겪는 어려움이 많았지만, 하나씩 문제를 해결해 나가는 과정 속에서 한층 더 성장할 수 있었습니다. 엑셀이 아닌 파이썬 판다스를 이용해 데이터를 분석하는 것도 새로운 경험이었습니다. 하지만 무엇보다 데이터 분석의 핵심 목표를 도출하고 논리구조를 다져나가는 과정 속에서 많은 것들을 배울 수 있었습니다. 함께 밤새워 고생한 조원들에게 고맙다는 말을 전하고 싶습니다.

## 6. Web 시연 언더라이팅 웹

용도 : 간편 보험설계를 위한 언더라이팅 웹

Tool : Python Django, Django Rest Framework, Heroku(배포)

The screenshot shows the 'Underwriting' form with the following fields and labels:

- 성별 \* (Gender): dropdown menu
- 나이 \* (Age): text input
- 키 \* (Height): text input
- 몸무게 \* (Weight): text input
- 허리둘레 \* (Waist Circumference): text input
- 혈압 판정결과 \* (Blood Pressure Assessment Result): dropdown menu
- 맥박 판정결과 \* (Pulse Assessment Result): dropdown menu
- 식전혈당 (bt\_gluc) \* (Fasting Blood Sugar): text input
- 총 콜레스테롤 (bt\_chol) \* (Total Cholesterol): text input
- 트라이글리세라이드 (bt\_trig) \* (Triglycerides): text input
- 혈색소 (bt\_hb) (Hemoglobin): text input
- 혈청크레아티닌 (bt\_crea) (Serum Creatinine): text input
- 혈청지오티 SGOT (bt\_sgot) (Serum Aspartate Aminotransferase): text input
- 혈청지피티 SGPT (bt\_sgpt) (Serum Alanine Aminotransferase): text input
- 감마지티피 (bt\_rgpt) (Gamma-GT): text input
- B형 간염 항원 (bt\_hbsa) (Hepatitis B Surface Antigen): dropdown menu
- 흡연상태 (Smoking Status): dropdown menu
- 음주여부 (Alcohol Consumption): dropdown menu

A blue button labeled '결과 확인' (Check Result) is at the bottom right.

간편 언더라이팅 웹 입력 창

The screenshot shows the same 'Underwriting' form with sample data entered:

- 성별 \*: 남자 (Male)
- 나이 \*: 27
- 키 \*: 170
- 몸무게 \*: 70
- 허리둘레 \*: 32
- 혈압 판정결과 \*: 정상 (Normal)
- 맥박 판정결과 \*: 정상 (Normal)
- 식전혈당 (bt\_gluc) \*: 80
- 총 콜레스테롤 (bt\_chol) \*: 120
- 트라이글리세라이드 (bt\_trig) \*: 100
- 혈색소 (bt\_hb): 28
- 혈청크레아티닌 (bt\_crea): 0.7
- 혈청지오티 SGOT (bt\_sgot): 30
- 혈청지피티 SGPT (bt\_sgpt): 40
- 감마지티피 (bt\_rgpt): 70
- B형 간염 항원 (bt\_hbsa): Negative
- 흡연상태: 흡연 (Smoking)
- 음주여부: 음주 (Drinking)

A red box highlights the '결과 확인' button with the text 'Click!!'.

가입심사를 위한 혈액검사 데이터 입력

\*언더라이터: 피보험자의 신체적, 재정적 위험 등을 종합적으로 고려해 보험계약 체결 여부를 심사하는 직업

## 6. Web 시연 언더라이팅 웹

데이터 입력 후 혈액검사 기반 가입심사 모델 적용

데이터가 기준 내 있을 경우: 승인

데이터가 기준 내 없을 경우: **거절**



해당 고객의 예상 판정 결과는 '**승인**'입니다.

고객 응대 시, 빠른 언더라이팅을 위한 서비스입니다.  
필수 필드는 반드시 입력해주시고, 나머지 필드는 선택적으로 입력해주세요.



