

제2회 데이터기반 리스크관리 경진대회

자동차보험 사고율 예측에 기반한 고객 맞춤 특약 추천 마케팅 제안

용앤리치

강세현 심현구 안은선 최용원



목차

| 서론

- 분석 배경
- 분석 흐름

| 결론

- 결과 적용 방안 제안
- 시사점 및 제언

| 본론

- 데이터 개요 및 EDA
- 데이터 전처리
- 사고율 예측 모델링
- 특약 연관규칙분석

전략 재수립을 위한 사고율 예측 모델의 필요성

“자동차보험 상반기 손익 11% 감소...
이동량 증가로 손해율 상승”

경향신문, 2023.09.11

손해보험사들의 올해 상반기 자동차 보험 손익이
지난해 상반기보다 감소한 것으로 나타났다.

이동량 증가로 손해율이 올라갔지만,
코로나19 이전과 비교하면 여전히 낮은 수준을 유지했다.

...

가입 대수 증가로 매출이 증가했지만,
사고율 증가에 따라 지급보험금이 증가하고
보험료가 인하된 영향이다.

- 코로나19가 완화된 이후 자동차의 통행량이 증가함에 따라
자동차보험 손해율이 다시 지속적으로 증가하는 추세
- 23년 상반기 기준 손해보험사들의 손해율은 78.0%로,
전년 동기 (77.1%)에 비해 약 0.9%p 상승한 모습을 보임



보험사의 대응 방안

환경 변화까지 복합적으로 고려한 사고 예측 모델링을 통해
선제 대응 방안 마련 및 손해율 관리 필요

특약을 활용한 초개인화 서비스의 필요성

“지속 강화되는 차보험 소비자 혜택 경쟁 심화”

보험매일, 2023.10.19

삼성화재의 **안전운전 관련 특약 혜택 확대**는
대형사 중심의 점유율 경쟁이 치열해지면서
이에 대응하기 위한 것으로 풀이된다.

이는 **안전운전에 따른 보험료 할인**을 제공하면서
소비자의 사고율을 경감시키고 그로 인해 **자동차보험 손해율이
낮아지는 효과**로 이어지기 때문이다.

“삼성화재, 고연식차량 대상 블랙박스 할인 폐지”

대한금융신문, 2020.03.13

자동차보험의 블랙박스 할인이 점차 축소되고 있다.

...

삼성화재 관계자는 “12년 이상 된 차량의 경우,
**블랙박스 장착 차량이 미장착 차량보다 오히려 사고율이 높아
할인특약을 없애게 됐다**”고 말했다.

- 보험업계에서 낮은 사고율과 연관된 특약으로
우량 고객을 모집하려는 ‘특약경쟁’ 심화
- 그러나, 블랙박스 특약의 경우 가입자의 사고율이
미가입자보다 높아 특약의 실효성에 대한 문제 제기



보험사의 대응 방안

사고율 완화에 도움이 되는 **특약 가입 고객의 특성**을 파악해
해당 고객들을 대상으로 하는 **집중적인 특약 추천** 마케팅을 통한
특약의 실효성 제고 필요

1. 고객 이진 분류 모델링

Logistic Regression

- AUC
- F1-Score
- Macro F1-Score

고위험군 고객

저위험군 고객

2. 사고율 예측 모델링

Model Ensemble

- Random Forest
- Gradient Boosting
- Decision tree Regressor

고위험군 모델

저위험군 모델

모델 해석

- Feature Importances
- SHAP

3. 연관분석

특약 가입자 특징 Filtering

- Apriori Algorithm

예측사고율 감소 검정

- paired t-test



리스크 관리 방안 제시

데이터 설명

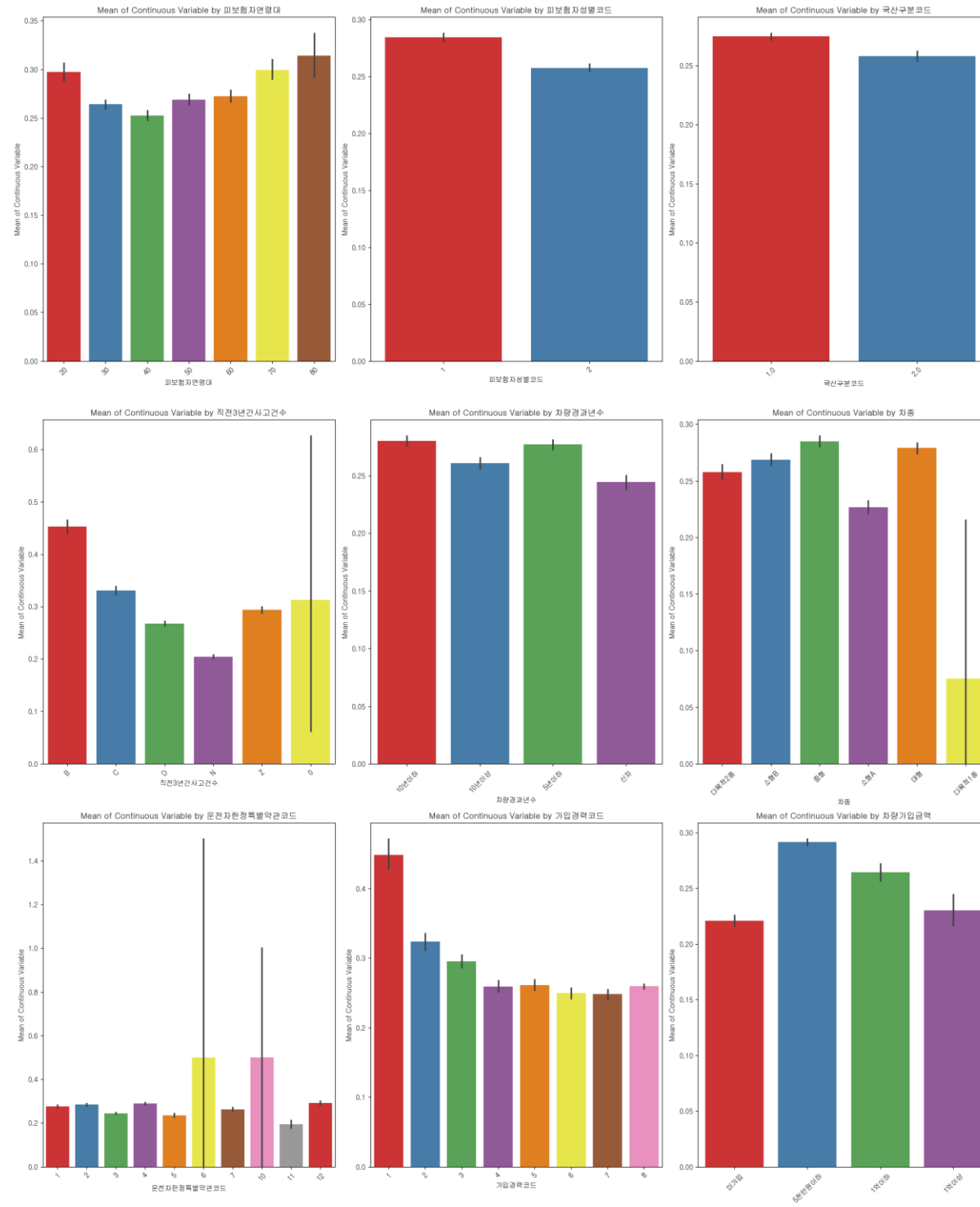
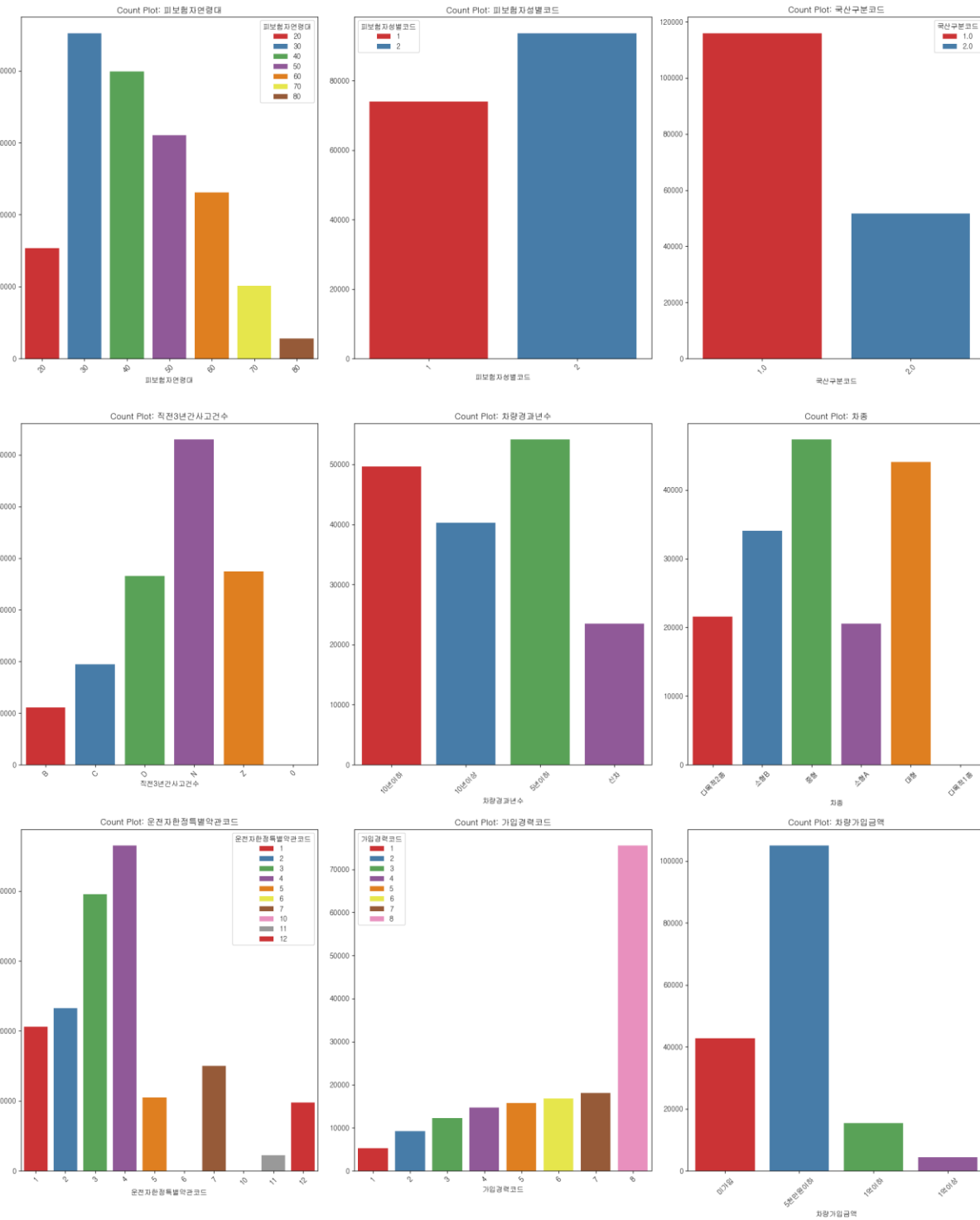
- 관측 기간 : 2022.07 ~ 2023.06
- 관측값 : 167,884

