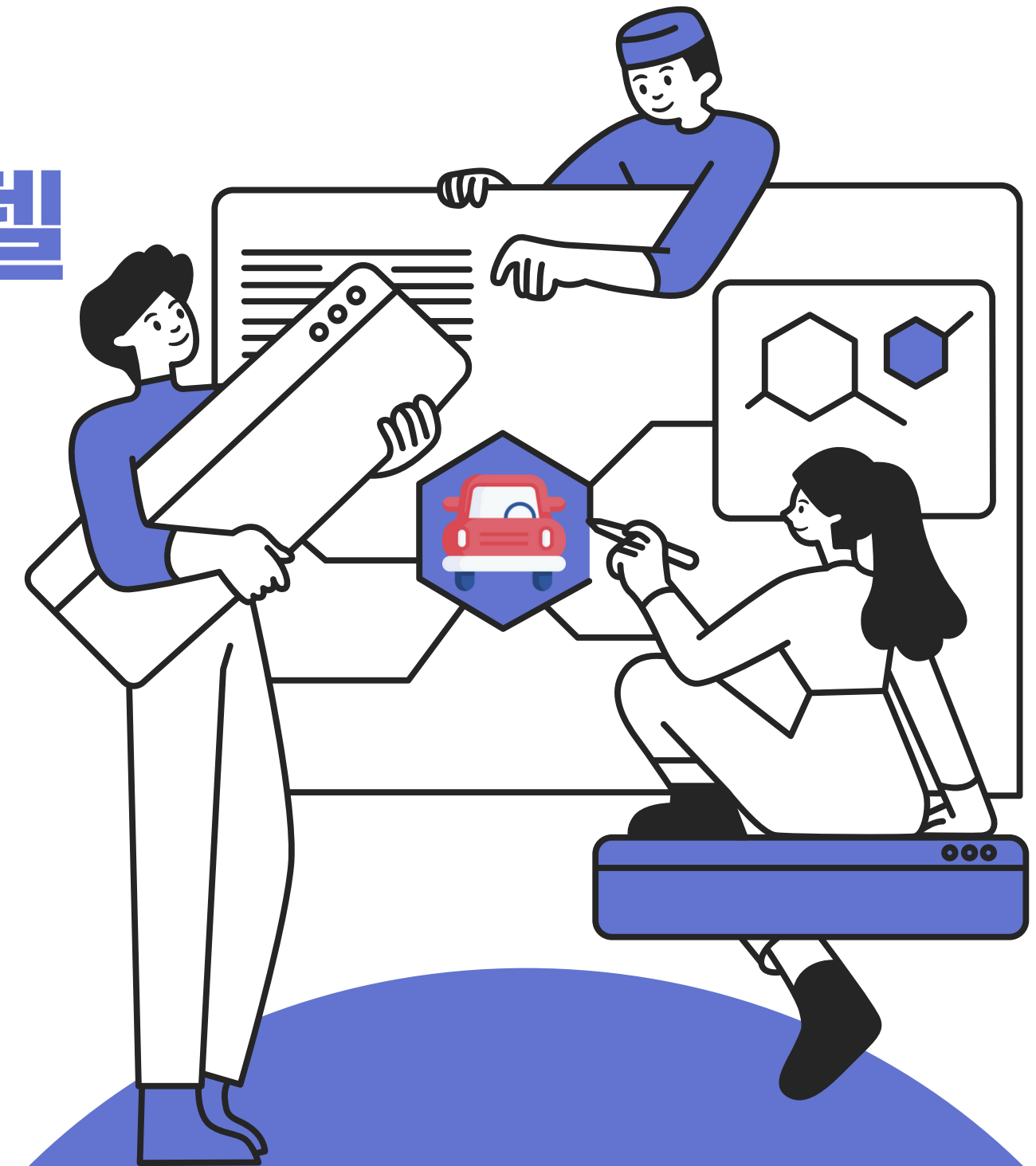


자동차 할부금융 이용고객 대상 연체예측 및 자체 신용평가 모델 개발을 통한 연체율 저감



C3 박경남 강민주 김영국 이정희 이찬영 이충현

CONTENTS

1 회사 소개

2 과제정의

01 추진배경

02 현상 및 개선기회

3 데이터 수집 및 정제

01 데이터 수집

02 데이터 정제

4 데이터 분석

01 분석계획

02 분석결과

5 개선안

6 배운점

01 회사소개

1. 회사소개



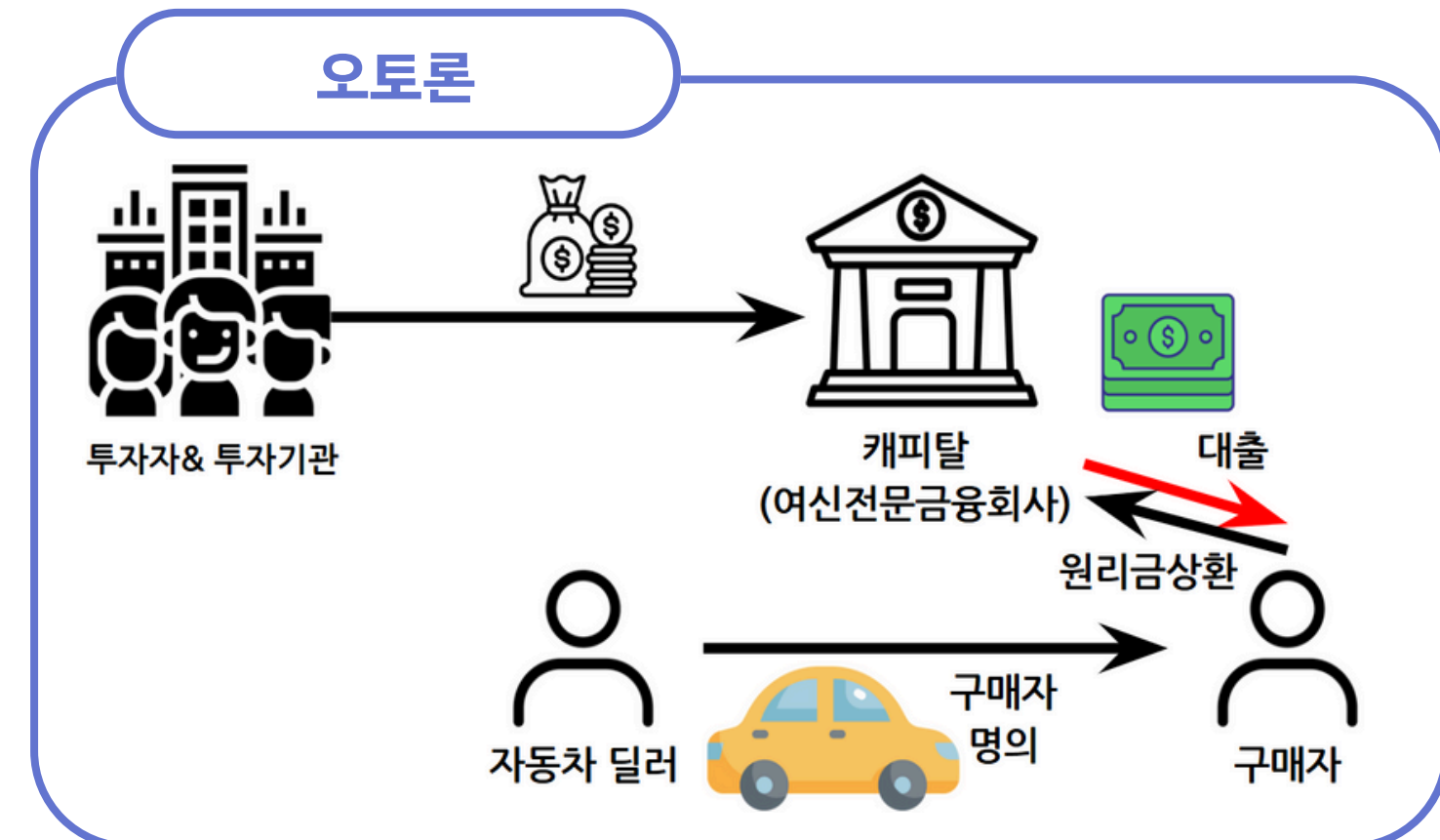
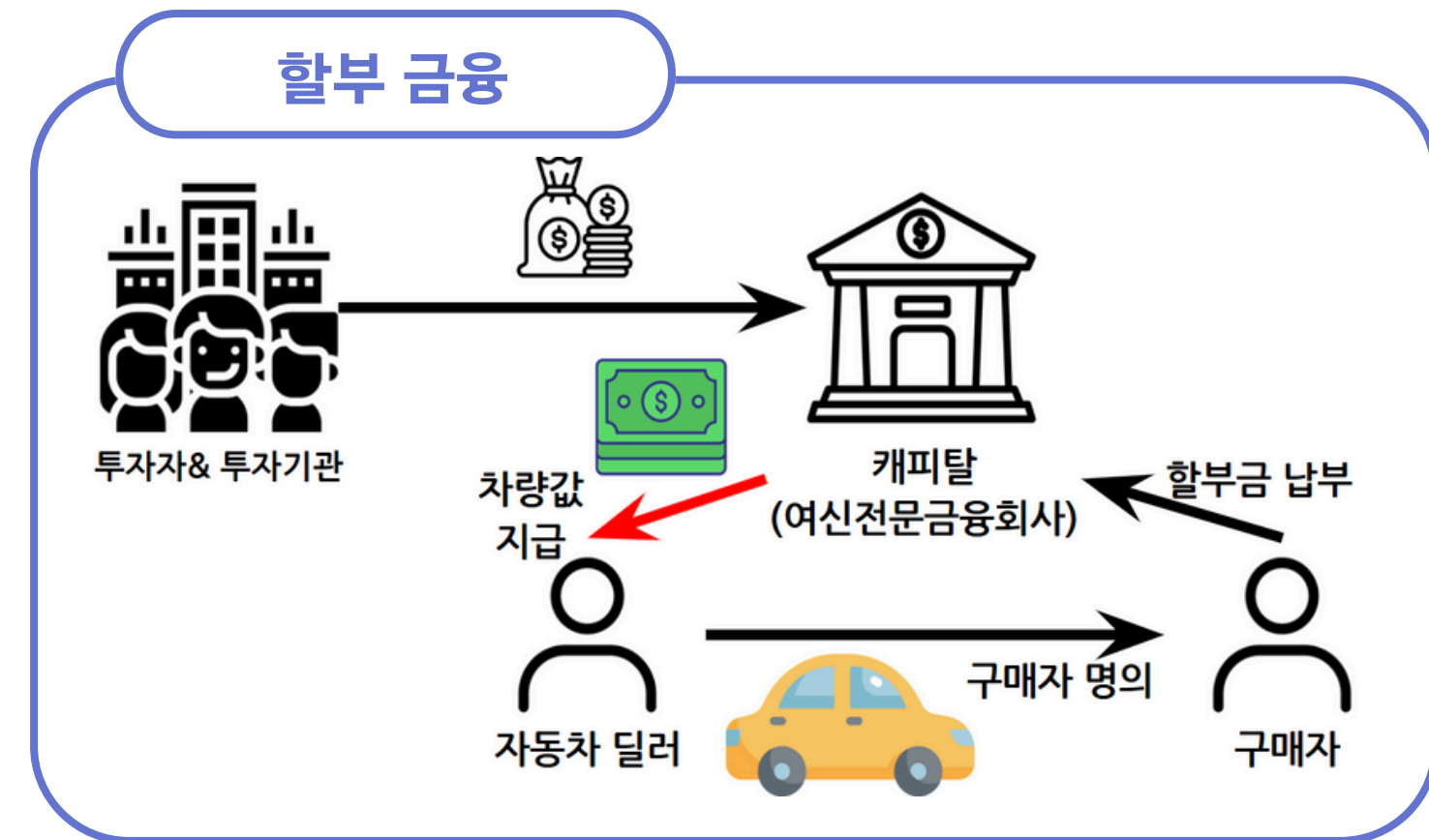
회사명 : 리본 (Re:Born) 캐피탈

