

인터넷전문은행 출범 후
유입 고객 및 비유입 고객 분석
- 소액 대출을 중심으로

2019.9.3.

KCB 인턴 4조

목 차

I. 서론

II. 주요 분석 내용

(1) 소액대출 시장의 현황

(2) 인터넷전문은행 출범 후 유입된 고객

(3) 유입 고객과 비유입 고객의 특징

III. 요약 및 결론

I. 서론

◆ 분석 계기

인터넷전문은행 카카오뱅크 1000만 고객 달성

27일 출범 2주년 앞두고 의미있는 성적
젊은 고객 주류...여수신 규모도 급성장

제3·4 인터넷은행 2020년 출범...인가 티켓 거머쥘 곳은 어디?

3분기 중 인가 재추진..."토스·키움 뿐 아니라 새 참여자 가능"

카카오뱅크 대출건수 절반은 '비상금 대출'

대출건수 52.7%가 50만~300만 마이너스대출

20대 대출건수 비중 17%, 대출액 기준은 6%

I. 서론

◆ 소액대출 분석 계기

대출 규제에 '모바일 소액대출' 눈 돌리는 시중은행

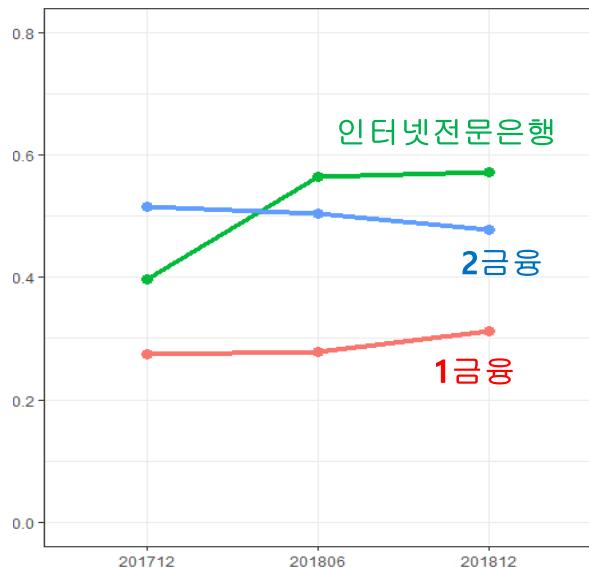
젊은 고객 뺏길라...'소액대출' 판 키우는 은행

출시일자	은행	상품	한도
2017/07/26		KB리브 간편대출	300만원
2017/09/18		쓸편한 포켓론 (소액 새희망 홀씨)	500만원
2018/09/18		하나핀크 비상금대출	50만원
2019/05/28		쓱쓱 간편대출	500만원
2019/07/11		우리 비상금대출	300만원

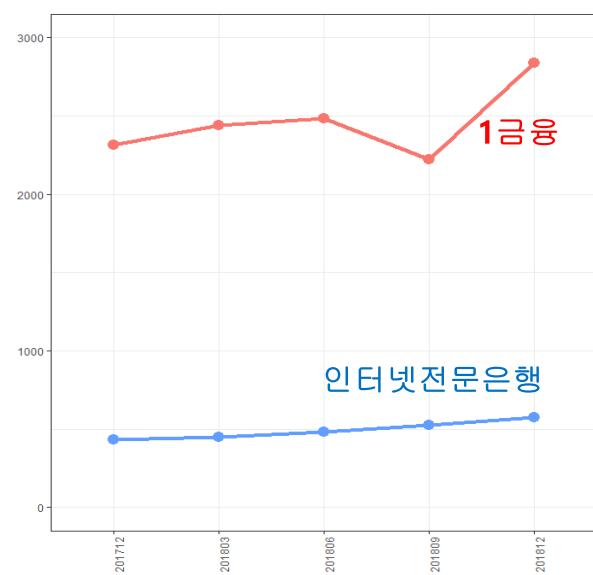
I. 서론

◆ 왜 소액대출인가?