변수 설명

- 피보험자연령대 : 자동차 계약 보험 증권에 피보험자로 기재되어 있는 기명 피보험자의 연령대
- 직전3년간사고건수 : N:무사고, D:3년1회, C:3년2회, B:3년3회, ZZZ:기타(가입 3개월이내 신규계약)
- 마일리지약정거리 : 마일리지 약정별 예상 할인율
- 유효대수 : 경과일수에 따른 가입대수
- 사고건수 : 유효대수 기간 중에 발생한 사고건수

상품 대분류	피보험자 연령대	피보험자 성별코드	국산구분 코드	직전 3년간 사고건수	차량 경과년수	배기량별 차종분류	운전자 한정특별 약관코드	운전자 한정 특별약관	가입경력 코드	차량가입 금액	블랙박스 특약요율	마일리지 약정거리	유효대수	사고건수
A10	20	001	1	B	10년 이상	소형 A	3	부부한정	02	5천만원 이하	미가입	미가입	0	0
A10	20	001	1	B	10년 이상	소형 B	3	부부한정	04	5천만원 이하	미가입	미가입	0	1
...
A10	40	002	1	C	10년 이상	소형 A	4	기명 피보험자 1인한정	06	미가입	미가입	15000K	2	0

EDA



사고율과 변수 간의 관계를
파악하기 위해
변수별 시각화 진행

데이터 클렌징

- 유효대수가 0인 관측치
 - 피보험자연령대에 10대, 90대 존재
 - 국산구분코드 항목에 11개의 결측치 존재

→

사고율 계산이 불가해 유효대수가 0인 관측치 제거

→

이상치로 판단해 제거

→

전체 약 17만개의 데이터에 대해 영향력이 미미한 것으로 판단해 결측치 제거

자료형 변환

범주형 변수 중 문자 자료에 대해 Encoding 진행

변수명	Encoding	상세
차량경과년수	Label Encoding	각 범주 간 위계 존재 가장 작은 값 0(미가입 포함), 그 크기에 따라 1씩 증가하는 정수형 변수로 표현
차종		
차량가입금액		
마일리지약정거리		
블랙박스특약요율	One-hot Encoding	가입자 = 0, 미가입자 = 1으로 표현
운전자한정특별약관		각 범주 간 위계가 존재하지 않음
직전3년간사고건수		'기타'와 같이 순서를 부여하기 어려운 범주 존재

파생변수 생성

- 사고율 : $\text{사고건수} \div \text{유효대수}$, 모델링의 종속변수로 사용
- 마일리지할인율 가입여부 : 마일리지할인특약을 가입한 고객과 가입하지 않은 고객을 분류
- 자차보험가입여부 : 차량가입금액을 이용해 자차보험을 가입한 고객과 가입하지 않은 고객을 분류
- 고경력운전자 : 가입경력 7년 이상 고객을 고경력운전자로 간주
- 저경력운전자 : 가입경력 1년 미만 고객을 저경력운전자로 간주

사고율 예측 모델링 과정

1

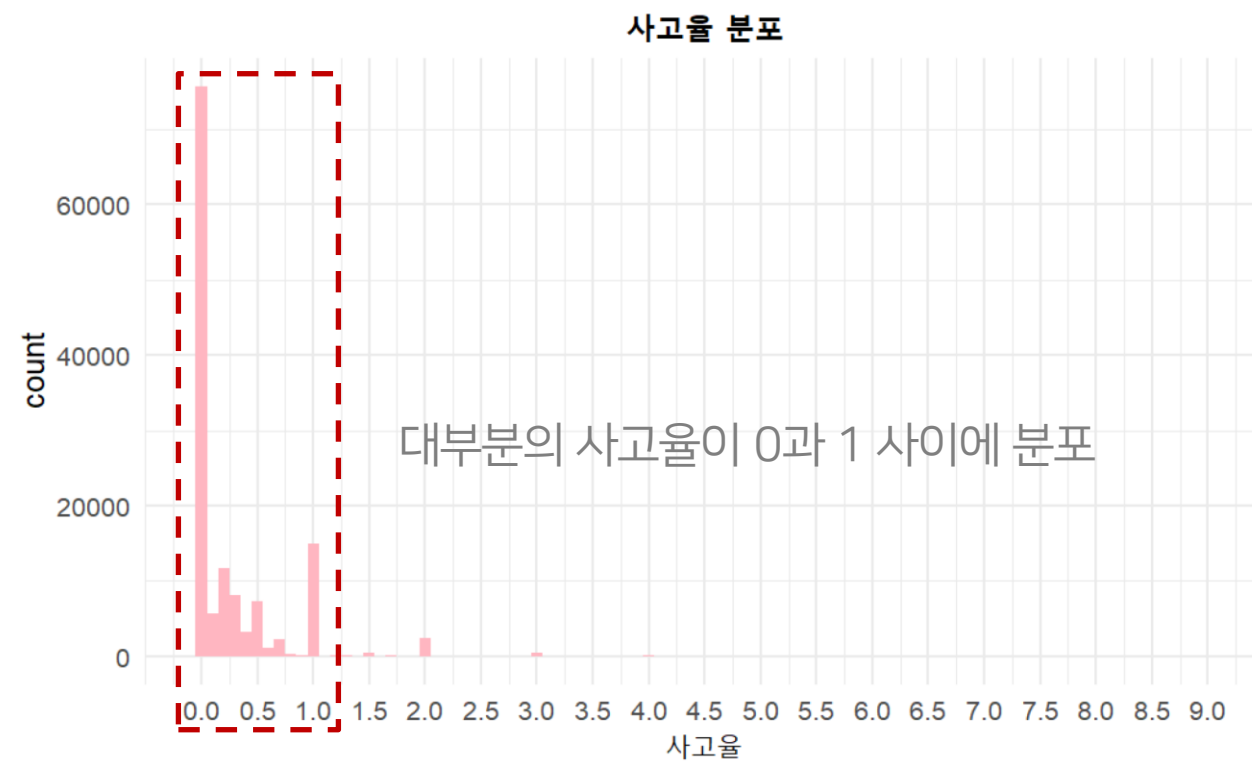
전체 데이터의 80%를
학습데이터로,
20%를 검증데이터로 나누어
예측 모델링 진행