설립일 : 2005년 07월 25일

대표이사 : 박경남

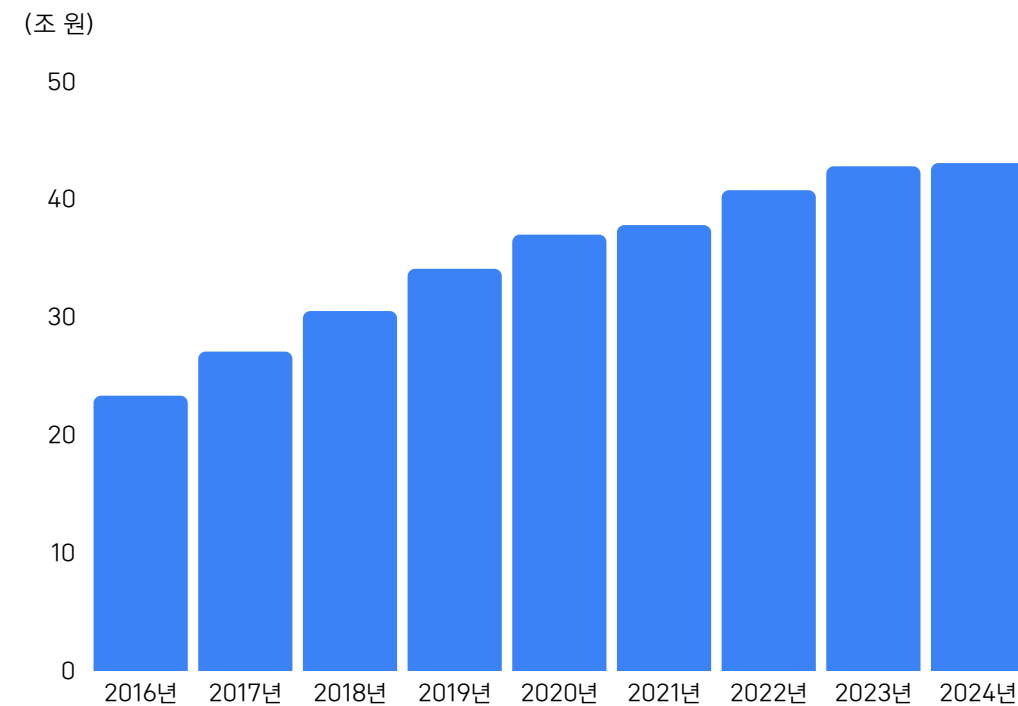
사업장 : 전국 (시도단위 지점)

주요 서비스: 할부금융 / 오토론



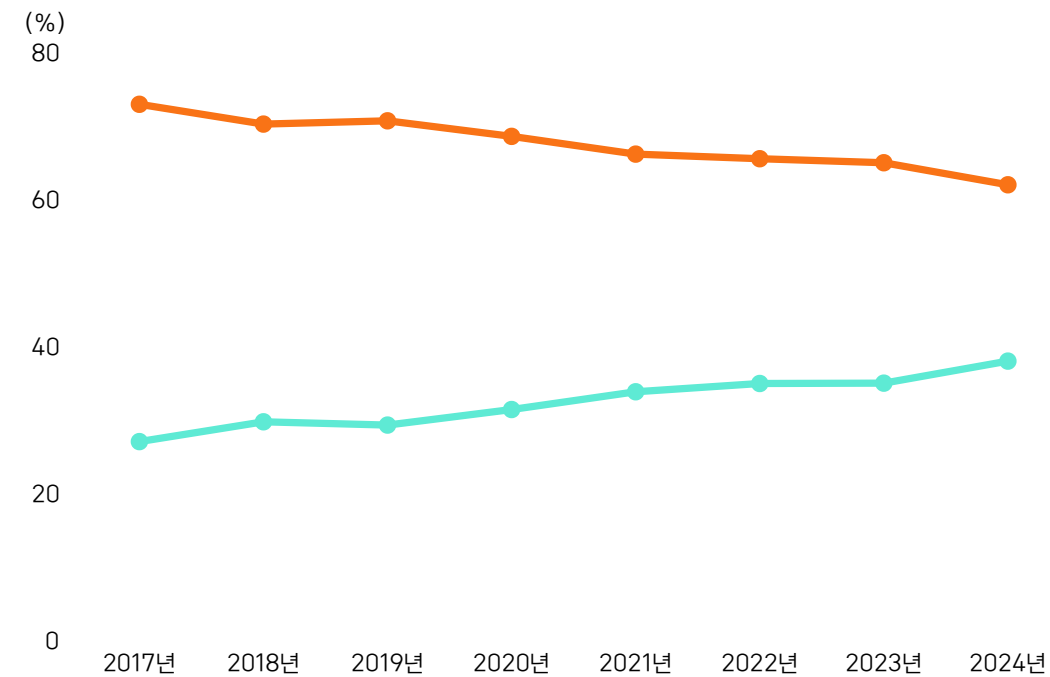
자동차 할부금융 시장은 성장하는 반면, 캐피탈사의 점유율 하락(21%p ↓)과 자사의 높은 연체율(7.97%)

자동차 할부금융 시장의 성장



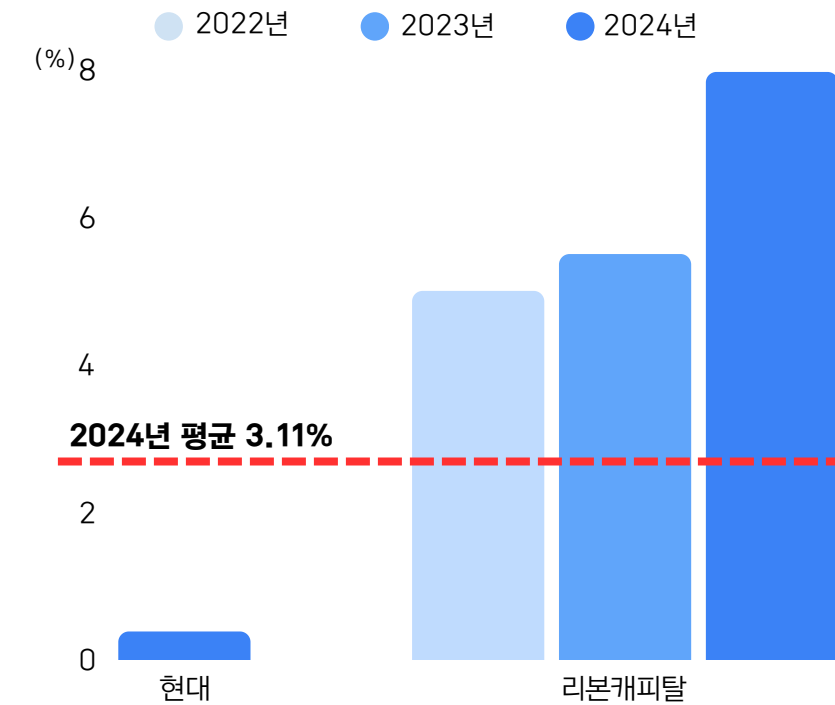
16년 대비 자동차 할부금융
시장 규모 85.3% 증가

캐피탈사의 점유율 하락



경쟁사 저금리 제시 → 캐피탈사 점유율 감소
16년 85% → 24년 64%로 지속하락

타사 대비 높은 연체율



현대캐피탈(0.38%) 대비 약 21배
업계 평균(3.11%) 대비 약 2.6배

“ 연체 고객 분석을 통한 연체율 저감 방안 수립이 필요함 ”