해당 고객의 예상 판정 결과는 '**거절**'입니다.

고객 응대 시, 빠른 언더라이팅을 위한 서비스입니다.  
필수 필드는 반드시 입력해주시고, 나머지 필드는 선택적으로 입력해주세요.

혈액검사 결과와 음주 및 흡연 여부의 입력만으로 간편하게 가입 가능 여부 확인

**“보험 가입 절차 간소화 실현”**

## 첨부 데이터 항목 및 활용

	데이터 셋	데이터 수	의미	목적	파생변수
메인	보험/청구 지급 정보	24 variables 49,449 rows	고객의 보험 청구 금액과 지급 정보	기존 고객 및 보험 수익성 파악 위험 고객 군 재정의	bmi whtr
	보험가입 사전 검진 정보	48 variables 14,938 rows	보험 가입 신청자의 혈액검사 일반검사 결과 및 판정결과	기존 거절 고객들 중 저 위험 고객 발굴 고위험 고객의 보험료 할증 방안 수립	수익-위험 지수 건강-위험 지수
참조	국민건강검진 결과	24 variables 177,346 rows	국민건강 검진결과	혈액검사 대체 예측모델 구현 고객 가입 심사 평가 기준 추가 개발 및 세분화	
	병명 코드	7 variables 13,755 rows	상병 정보		
	보험 상품별 보험료	3 variables 29 rows	보험 상품별 보험료		

\* bmi = 키/체중<sup>2</sup>

\* whtr = 허리둘레/신장



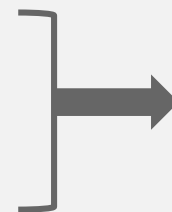
## 첨부 데이터셋 설명

### 보험 청구/지급정보 (24 variables / 49,449 rows)

데이터 셋	의미	항목	항목명	속성
Insu_request (24 variables / 49,449 rows)	고객의 보험 청구금액과 지급 정보	보험상품	insu_prod_id	이산형
		보험청구 금액	insu_req_amount	연속형
		보험지급 금액	insu_pay_amount	연속형
		누적 납입 보험료	insu_cum_amount	연속형

기존 고객 및  
보험 수익성 파악

- 과지급 보험 상품 및 고객 분석
- 누적 납입금 대비 지급 금액이 높은 위험 고객 분석
- 단기 보험금 수령 고객 분석



위험고객군  
재정의

## 첨부 데이터셋 설명

### 보험가입 사전 검진 정보 (48 variables / 14,938 rows)

데이터 셋	의미	항목	항목명	속성
insu_pre_review (48 variables / 14,938 rows)	보험 가입 신청자의 혈액검사 / 일반검사 결과 및 판정결과	심사결과	judge	범주형 (승인, 거절, 재검)
		고객 기본 인적사항	gender, age, height, weight, vust, waist	이산형, 연속형
		혈액검사 결과	bt_chol(콜레스테롤), bt_crea(혈청 크레아티닌)	연속형
		혈액검사 판정결과	judge_score (판정결과 합), bt_chol_judge, bt_crea_judge	이산형



1. 기존 거절 고객들 중 저위험 고객 발굴    2. 고위험 고객의 보험료 할증 방안 수립

## 첨부 데이터셋 설명

데이터 셋	의미	항목	항목명	속성
국민건강 검진결과 (24 variables / 177,346 rows)	국민건강 검진결과	건강 데이터	gender, age, weight, height, bmi, smoke_flag, bt_chol	이산형, 연속형
병명 코드 (7 variables / 13,755 rows)	상병 정보	질병유형	sick_cd_3, sick_desc_kor, gender	이산형, 연속형
보험 상품별 보험료 (3 variables / 29 rows)	보험 상품별 보험료	보험상품, 기본 보험료	insu_prod_name, base_price	이산형, 연속형