2

고객을
사고율이 1 이상인 고위험군 고객과
사고율이 1 미만인 저위험군 고객으로
나누어 이진 분류 모델링 진행

3

이진 분류로 예측한 값을 기반으로
고위험군 고객과 저위험군 고객에 대한
사고율 예측 모델링을 따로 진행



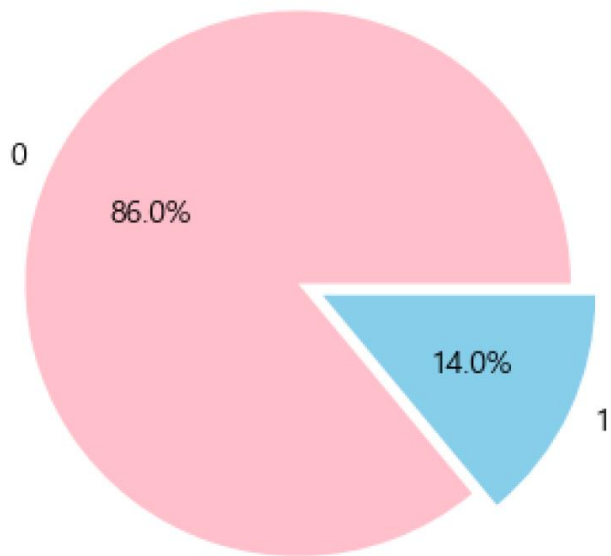
고객의 사고율이 주로 0과 1의 범위에 속해 있어
예측 모델이 높은 사고율을
예측하지 못하는 문제를 해결하기 위함

고위험군/저위험군 이진분류 모델링

클래스 불균형을 해결하기 위해 최적의 임계점을 설정한 로지스틱 회귀 모델로 고위험군/저위험군 이진분류 모델링 진행

클래스 불균형

학습데이터 중
저위험군 고객이 86%,
고위험군 고객이 14%로
클래스 불균형 문제가
발생한 것을 확인



변수 선택

유의수준 0.05 하에서
예측에 유의미한 변수 선택

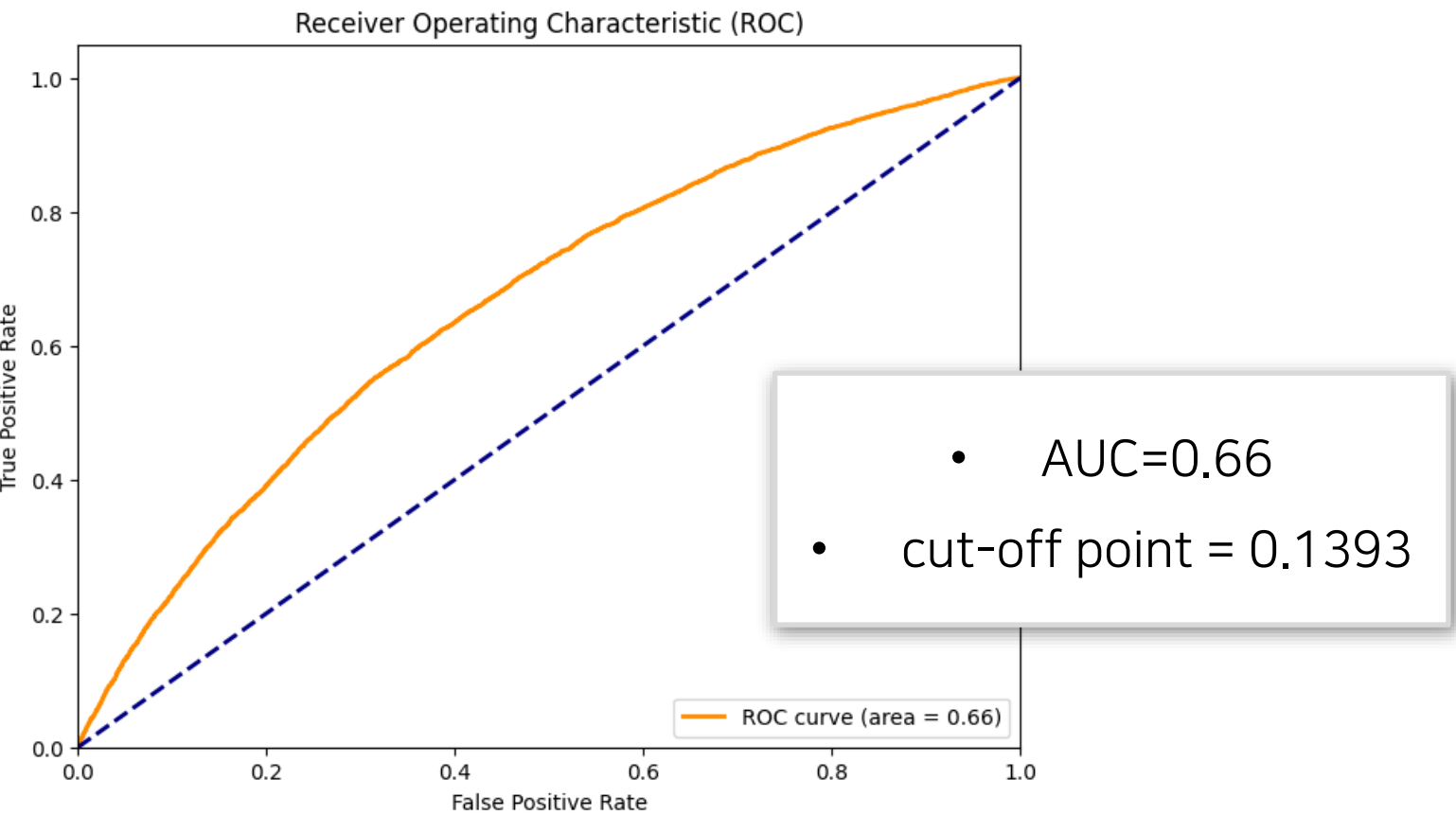
변수 간 다중공선성을 제거하고
분류 예측에 유의미한 영향을 미치는
변수를 파악하기 위함