02 현상 및 개선 기회

2. 과제정의

연체 고객에 대한 특성 파악 미흡, 자체 신용평가 모델 부재, 신규고객 유치 활동 미흡으로 시장 경쟁력 확보 어려움

연체 고객에 대한 특성 파악 미흡

연체 고객이 가진 공통된 특성에 대한
파악이 미흡하여 위험관리가 어려움



리스크 예측부터 채권 회수까지
체계적 자산건전성 관리에 집중해
1.08%의 낮은 연체율 기록 중

→ 연체 예측 모델 도입을 통해
연체 고객에 대한 특성 파악

자체 신용평가 모델 부재

외부기관 신용평가 모델의
낮은 정확도(59%)로 인한 연체예측 한계



비금융 데이터를 활용한 AI 기반 자체
신용평가 모델을 구축해 싼파일러 대상
평가 정확도 강화

→ 캐피탈사의 業 특성이 반영된
자체 신용평가 모델 구축

신규고객 유치 활동 미흡

혜택 부족 및 높은 금리로 인해
신규 고객 확보 어려움



현대카드는 자사 차량 할부결제
도입으로 자동차 금융 시장 진입을
가속화 중

→ 고객 맞춤형 할인금리 및 프로모션
제공을 통해 신규 고객 유치 경쟁력 확보

개선 목표 : 연체 예측 및 자체 신용평가 모델 도입으로 25년 말까지 연체율 7.9% → 0.9%(7%p ↓)



고객 데이터

변수명	데이터명	속성
연령	Age	연속형
성별	Gender	범주형
수입형태	IncomeType	범주형
소득	Incomeclass	연속형
결혼여부	MaritalStataus	범주형
학력	Education	범주형
	...	

18개 변수 X 64,464 행



대출 데이터

변수명	데이터명	속성
대출유형	LoanType	연속형
대출기간	LoanTerm	연속형
잔여기간	LoanRemainTerm	연속형
대출금리	InterestType	연속형
대출금	LoanAmount	연속형
월납입액	InstallAmount	연속형
	...	

26개 변수 X 64,464 행

연체 예측을 위한 데이터

=



Total

44개 변수 X 64,464 행

근속연수의 유무 및 평균 금리, 평균 나이 등에서 재직자와 퇴직자의 특성이 상이하여, 집단 간 데이터 왜곡을 방지하고자 데이터를 분할하여 분석을 진행함

구분	재직자	퇴직자
근속연수	有	無
평균금리	4.69%	5.19%
평균나이	41세	57.9세
연체율	8.10%	5.41%

02 데이터 정제

3. 데이터 수집 및 정제

1. 재직자

- **근속연수 결측치 제거 → 2,770건 제거**
근속연수는 임의로 대체 시 분석 결과 왜곡 우려 있어 제거

2. 퇴직자

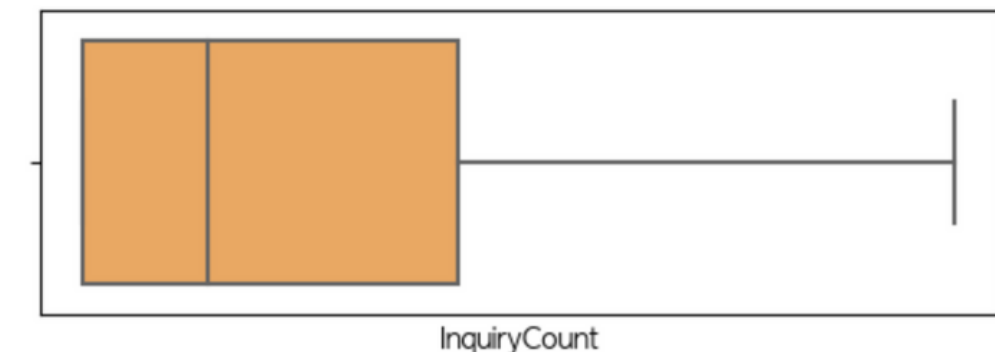
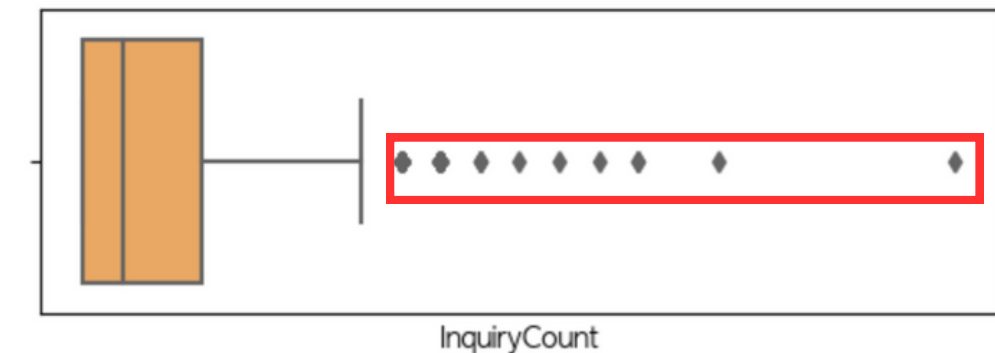
- **근속연수 결측치 제거 → 열 제거(전원 결측)**
근속연수는 퇴직자에게 해당되지 않아 제거

· 결측치 제거

주거등급, 주거형태, 신분증 변경 경과 기간, 전화번호 변경 경과 기간 결측치 제거
데이터의 결측 비율이 낮아(5% 미만) 결측값 제거

· 이상치 제거

대출 전년도 문의건수, 고객정보 변경 경과 기간 이상치 제거
이상치는 변수 간 관계 왜곡 및 모델 과적합을 유발할 수 있어 이상치 제거



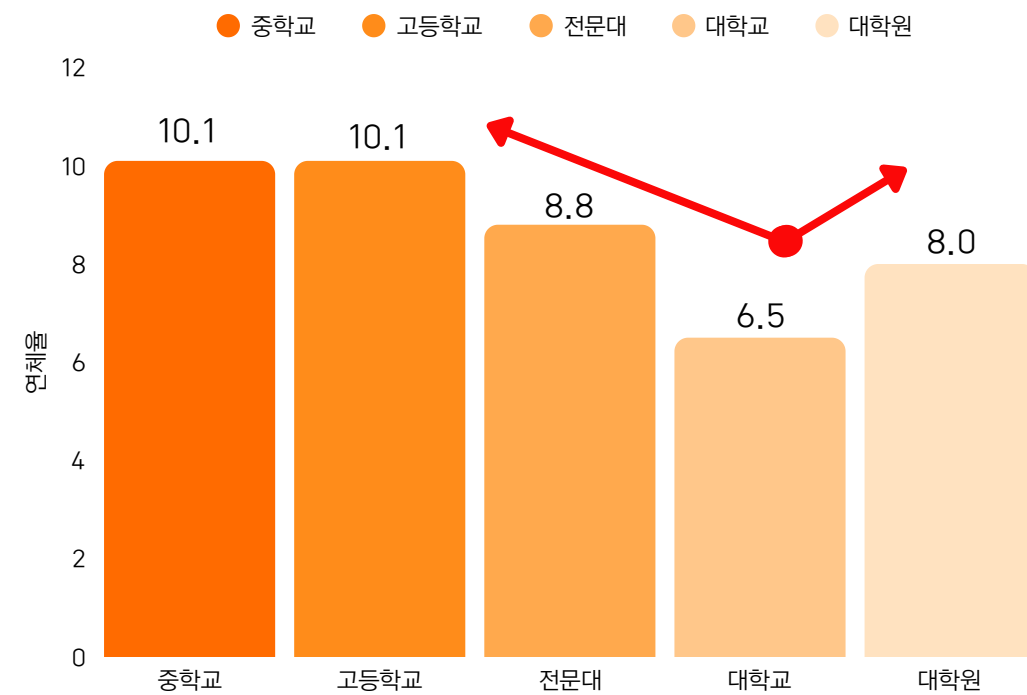
대출 전년도 문의건수

목적	분석방법	분석내용	담당자
고객 특성 파악	Histogram	고객 특성 기반으로 연체 예측의 주요 핵심 인자 파악	이충현
연체 예측 모델 개발	Decision Tree	고객의 다양한 특성을 바탕으로 신용점수에 영향을 미치는 주요 요인을 도출 및 이를 통해 신용 평가의 예측력과 정확도를 향상시킴	김영국 박경남
	CatBoost		
	LightGBM		
자체 신용평가 모델 개발	Histogram	기존 신용 점수와 연체 여부의 관계를 파악	이정희
	Gradient Boosting	고객 특성 기반으로 신용점수의 주요 핵심 인자 파악	강민주
신규고객 유치를 위한 고객 맞춤형 금융상품 추천	Histogram	자사와 타사의 신규 고객 유입 인원의 차이를 비교하기 위해 시각화	이찬영