[업종별 소액 대출의 비중]



[업종별 소액대출 신규 계좌 개설 수]



- ▶ 인터넷전문은행은 1금융권임에도 소액대출의 비중이 2금융권과 비슷함
- ▶ 인터넷전문은행은 **소액대출을 중심으로 성장**하였다
고 볼 수 있음

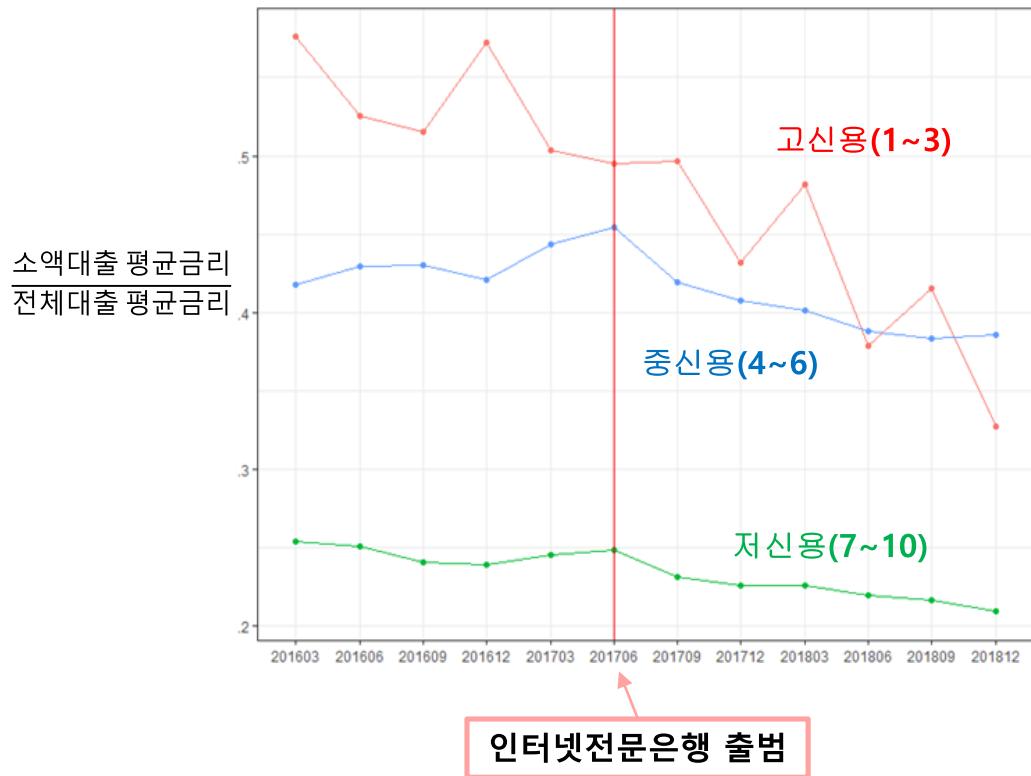
- ▶ 은행과 인터넷전문은행 모두 증가추세

⇒ 소액대출 시장에서 인터넷전문은행에 유입된 고객들과 그렇지 않은 고객들에 대한 분석 필요

II. (1) 소액대출 시장의 현황

◆ 소액대출 시장의 경쟁

[신용등급별 소액대출 금리 추이]



$$* \text{금리} = \frac{\text{상환이자}}{(\text{잔액} + \text{상환원금} + \text{상환이자})} * 100(%) * 12(\text{개월})$$

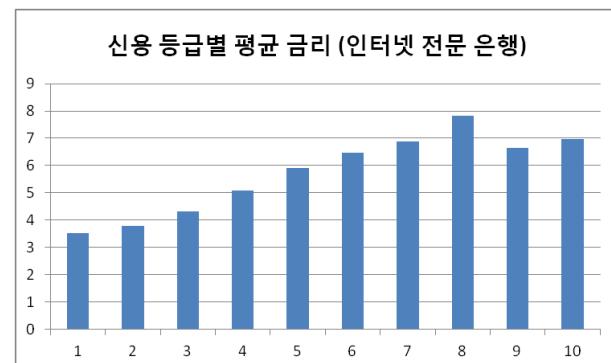
- 단, 미연체 차주 대상
- 계산된 모든 금리의 평균값
- 고정 금리로 가정