로지스틱 회귀

고위험군 예측 성능을 극대화하는
임계점을 이용해 분류 진행

클래스 불균형으로 인한
문제를 해결하기 위함

고위험군/저위험군 이진분류 모델링 결과



성능 평가 지표 : Macro F1-Score

이진 분류 모델 성능 : 0.5232

각 클래스별로 F1-Score의 평균을 내 동등하게 클래스를 고려하는 성능 지표
소수의 클래스를 얼마나 잘 찾는지도 고려하기 위함

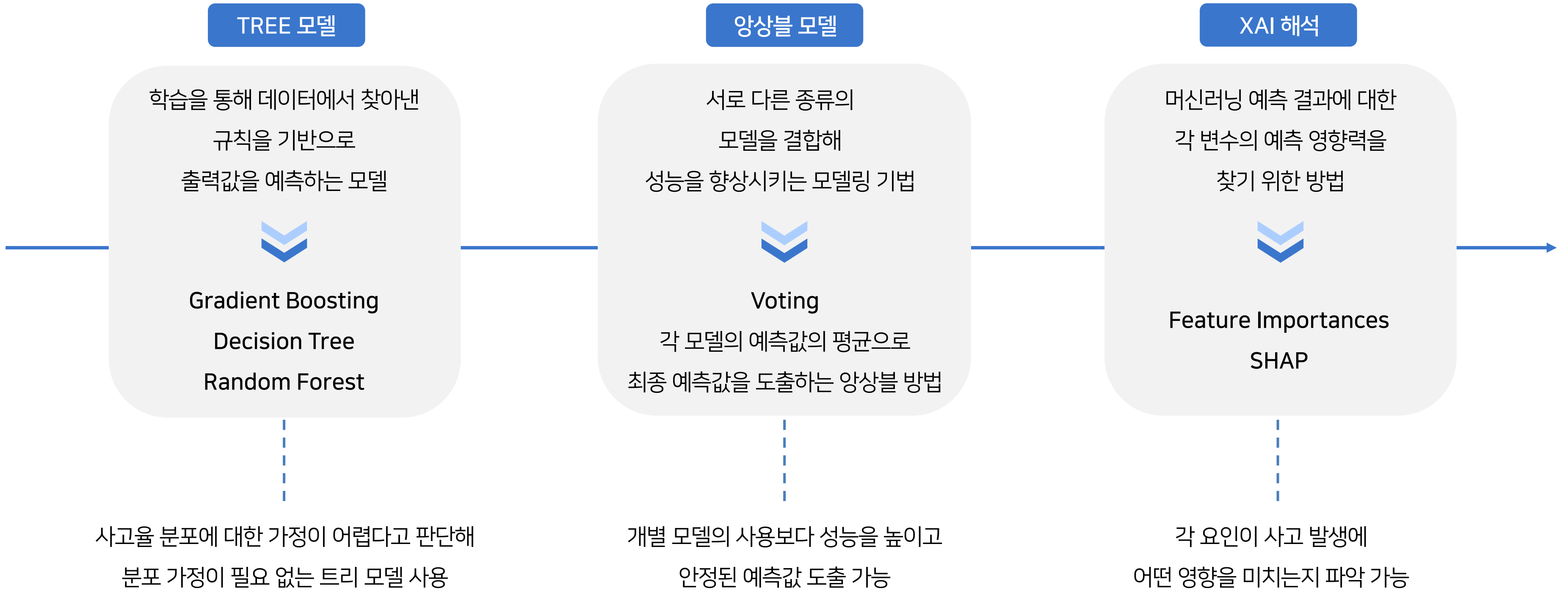
변수명	계수	변수명	계수
피보험자연령대	0.0881	직전3년간 무사고	-0.4801
피보험자성별	-0.2145	직전3년간 사고1건	-0.0263
국산구분	0.1641	직전3년간 사고2건	0.4396
가입경력	-0.1014	직전3년간 사고3건	1.0050
차량가입금액	0.0838	운전자한정특별약관 부 부한정 가입	-0.4307
블랙박스 특약요율	-0.1239	운전자한정특별약관 기 명피보험자1인한정 가입	-0.2457
마일리지할인특약 가입여부	-0.3188	운전자한정특별약관 1인및지정1인 가입	0.0718
차종	-0.0180	운전자한정특별약관 임 직원한정 가입	0.1807
고경력운전자	-0.4260	운전자한정특별약관 부 부및자녀한정 가입	0.0990

고위험군/저위험군 이진분류 모델 해석 결과

- 피보험자연령대가 높을수록, 직전 3년간 사고건수가 많을수록 고위험군일 확률이 높아지는 것을 파악
 - 가입경력이 길수록, 할인 특약에 가입할수록 저위험군일 확률이 높아지는 것을 파악

| 사고율 예측 모델링

트리 기반 모델인 Gradient Boosting, Decision Tree, Random Forest을 Voting 방법으로 결합한 앙상블 모델로 사고율 예측 모델링 진행



사고율 예측 모델링 결과

성능 평가 지표 : MSE(Mean Squared Error)