02 분석결과_고객특성과악(재직자)

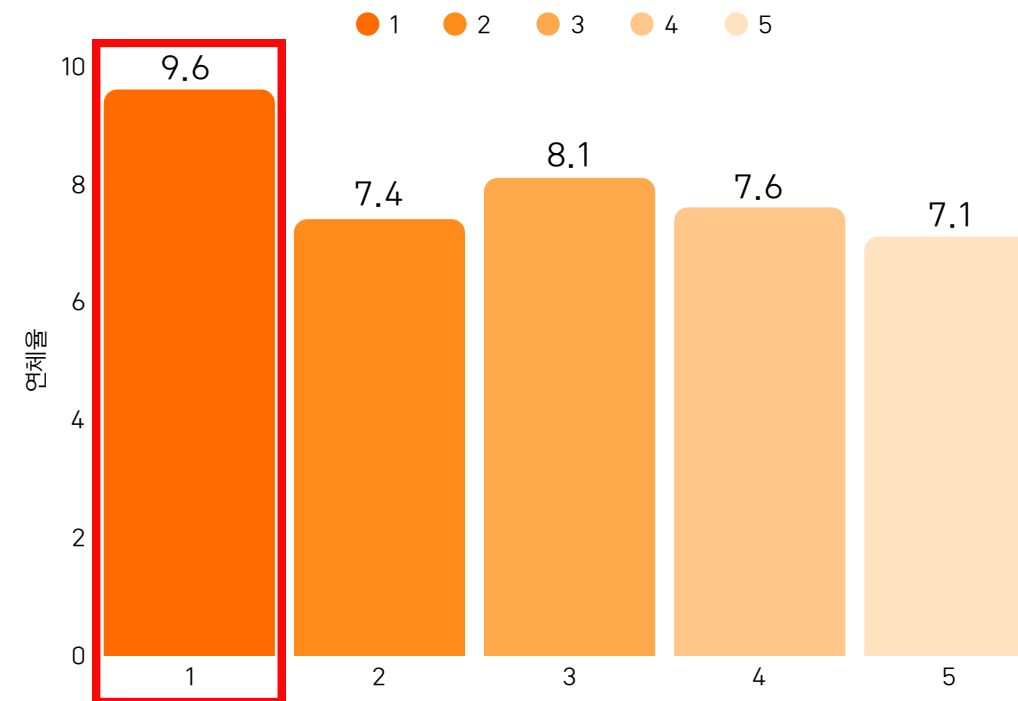
4. 데이터 분석

학력



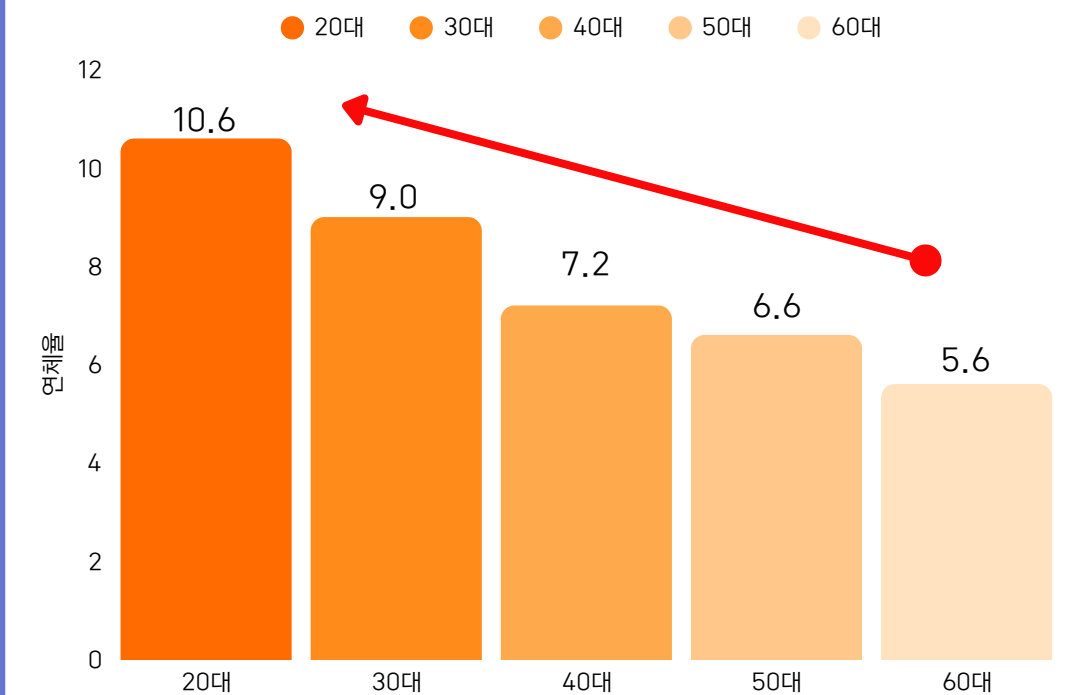
학력이 낮을수록 연체할 확률이 높다
대학원은 대학교보다 연체율이 높다

가족구성원 수



1인 가구가 연체할 확률이 높다

연령



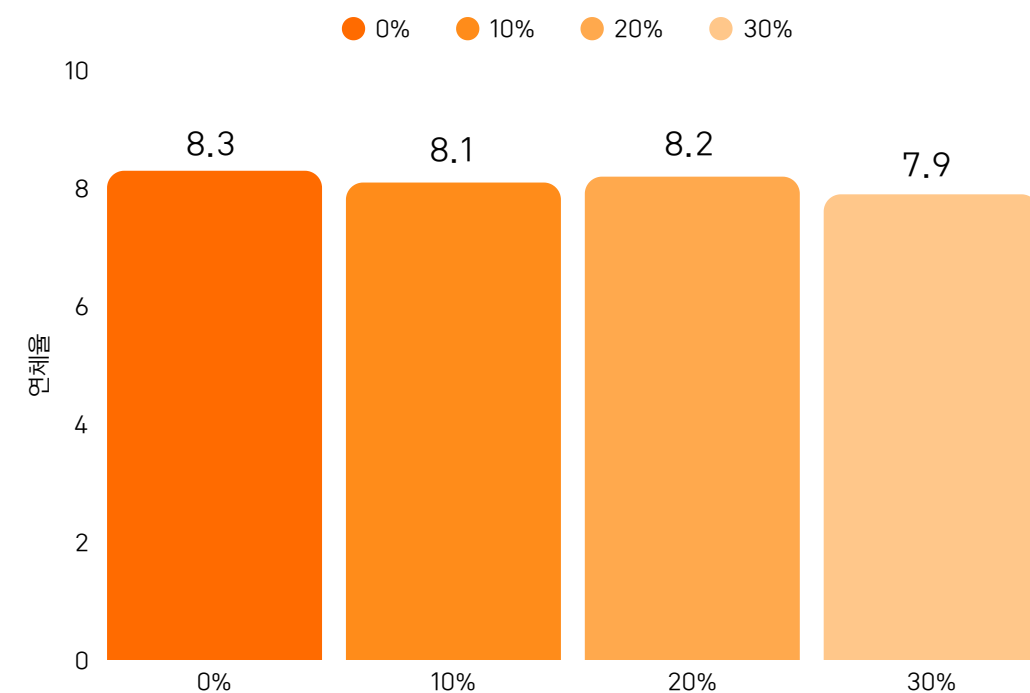
연령이 낮을수록 연체할 확률이 높다

1인가구는 대부분 사회 초년생이기에 연령과 연관지어 연체할 확률이 높다는 것을 알 수 있음

02 분석결과_고객특성파악(재직자)

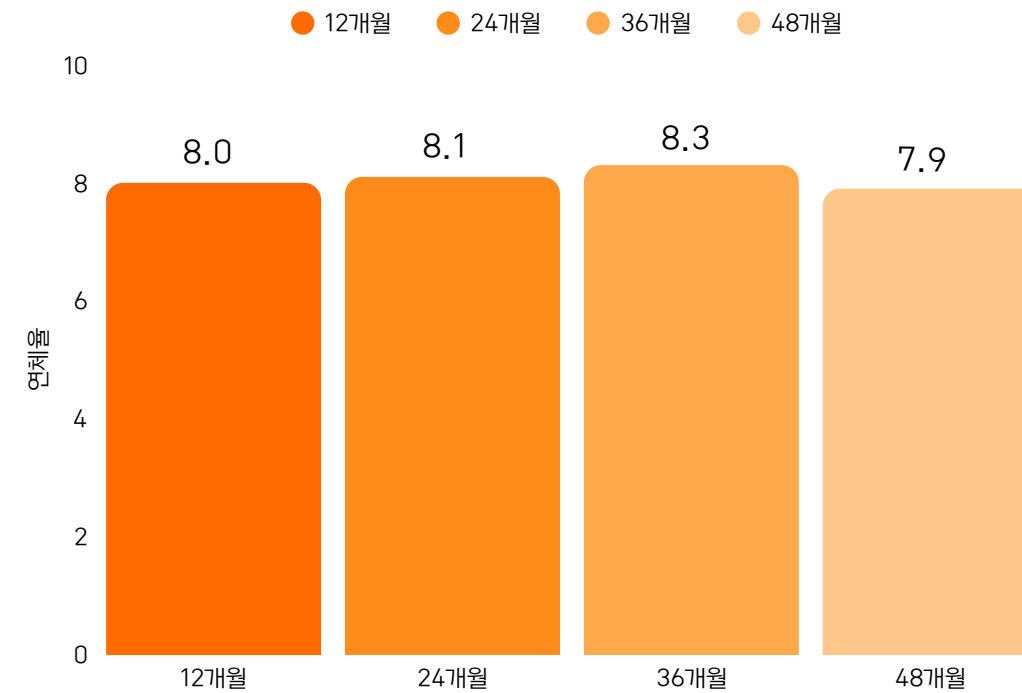
4. 데이터 분석

선수금 납부 비율



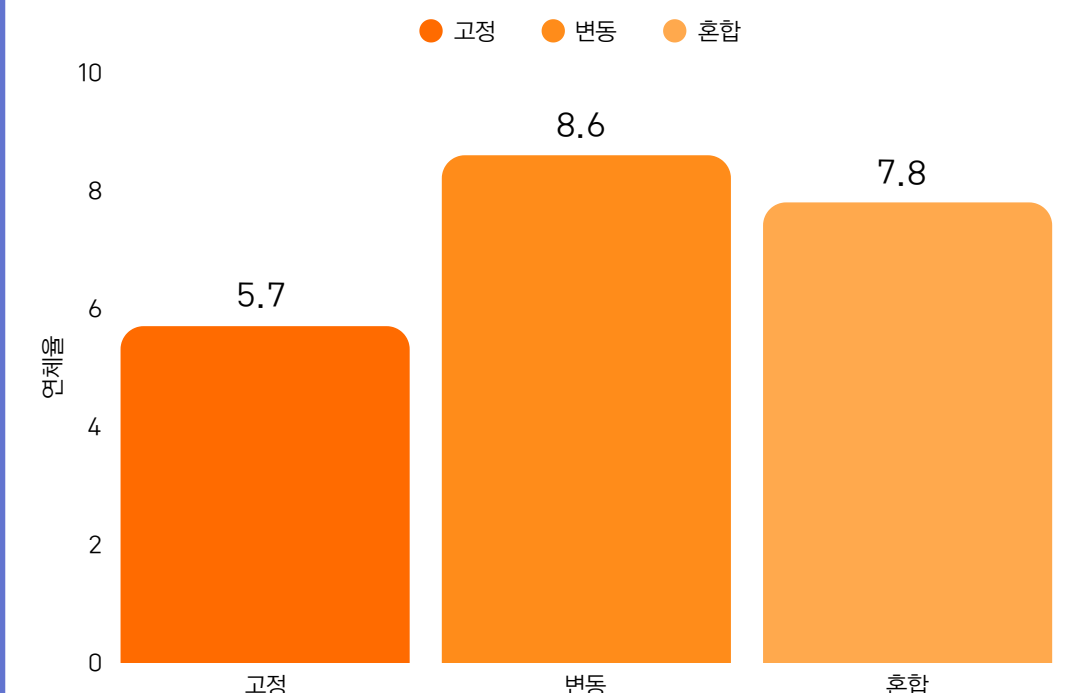
선수금 납부 비율이 연체율과는
관련이 없다

대출 기간



대출기간이 연체율과는 관련이 없다

금리 유형



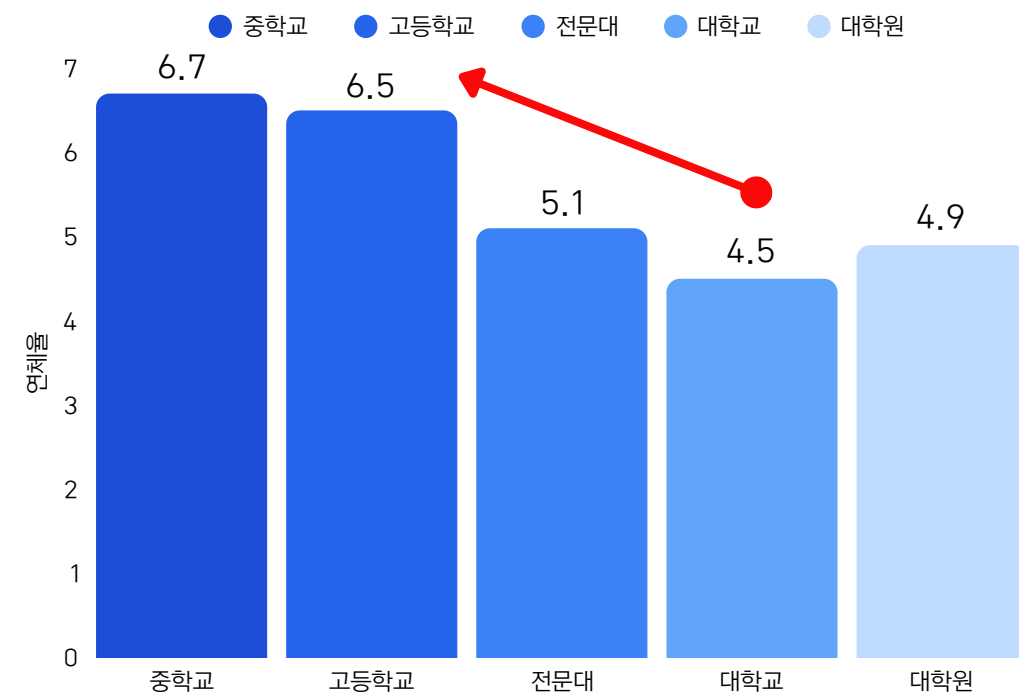
변동금리일수록 연체확률이 높다

재직자들은 고정금리보다 변동금리 선택 비중이 높아, 금리 인상 시 부담이 클 수 있음

02 분석결과_고객특성파악(퇴직자)