저위험군 예측 모델 성능 : 0.60

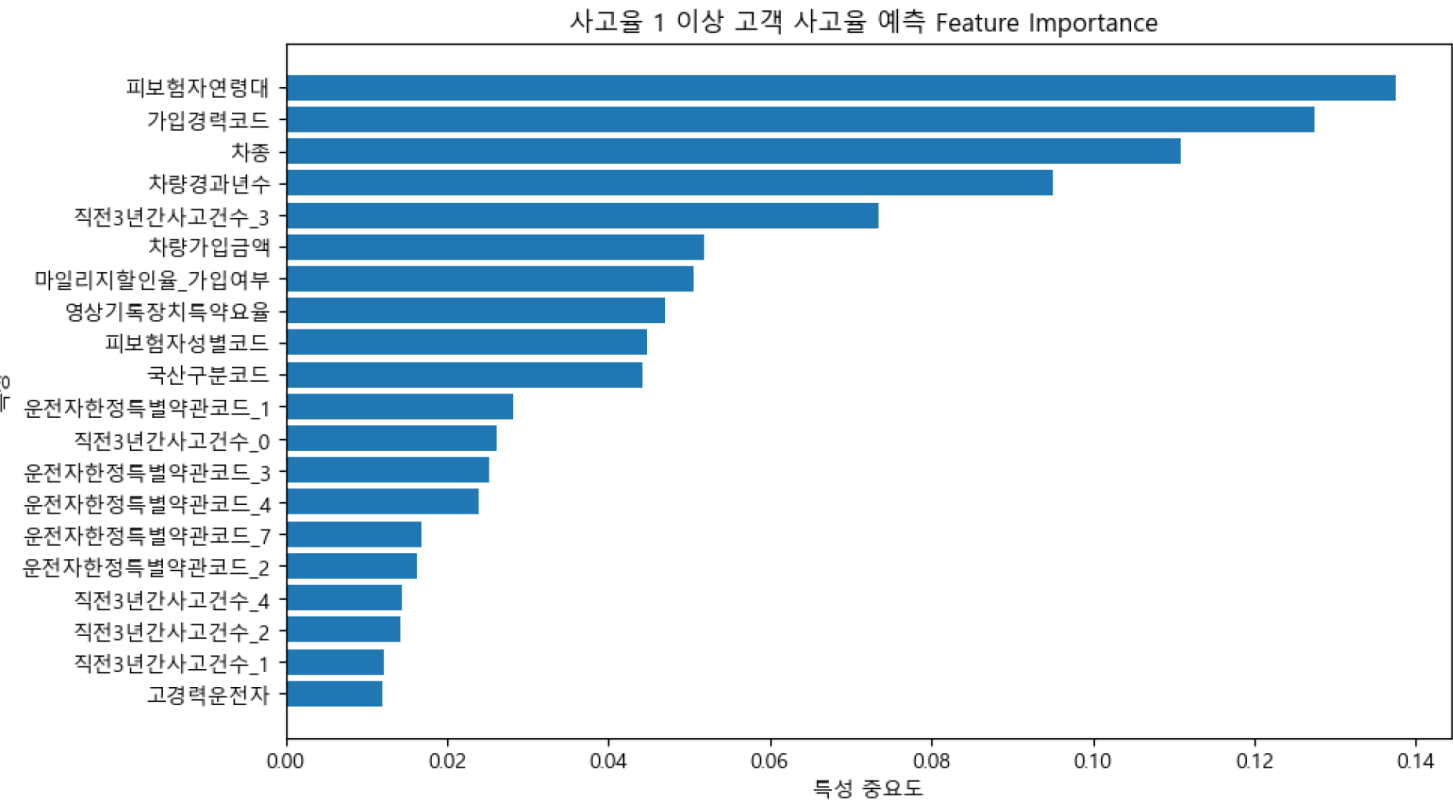
고위험군 예측 모델 성능 : 0.73



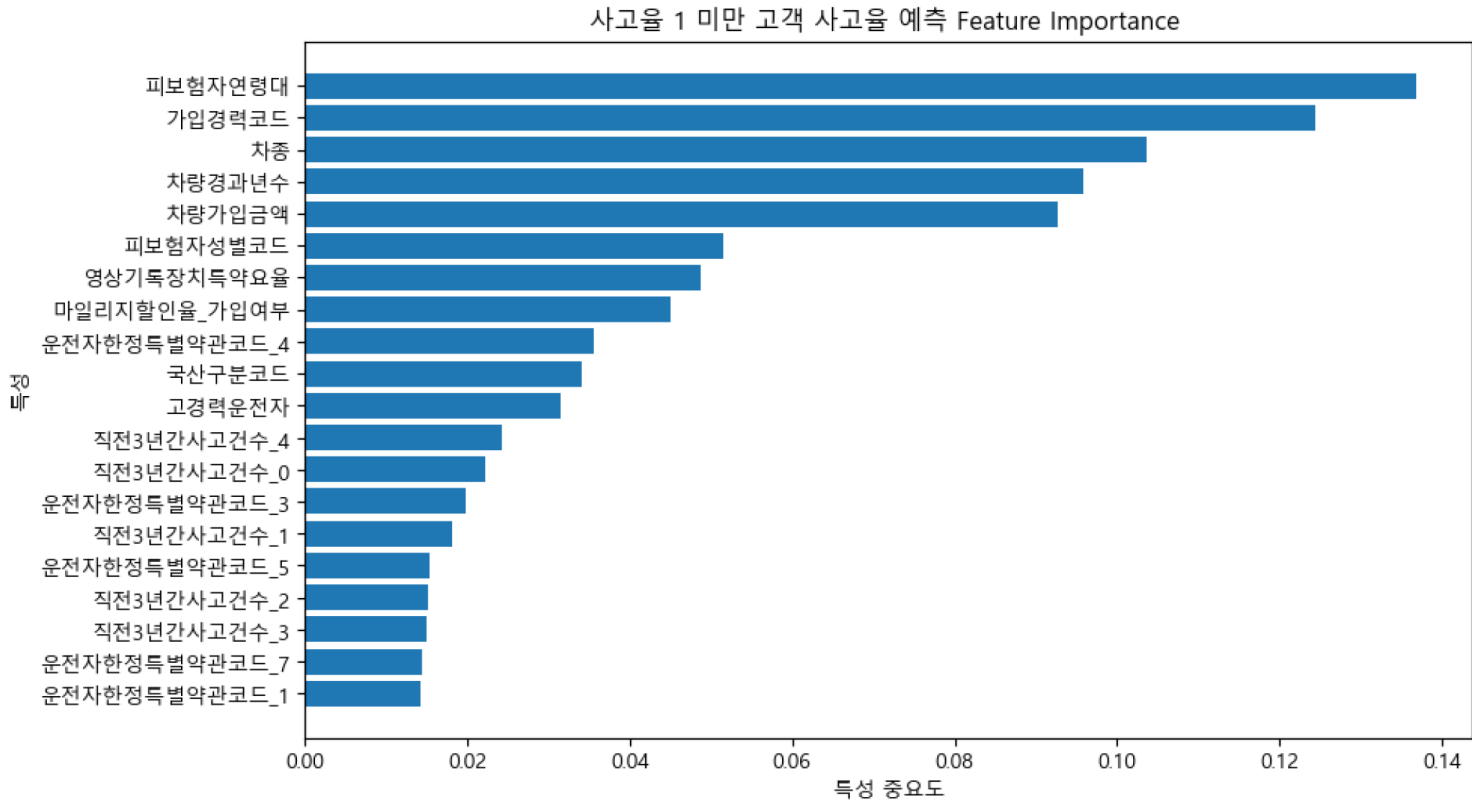
최종 사고율 예측 모델 성능 : 0.61

고위험군/저위험군 분류가 잘 이루어졌다고 가정했을 시 사고율 예측 성능은 0.076

고위험군 사고율 예측 모델 결과 해석



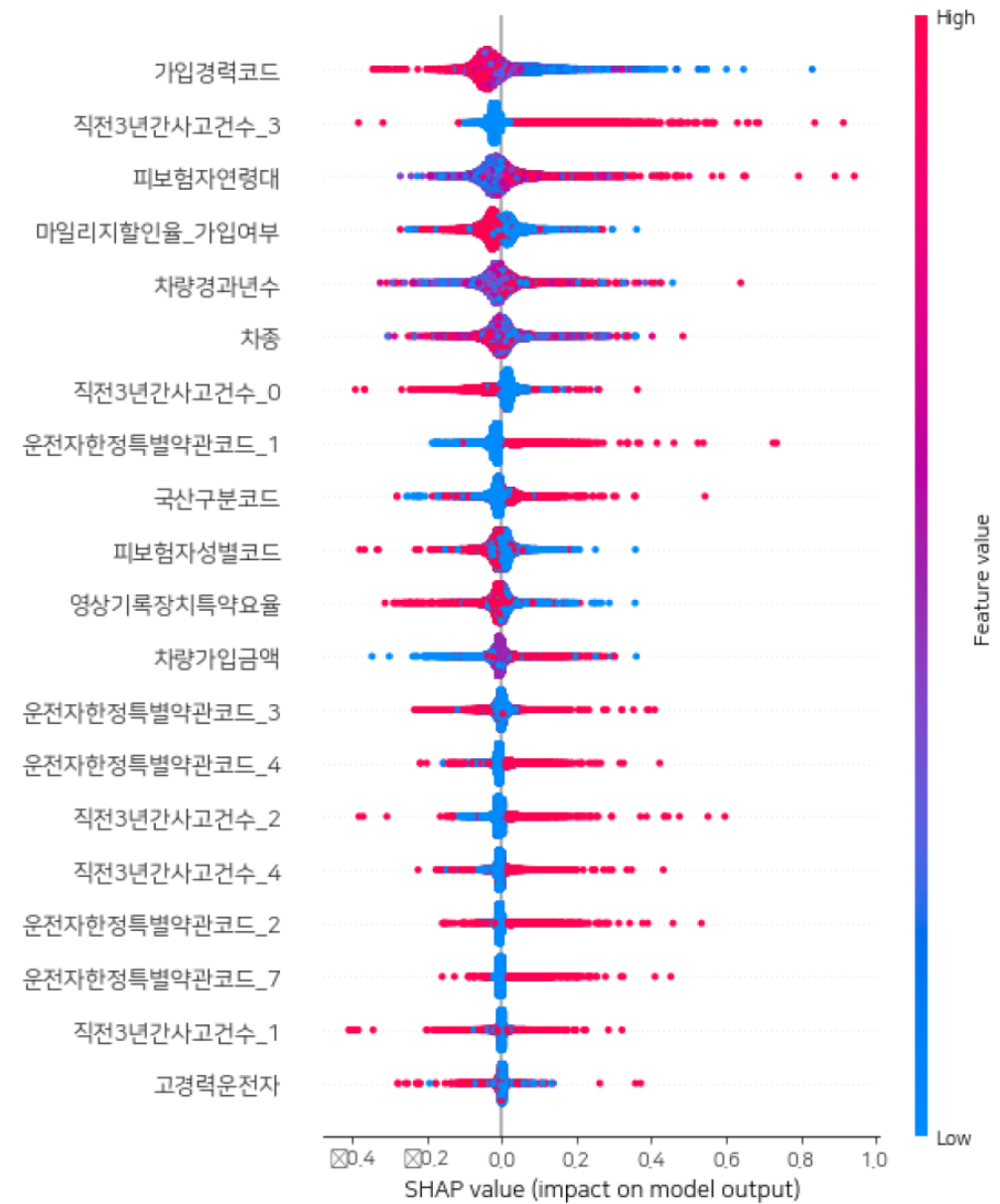
저위험군 사고율 예측 모델 결과 해석



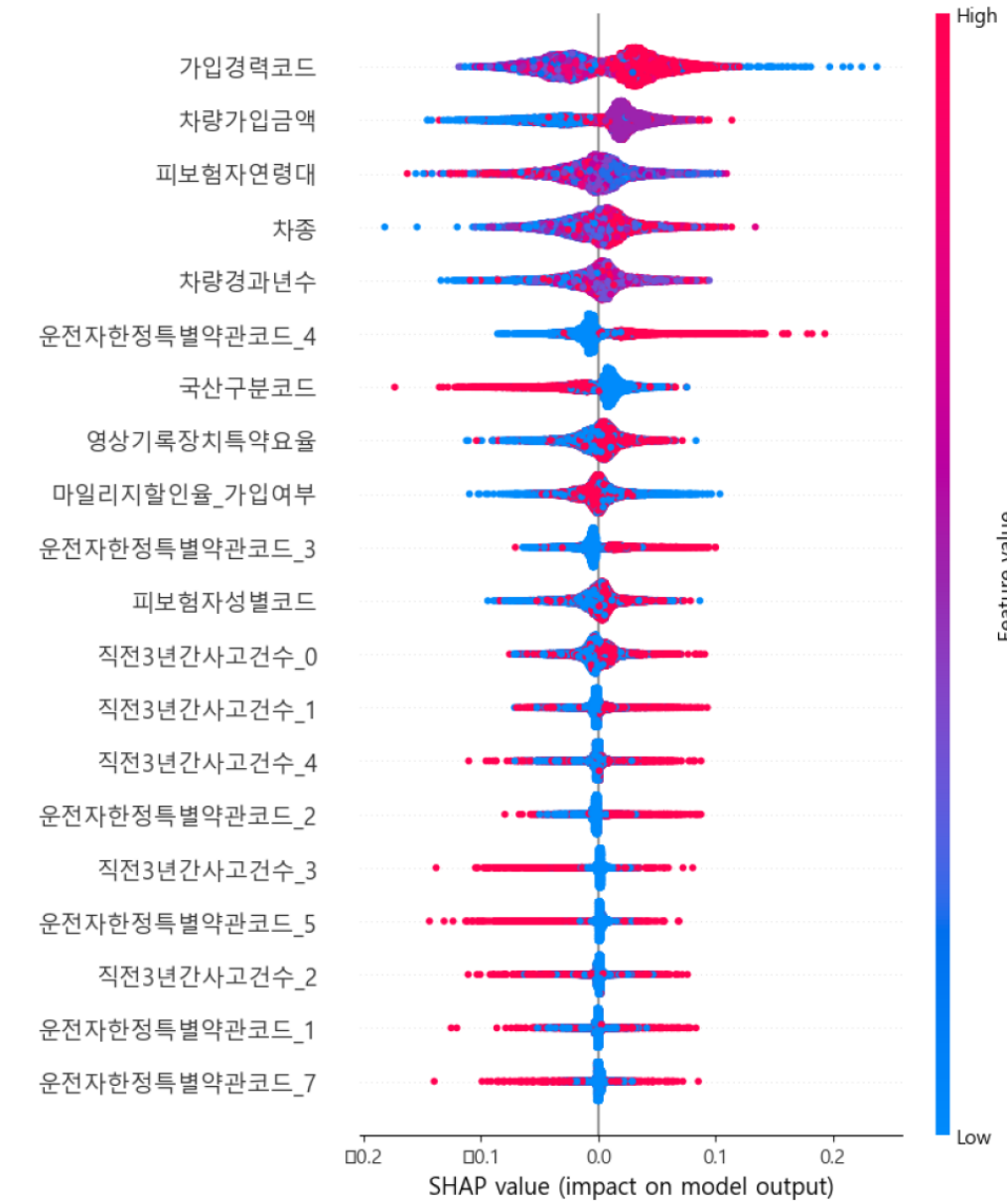
Feature Importance 해석 결과

- **피보험자연령대, 가입경력, 차종, 차량경과년수, 할인특약 가입여부**가 공통적으로 사고율 예측에 중요한 역할을 한다는 것을 파악
- 저위험군 고객보다 **고위험군** 고객에서 **직전 3년간 사고건수**가 상대적으로 사고율 예측에 중요한 역할을 한다는 것을 파악

고위험군 사고율 예측 모델 결과 해석



저위험군 사고율 예측 모델 결과 해석



SHAP 해석 결과

- 고위험군 고객의 경우, 가입경력, 피보험자연령대, 직전3년간사고건수가 높을수록 사고율이 높아지며 할인 특약 가입할수록 사고율이 낮아지는 것을 파악
- 저위험군 고객의 경우, 기명피보험자1인한정 및 부부한정을 가입할수록 사고율이 높아지며 차량이 국산차인 경우 사고율이 낮아지는 것을 파악

| 연관규칙분석 활용 배경

사고율 예측 모델링을 통해
고객이 특약에 가입할 수록
사고율이 낮아지는 상관관계가 존재하는 것을 파악



특약에 가입할 의사가 높은 동시에,
특약 가입 시 사고율이 유의미하게 감소할 것으로 예측되는
고객들을 대상으로 특약 가입을 추천한다면
손해율 관리와 특약 가입자 증대를 동시에 실현 가능할 것으로 기대

| 연관규칙분석 및 사고율 검정 과정

1

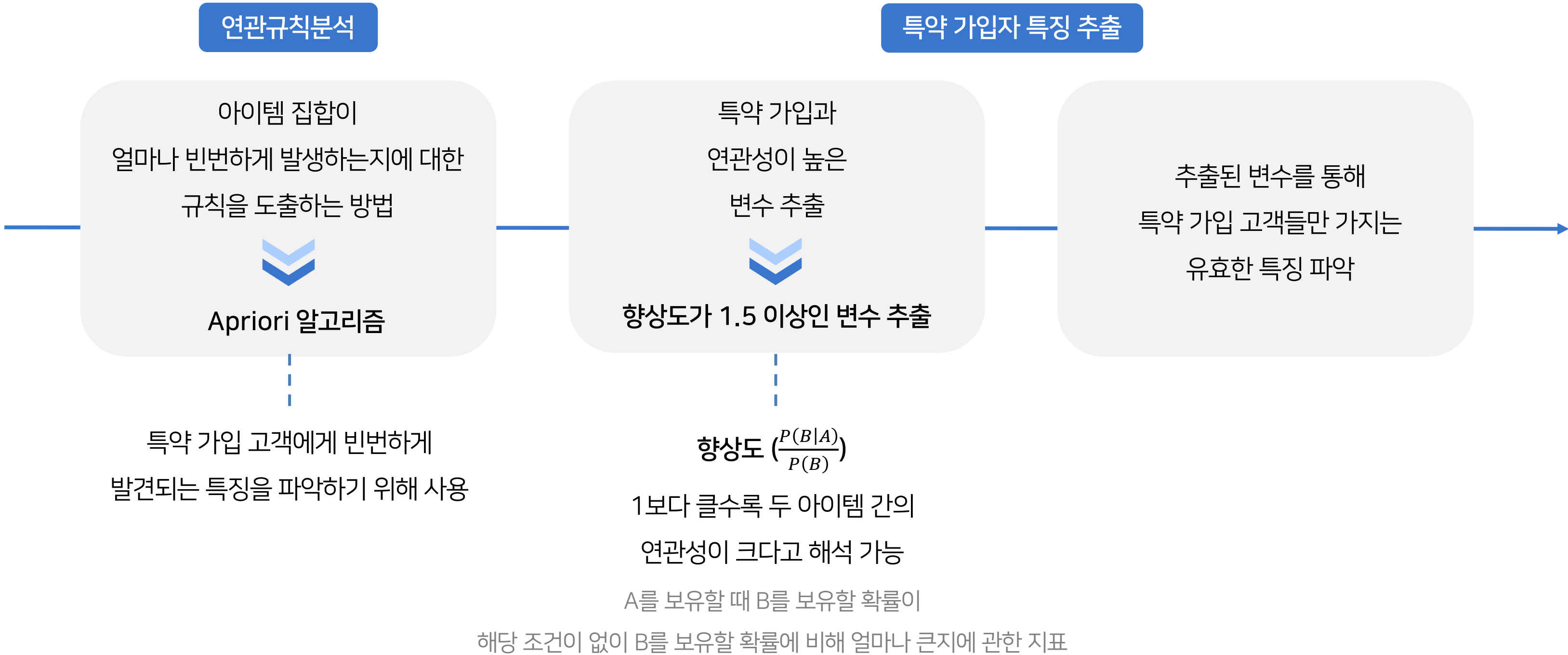
특약에 가입할 의사가 높은
고객군을 파악하기 위해
연관규칙분석을 통해
특약 가입 고객들이
공통적으로 갖는 특징 파악

2

특약 추천 대상인 고객 중 미가입 고객이
특약 가입 시 사고율이 유의미하게
감소하는지 검토하기 위해
대응표본 T-검정 시행

연관규칙분석