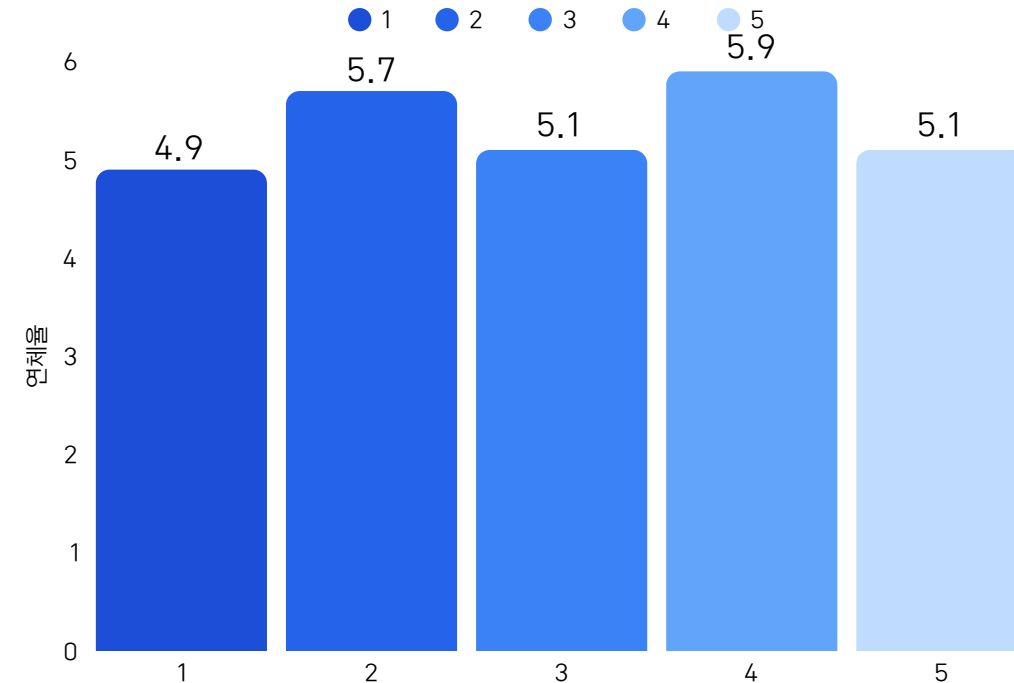
4. 데이터 분석

학력



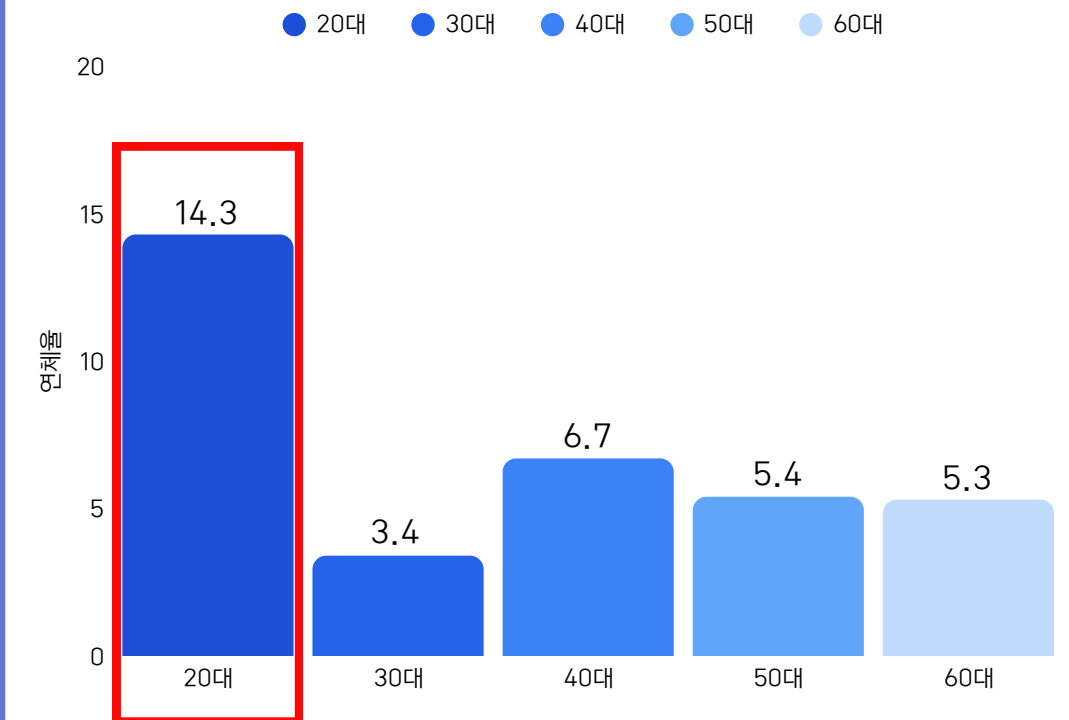
학력이 낮을수록 연체율이 높아지는 경향이 있다

가족구성원 수



가족 구성원 수는 2인, 4인이 연체율이 높다.

연령



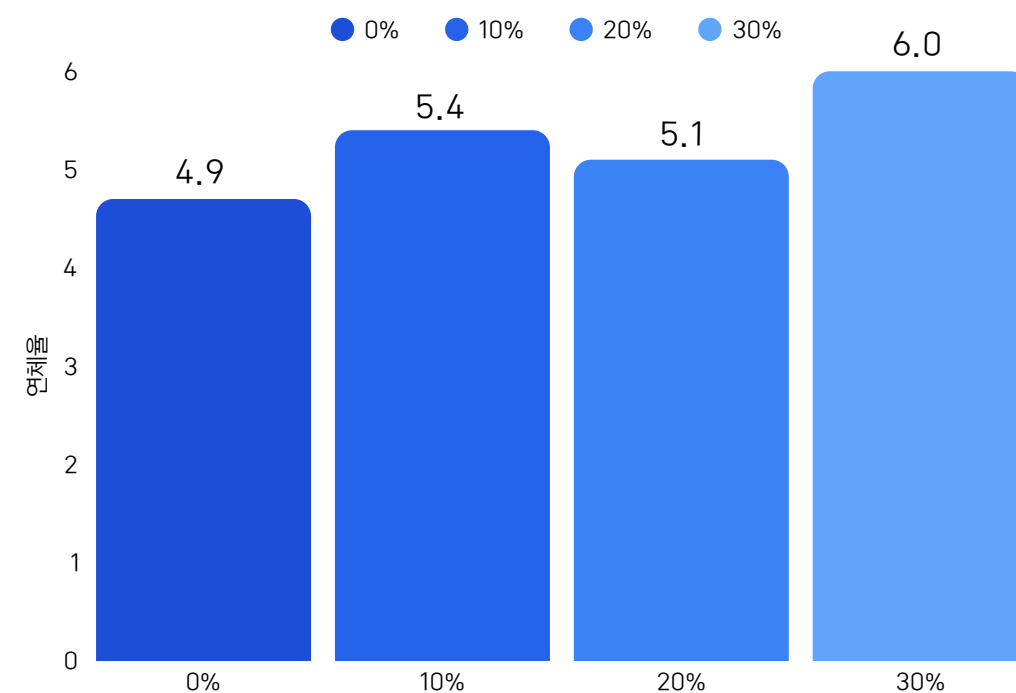
20대 연령층에서 높은 연체율을 보인다

퇴직자 중 20대는 소수에 해당하나, 연체율이 다른 연령대 대비 이례적으로 높음.

02 분석결과_고객특성파악(퇴직자)

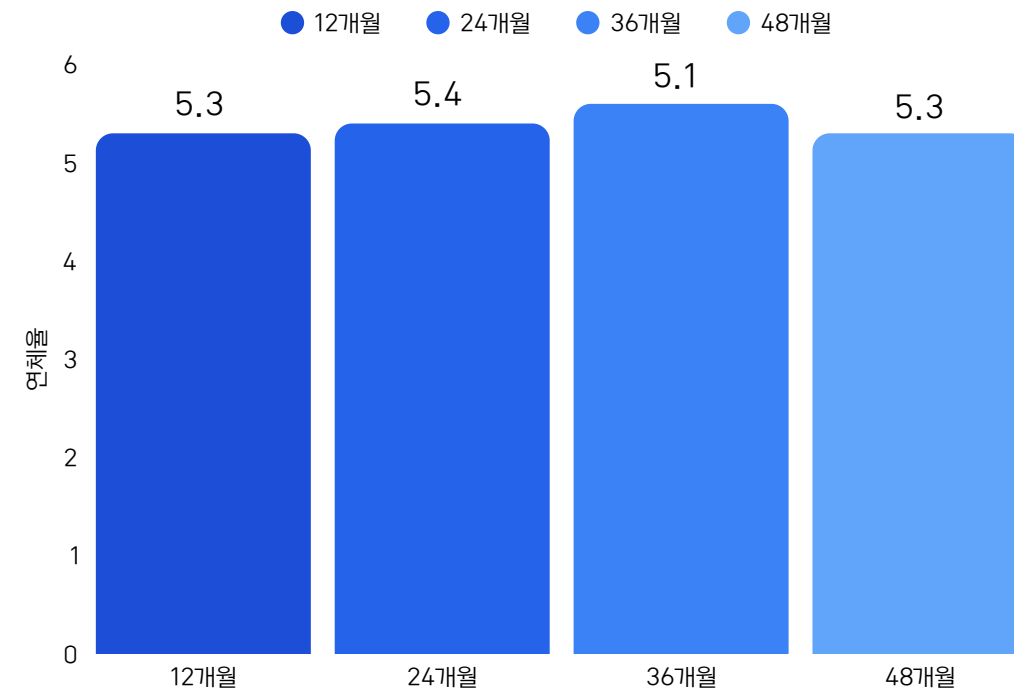
4. 데이터 분석

선수금 납부 비율



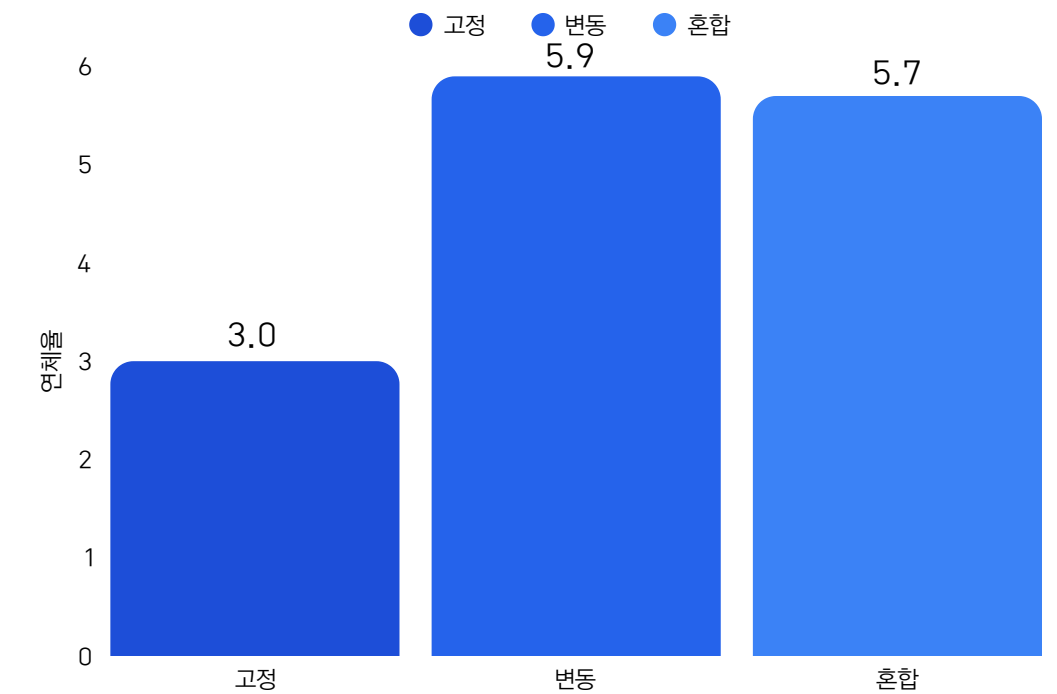
현금비율이 높을수록 연체 확률이 높다

대출 기간



대출 기간은 연체율과 관련이 없다

금리 유형



변동금리일수록 연체 확률이 높다

퇴직자들 또한 변동금리일수록 연체확률이 높기 때문에 대부분의 고객은 금리 인상시 부담이 클 수 있음

변수	유의성 검정		모델링					
	카이제곱	t검정	재직자			퇴직자		
			DT	CatBoost	LGBM	DT	CatBoost	LGBM
나이	-	o	4	6	1	4	5	1
학력	o	-	2	3		2	6	
수입형태	o	-						
근속연수	-	o	1	1	2	1	1	2
성별	o	-						
신분증 변경 후 경과기간	-	o		7	3	10	7	3
대출전년도 문의건수	-	o		8		9	8	
대출잔액	o	-	9		7	7		7
고객 정보 변경 후 경과기간	-	o	7		4	5		4
대출신청 차량 이외의 자동차 소유 여부	o	-	5					
결혼상태	o	-						
대출잔여기간	-	x						
대출금	-	x						
고객직업군	o	-		4	5		3	5
직장유형	o	-			6			6
주거등급	o	-	6	2		6	2	

→ 나이, 학력, 근속연수 외 6개 핵심인자 추출

재직자				샘플링여부	퇴직자			
DT	LightGBM	CatBoost	모델평가		모델평가	CatBoost	LightGBM	DT
0.94	0.93	0.94	정확도	Test	정확도	0.94	0.94	0.95
0.30	0.32	0.35	재현율		재현율	0.18	0.20	0.24
0.41	0.43	0.46	F1-score		F1-score	0.26	0.28	0.31
0.85	0.91	0.91	정확도	SMOTE 적용 후 Test	정확도	0.91	0.92	0.85
0.86	0.92	0.92	재현율		재현율	0.92	0.93	0.86
0.88	0.92	0.92	F1-score		F1-score	0.92	0.93	0.89

연체여부가 1인 데이터가 0인 데이터보다 더 적으므로 데이터 불균형이 발생함
이를 보완하기 위해 학습 데이터에 SMOTE 기법을 적용해 소수 클래스(연체자)를 증가시킴

02 분석결과_자체 신용평가모델 개발

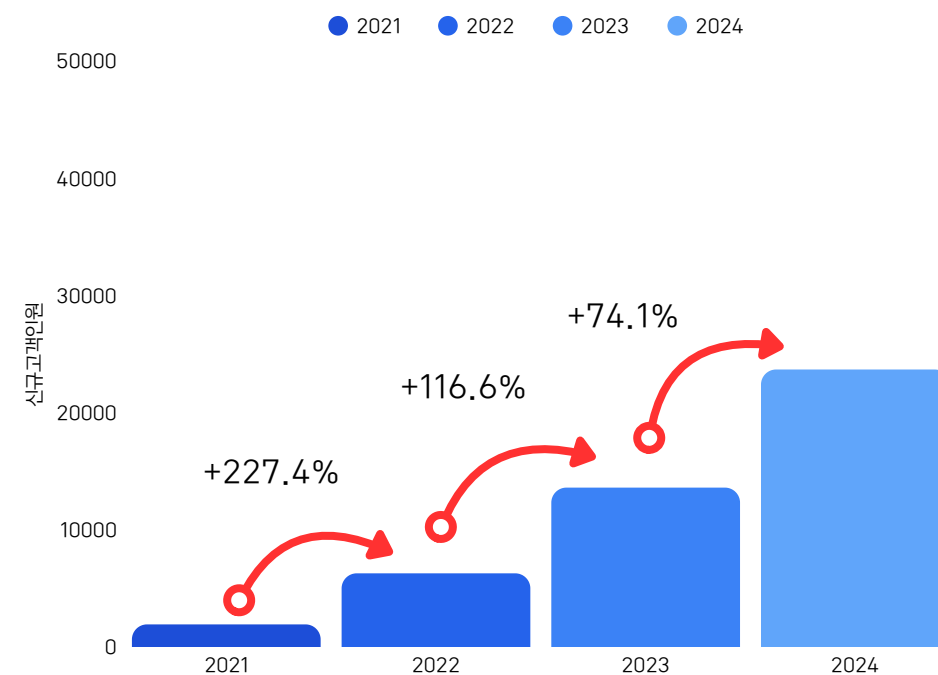
4. 데이터 분석

재직자			변수	퇴직자		
Score A	Score B	Score C		Score A	Score B	Score C
1	1	1	연령	1	2	3
2	2	2	주거등급	7	5	4
8	8	8	고객정보 변경 후 경과기간	2	4	5
4	3	3	휴대전화 변경 후 경과기간	4	3	1
10	7	6	대출전년도 문의건수	6	7	6
3	4	4	근속연수	-	-	-
7	-	-	성별	9	-	-
5	5	5	교육수준	8	8	9
연령, 주거등급, 고객정보 변경 후 경과기간, 휴대전화 변경 후 경과기간, 근속연수, 교육수준			종합 핵심 인자	연령, 주거등급, 고객정보 변경 후 경과기간, 휴대전화 변경 후 경과기간, 대출전년도 문의건수		

02 분석결과_신규고객 인원 비교

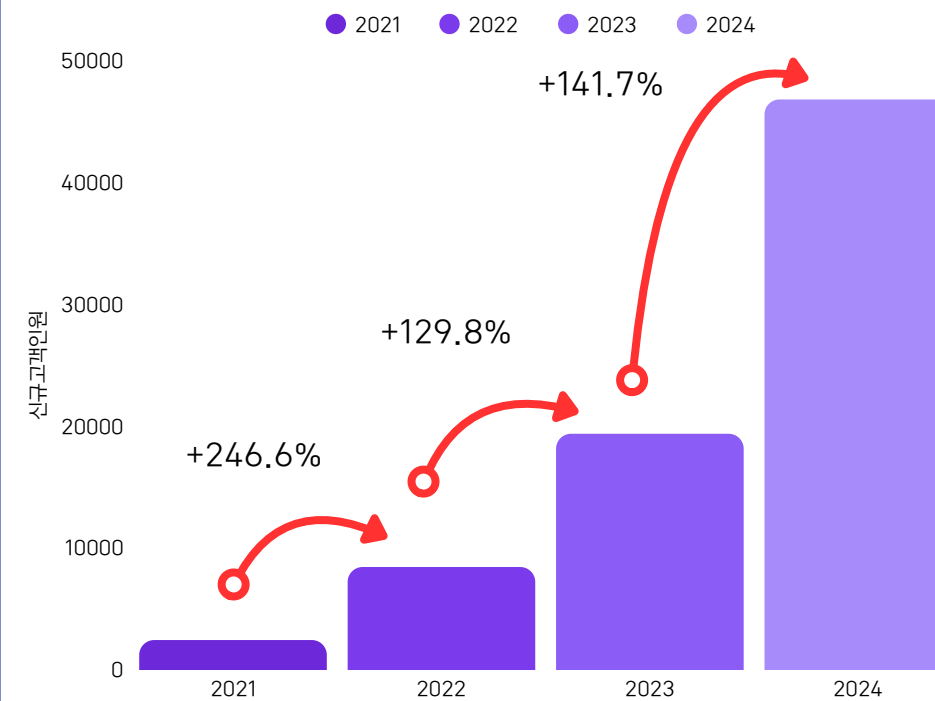
4. 데이터 분석

리본 캐피탈



고객 맞춤 프로모션 부재로
신규고객 인원 하락

현대카드



고객 맞춤형 할인금리 제시로
신규고객 인원 상승

현황분석

신규 고객 수 증가

→ 현대카드 대비 유입 수준은 낮음

문제인식

현대카드 대비 신규 유입 수 낮음

→ 프로모션 경쟁력 부족으로 판단

개선 방향

고객 맞춤형 프로모션 도입

→ 신규 고객 유입 수 향상 기대

1. 변수 가중치

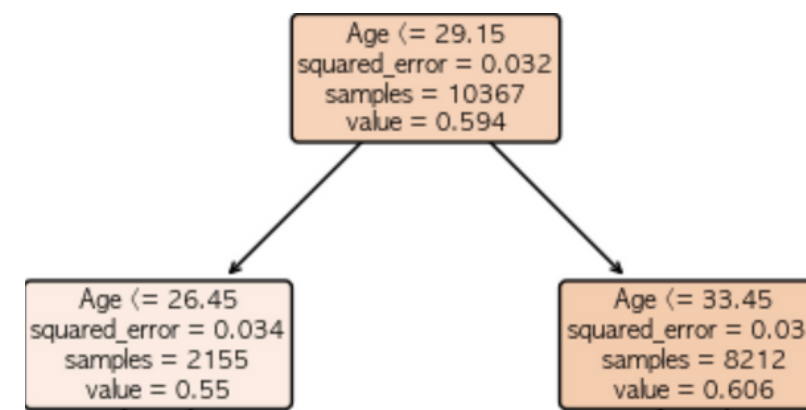
DT 모델을 활용한 Y 목표변수(Score_Re) 예측
변수 중요도 확인 후 변수 10개 선정, 비율 선정(합계 1)
각 변수에 대한 가중치

$$\text{Score_Re} = w1*x1 + w2*x2 + \dots w10*x10$$

변수	값
Score_Re	0.345
	0.572
	0.721
	0.654

2. 값 가중치

1에서 선정된 변수 1개씩 Y 변수(Score_평균) 예측 실시
DT 모델 확인하여 값 중요도 확인 후 가중치 부여



변수	값	가중치
나이	20대	0.3
	30대	0.4885
	40대	0.4718
	50대 이상	0.4572

3. 개선안 결론

- 자회사만의 신용 평가 모델 구축
- 모델에서 도출된 핵심 인자뿐만 아니라 비금융 정보 요소도 추가할 필요 있음
- 자체 모델 개발 초기 상태이기에 타 신용평가사와 병행이 필요함