연관규칙분석을 통해 특약 가입자가 가진 특징을 파악해 해당 특징을 가진 고객들을 대상으로 특약 추천 진행



특약 가입고객 특징 파악

마일리지할인특약과 블랙박스특약 가입고객 특징을 연관규칙분석을 통해 파악

마일리지할인특약 가입고객 특징 변수

- 피보험자연령대 50대
- 차량경과년수 10년 초과
- 차종 대형 (배기량 2000cc 초과)

블랙박스특약 가입고객 특징 변수

- 피보험자연령대 60대
- 피보험자연령대 50대
- 차량경과년수 5년 이하(신차 제외)
- 차량경과년수 5년 초과 10년 이하

연관규칙분석 결과

- 피보험자연령대 50대, 배기량 2000cc 초과, 차량경과년수 10년 초과인 고객에게 마일리지할인특약 가입을 추천
 - 피보험자연령대 50-60대, 차량경과년수 10년 이하(신차 제외)인 고객에게 블랙박스특약 가입을 추천

연관규칙분석을 통해 파악한 **특약 가입 추천 대상 고객**들이 특약 가입했다고 가정했을 때 **사고율**이 **감소**하는 **고객**이 **다수 존재**하는 것을 확인

	특약 종류	가입 전 예측 사고율	가입 후 예측 사고율
고객1	블랙박스특약	3.4957	2.9143
고객2	블랙박스특약	2.3703	1.5501
고객3	마일리지할인특약	2.1779	1.1194



고객들의 사고율 감소가 통계적으로 유의한지에 대한 검토 필요

추천 시 사고율 변화 검정

연관규칙분석과 예측모델을 통해 특약 추천 대상 고객들의 사고율 감소가 통계적으로 유의한지 검토하기 위해 **대응표본 T-검정** 시행

대응표본 T-검정

- 동일한 객체에 대해서 처치 전과 후에 측정된 자료 간에 평균의 차이가 있는지 검정하는 방법
- $p\text{-value} < \alpha$ 이면 유의수준 α 하에서 귀무가설 기각한다고 해석

블랙박스특약 추천 시 사고율 변화 검정 결과

모든 고객을 블랙박스특약에 가입 고객으로 전환했을 때 사고율 변화와
블랙박스특약 가입 추천 대상인 미가입 고객을 가입 고객으로 전환했을 때 사고율 변화에 대한 대응표본 T-검정 시행해 검정 결과를 비교

검정 결과

모든 블랙박스특약 미가입 고객을 가입 고객으로 전환한 경우

H_0 : 블랙박스특약 가입 전과 후 사고율에 차이가 없다.
 H_1 : 블랙박스특약 가입 전보다 가입 후 사고율이 작다.

검정통계량	P-value
-0.35338	0.7238

검정 결과 해석

p-value > 0.1로 유의수준 0.1 하에서 귀무가설을 기각 불가능
즉, 블랙박스특약 가입 전에 비해
가입 후에 사고율이 유의하게 감소한다고 할 수 없음

검정 결과

블랙박스특약 가입 추천 대상인 미가입 고객만 가입 고객으로 전환

H_0 : 블랙박스특약 가입 전과 후 사고율에 차이가 없다.
 H_1 : 블랙박스특약 가입 전보다 가입 후 사고율이 작다.

검정통계량	P-value
2.1745	0.0149

검정 결과 해석

p-value < 0.1로, 유의수준 0.1 하에서 귀무가설을 기각 가능
즉, 블랙박스특약 가입 전에 비해
가입 후에 사고율이 유의하게 감소한다고 할 수 있음

마일리지할인특약 추천 시 사고율 변화 검정 결과

블랙박스특약과 같은 방법으로 마일리지할인특약 가입 추천 대상인 미가입 고객을 가입 고객으로 전환하여 사고율 변화에 대한 대응표본 T-검정 시행

검정 결과

H0: 마일리지할인특약 가입 전과 후 사고율에 차이가 없다.

H1: 마일리지할인특약 가입 전보다 가입 후 사고율이 작다.

검정통계량	P-value
1.4107	0.0792

검정 결과 해석

p-value < 0.1로 유의수준 0.1 하에서 귀무가설을 기각 가능
즉, 마일리지할인특약 가입 전에 비해
가입 후에 사고율이 유의하게 감소한다고 할 수 있음

할인특약 연관규칙분석 및 사고율 검정 결과

- 모든 특약 미가입 고객을 대상으로 사고율 변화 검정했을 때는 특약 가입 전과 후에 사고율 변화에 유의미한 차이가 없는 것처럼 비춰졌지만,
특약 가입 고객의 특징을 가진 미가입 고객을 선별했을 때 특약 가입 전과 후에 유의한 사고율 감소 확인
- 마일리지할인특약과 블랙박스특약 모두 가입 고객의 특징을 가진 미가입 고객들을 대상으로 한 타겟 마케팅을 통해 사고율 감소 효과를 기대

고객 맞춤 특약 추천 마케팅

각 고객의 사고율 예측과 연관규칙분석을 이용해 사고율 감소를 유도하는 할인 특약을 가입할 의사가 높다고 판단되는 고객군에게 해당 특약 가입을 추천하는 고객 맞춤 특약 추천 마케팅 제안

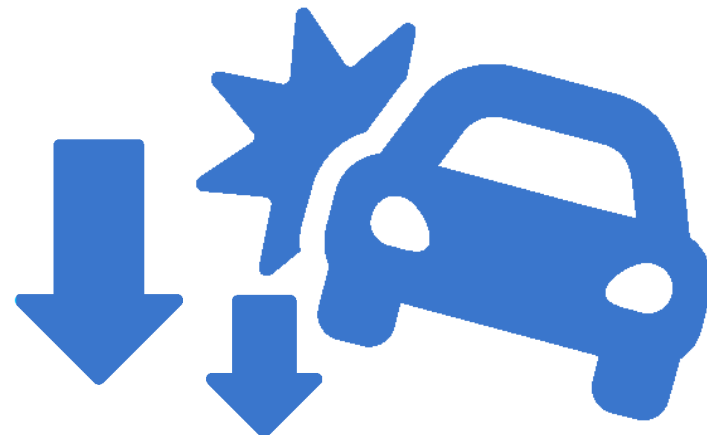
기대효과

마케팅 비용 절감 기대



모든 고객을 대상으로
추천하는 것이 아닌
특약 가입 의사가 높다고 예측되는
특정 고객군을 집중적으로 타겟팅

실질적 사고율 감소 효과



다수의 고객들이 특약을 가입했을 때
사고율 절감효과가 낮은 특약의 경우,
특약 시행에 따른 사고율 절감효과가
존재하는 일부 고객에게 선별적 추천

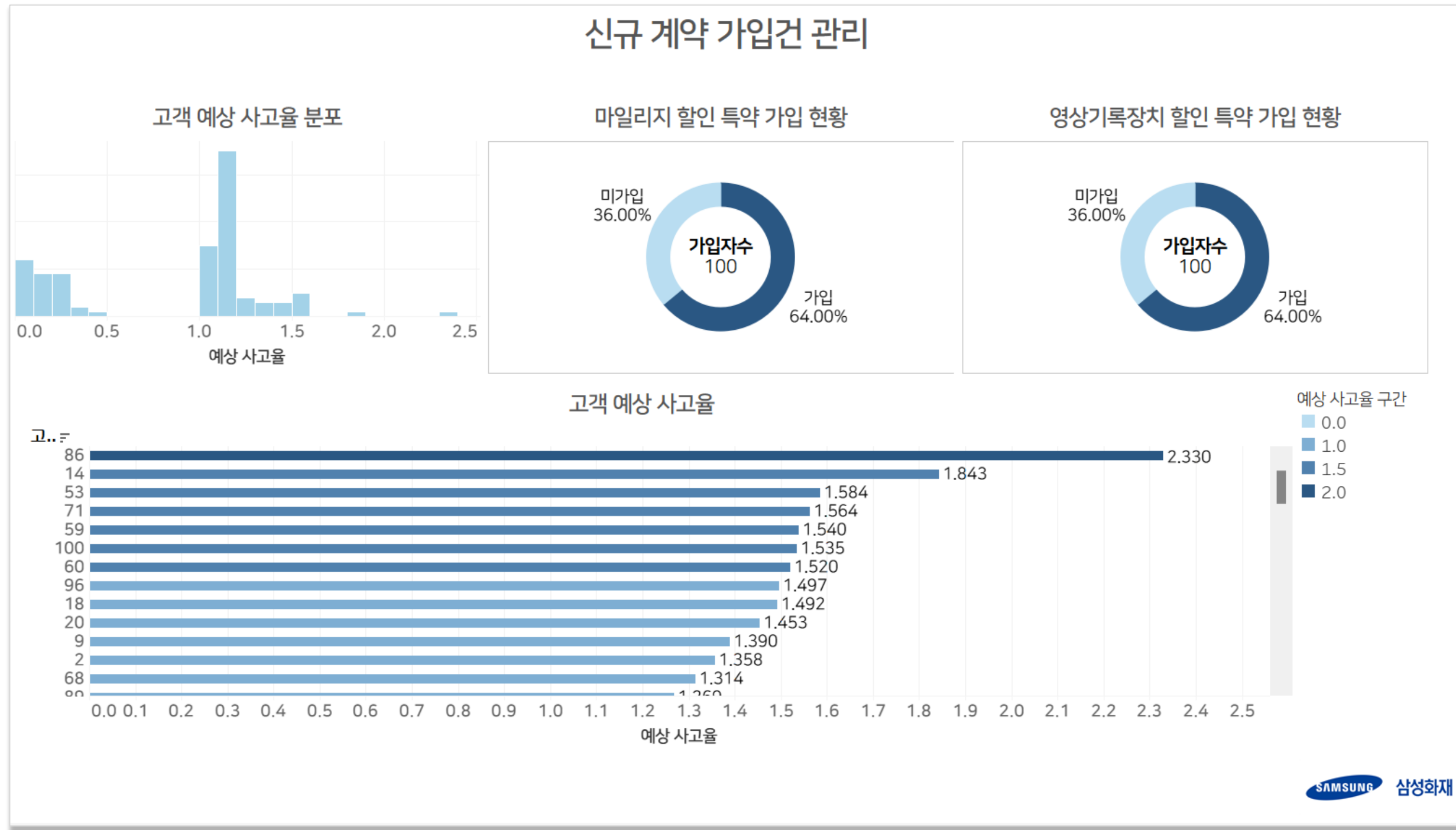
보험 가입 만족도 제고



보험사에서 개별 고객의 정보를 기반으로
고객이 선호할 것으로 예상되는
할인 특약과 혜택을
고객에게 맞춤적으로 제안

신규 고객 언더라이팅 고도화

- 사고율 예측과 사고 요인 분석을 통해 기존 계약 인수 심사에 고려되었던 사고 이력 외에도 **다양한 고객 정보를 보험 가입 승인 기준에 반영해 언더라이팅 고도화** 제안
- 계약 인수 담당자가 사고율 예측 모델링 결과를 쉽게 활용할 수 있는 **대시보드 활용** 제안