예측 확률 기반 고객 신용등급 분류

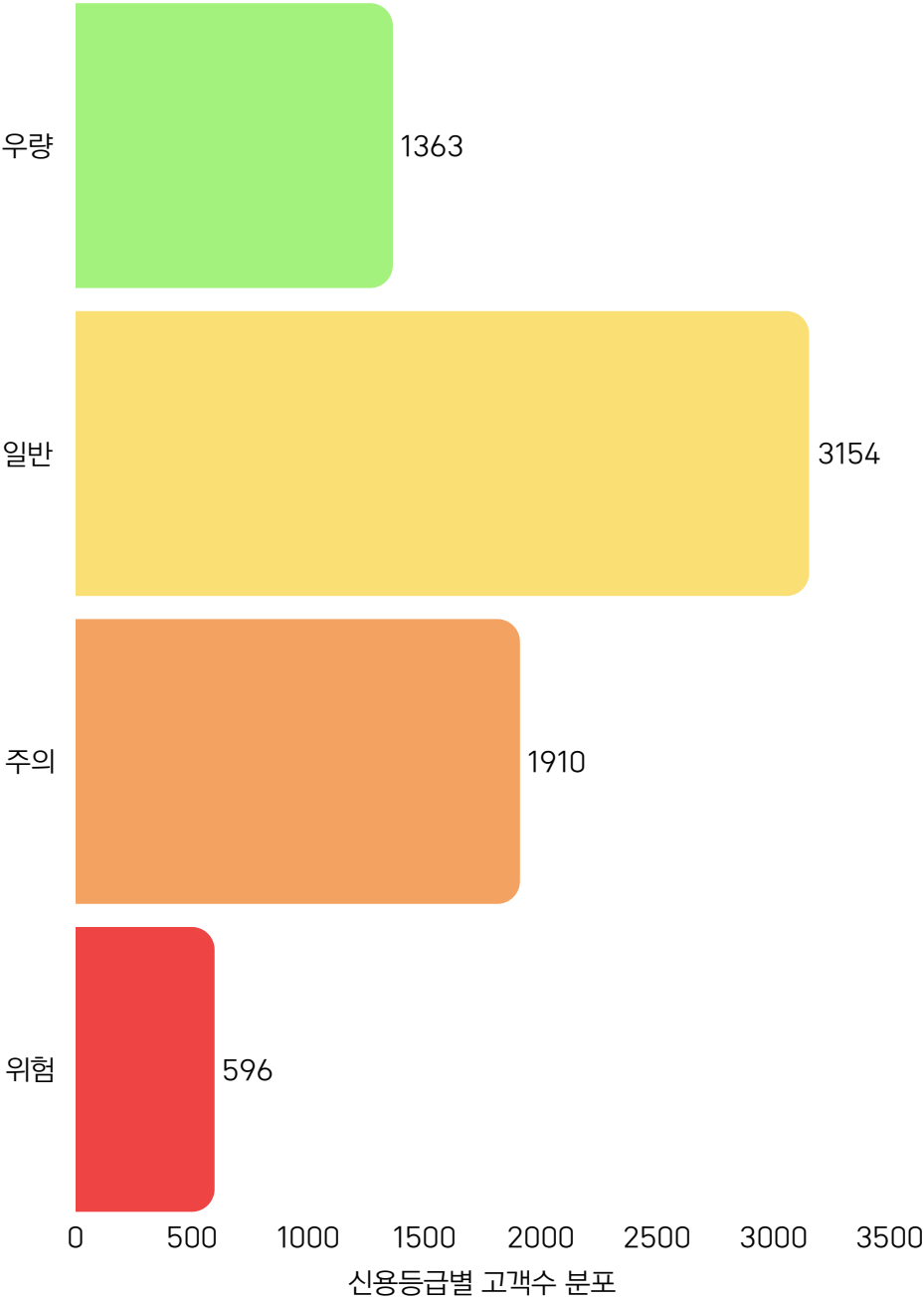
고객 신용 위험을 수치로 정량화
등급별로 우량~위험까지 구분 가능

예측 확률(p)		
0	0	0.052205
1	0	0.081020
2	0	0.081020
3	0	0.108303
4	0	0.051140
...
7406	0	0.052205
7407	0	0.051465
7408	0	0.023560
7409	0	0.031746
7410	0	0.052205

점수 환산 방법

$$\text{Score} = (1-p) \times 1000$$

등급	신용평가 점수	신용자 구분
1등급	942~1000	우량
2등급	891~941	
3등급	832~890	
4등급	768~831	
5등급	698~767	일반
6등급	630~697	
7등급	530~629	주의
8등급	454~529	
9등급	335~453	위험
10등급	0~334	



① 고객 맞춤형 할인금리 제시

신용구분별 금리 범위 및 평균금리

금리 범위(평균금리) [단위: %]	재직자	퇴직자
우량	2-7.43(4.69)	2-7.43(5.18)
일반	2-7.43(4.58)	2-7.43(5.07)
주의	2-7.43(4.70)	2-7.43(5.19)
위험	2-7.43(4.79)	2-7.43(5.30)
신용구분 전체 금리 평균	4.69	5.19

신용구분별 차등 금리 적용

우량

할인 금리 추천

저금리 확정 프로모션

리본 VIP Prime 고객 특별 할인금리 이벤트

- 최저 금리 2.6~3.0% 보장
- 추가 혜택: "첫 달 무이자", "중도상환 수수료 면제"

*카드사 최저, 최대 금리(업계 평균)

최저: 2.61%

최대: 4.59%

일반

평균 금리 추천

조건형 금리 인하 프로모션

리본 성실 납부 고객 금리 우대제

- 기본 금리(평균 4.58, 5.07%)에서 -0.5% 인하
- 추가 혜택: 1년 후 금리 재조정 보장

*개피탈사 최저, 최대 금리(업계 평균)

최저: 3.84%

최대: 5.37%

주의

높은 금리 및 고정금리 추천

성실 상환 리워드형 금리 프로모션

리본 올라가자! 금리 할인 챌린지

- 6개월 연속 상환 성공 시, 금리 -0.3%
- 누적 상환 성과에 따라 금리 누적 할인

위험

고금리 제시 혹은 대출 제한

② 중고차 제휴시장 할인금리 제시



중고차 회사인 POSCAR와 제휴 협약을 통해 신규고객 유치

[상호 혜택 구조]

POSCAR → 리본캐피탈: 할인금리·보증연장·사은품 제공 → 고객 유치 강화

리본캐피탈 → POSCAR: 금리 프로모션 → 판매량 증대

자동차

신용평가

박경남

도메인을 이해하고, 데이터를 분석해 인사이트를 도출하는 전 과정을 직접 경험해볼 수 있어 유익한 시간이었습니다. 또한 보고서 작성과 결과 해석의 중요성을 느끼며, 빅데이터 분석의 흐름을 체득할 수 있었습니다.

이정희

이번 프로젝트를 통해 데이터에 대한 이해뿐만 아니라 실무적인 문제를 풀어가는 사고력을 키울 수 있었습니다. 끝까지 함께 달려온 시간들이 의미있게 남을 것 같습니다.

김영국

금융에 대한 도메인을 공부해보고, 데이터 분석 모델 성능 개선을 위해 수차례 반복 테스트를 거치며 실전 감각도 키울 수 있었고, 그 과정에서 팀원들과의 협력은 큰 힘이 되었습니다. 서로의 아이디어를 공유하고 부족한 부분은 함께 채워가며, 단순한 분업이 아닌 진짜 협업을 경험할 수 있었습니다.

할부금융

이충현

도메인을 이해하고 문제를 정의, 데이터 분석의 한 흐름을 경험하며 각 단계의 중요성을 알 수 있었던 시간이었습니다. 바로 분석으로 가는 것이 아닌 변수를 이해하고 도메인을 연결하는 과정이 매우 중요하다고 느꼈습니다.

강민주

비즈니스 상황과 도메인에 대한 이해를 바탕으로 문제 해결 방향을 설정하는 것이 중요하다는 것을 배웠습니다. 데이터 분석 이전의 기획 단계가 프로젝트의 흐름과 성과를 좌우한다는 점을 체감했습니다. 또한 서로의 강점을 이해하고 조화를 이루는 협업이 프로젝트 완성도를 높인다는 것을 느꼈습니다.

이찬영

빅데이터 프로젝트를 통해 문제 정의부터 데이터 전처리, 모델 평가까지 전 과정을 경험하며 데이터 기반 문제해결력과 협업 능력을 키울 수 있었습니다.

리본캐피탈

THANK YOU