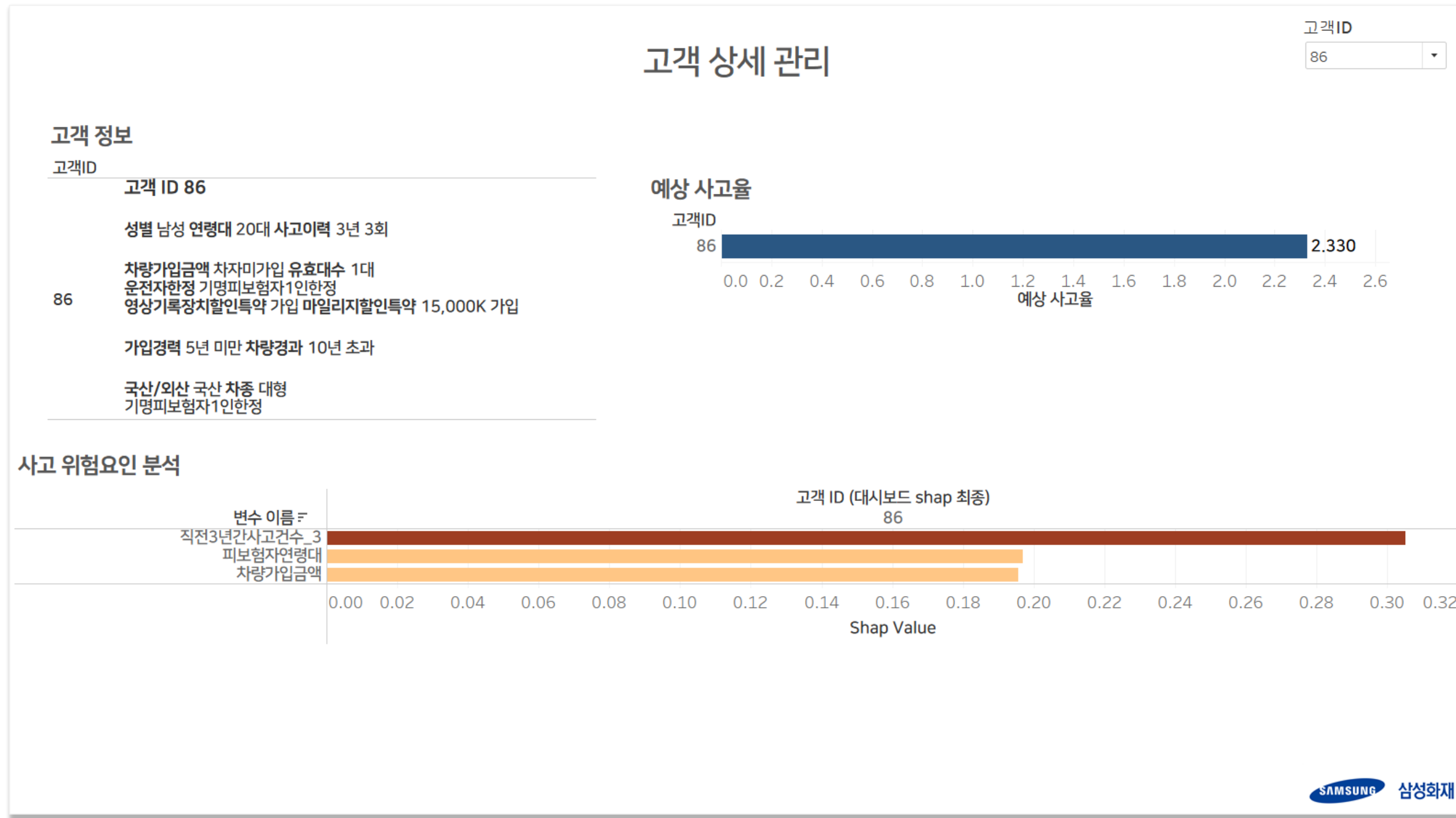


전체 신규 가입 고객 대시보드 구성 요소

- 고객 예상 사고율 분포
- 할인 특약 가입 현황
- 고객 예상 사고율 정보

신규 고객 언더라이팅 고도화

- 사고율 예측과 사고 요인 분석을 통해 기존 계약 인수 심사에 고려되었던 사고 이력 외에도 **다양한 고객 정보를 보험 가입 승인 기준에 반영해 언더라이팅 고도화** 제안
- 계약 인수 담당자가 사고율 예측 모델링 결과를 쉽게 활용할 수 있는 **대시보드 활용** 제안



신규 가입 고객 상세 대시보드 구성 요소

- 고객 정보
- 고객 예상 사고율
- 사고 위험요인 분석

신규 고객 언더라이팅 고도화

- 사고율 예측과 사고 요인 분석을 통해 기존 계약 인수 심사에 고려되었던 사고 이력 외에도 **다양한 고객 정보를 보험 가입 승인 기준에 반영해 언더라이팅 고도화** 제안
- 계약 인수 담당자가 사고율 예측 모델링 결과를 쉽게 활용할 수 있는 **대시보드 활용** 제안

기대효과

정교한 손해율 관리



단순히 과거 사고 이력만 아니라
다양한 고객 데이터를 기반으로 한
심사 시스템 구축

고객 만족감 증가



대시보드 및 사고 예측 결과를 활용해
가입 심사 과정 단순화를 통한
심사 소요시간 단축

심사 결과 신뢰성 제고



사고 요인 분석을 기반으로
고객에게 심사 결과에 대한
직관적인 근거 제시 가능

| 시사점 및 제언

1

카드사, 신용평가사 등의 다양한 금융기관과의
데이터 결합



사고율 예측에 활용되는 변수 다양화로
마케팅 아이디어 도출 및 예측 성능 개선 기대

2

AI 영상인식 기술 등으로 수집한
운전자별 주행습관 데이터를 활용한
초개인화된 사고율 예측 및 특약 가입 추천 마케팅



AI 영상인식 기술을 이용해 졸음운전, 전방주시 등
운전자의 주행습관 데이터 수집 및 정밀 분석을 통해
운전 상황별 보험료 할인 특약 적용 상품 개발

EX) BBI(주행습관기반) 특약

3세대 보험으로의 전환 기대

참고문헌

- 경향신문. (2023, 9월 11일). 자동차보험 상반기 손익 11% 감소...이동량 증가로 손해율 상승.
- 보험매일.(2023,10월 19일).지속 강화되는 차보험 소비자 혜택 '경쟁 심화'.
- 대한금융신문.(2020,3월 13일).삼성화재, 고연식차량 대상 블랙박스 할인 폐지.
- 기승도. (2023). 사고감소를 위한 자동차보험제도. KIRI 리포트 포커스.
- 최창희,홍민지. (2019). 빅데이터 활용 현황과 개선 방안. KIRI 보험연구원.
- 최양호, 이우주, 오승철 and 이동환. (2014). 일반화선형모형을 이용한 생명보험 지급금 분석: 암 발생 사고자료를 중심으로. Journal of The Korean Data Analysis Society, 16(6), 3093- 3106.
- 정중영, & 강중철. (2006). 자동차보험 손해율에 관한 연구. Journal of The Korean Data Analysis Society, 8(6), 2445-2456
- BizWatch. (2023, 4월 29일). [보푸라기]'보험 문지기' 언더라이터의 세계.
- 라이선스뉴스.(2022,4월 12일).“운전습관으로 보험료 측정”... BBI 보험, 車보험 高손해율 대안 될까?.

분석툴



python™



대시보드 링크

https://public.tableau.com/views/2023_16997846674410/sheet9?:language=ko-KR&:display_count=n&:origin=viz_share_link