▶ 인터넷전문은행 출범 이후 전반적으로 소액대출 금리 하락

▶ 고신용군에서 하락 폭이 커짐

⇒ 고신용자를 대상으로 한 소액대출을 중심으로 경쟁이 과열된 것으로 보임

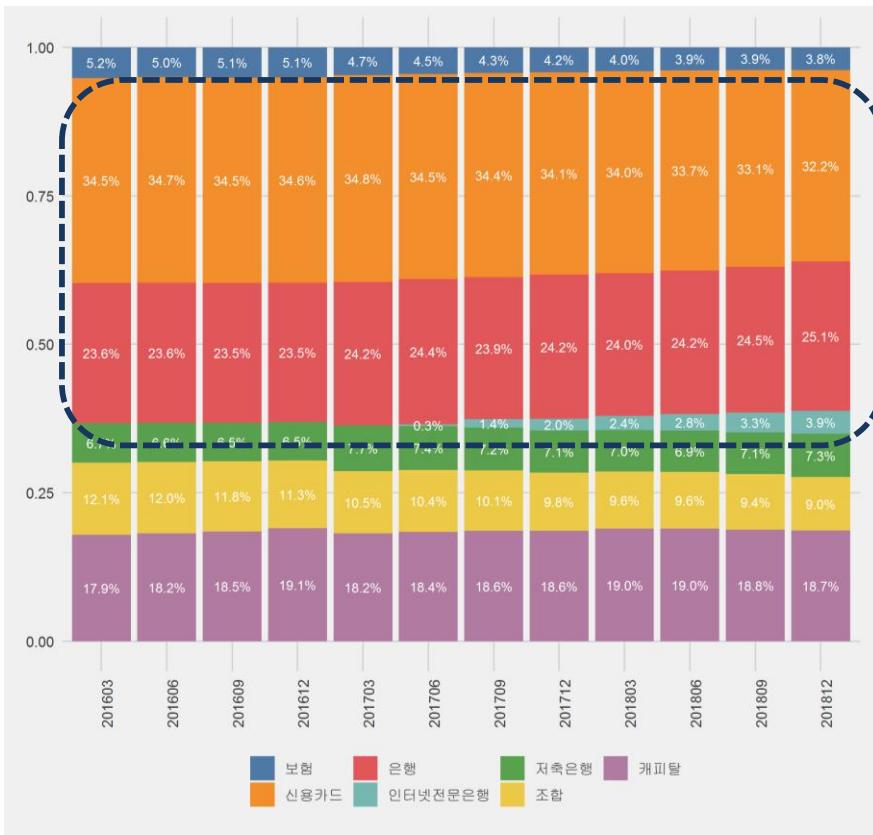
*금리 계산 예시



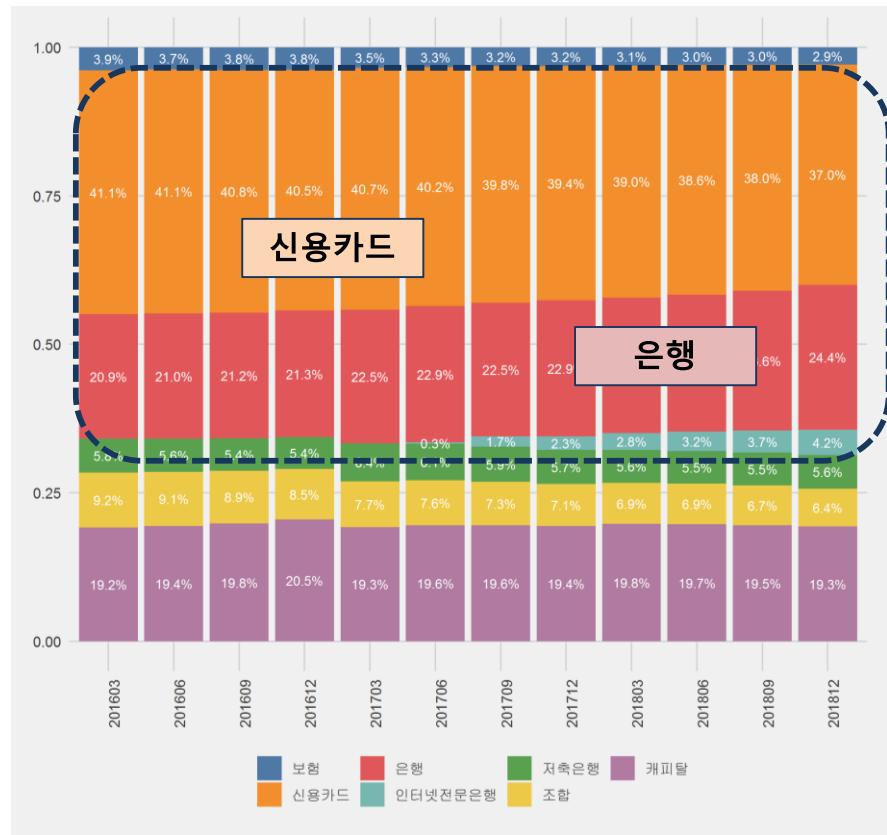
II. (1) 소액대출 시장의 현황

◆ 소액대출 시장의 경쟁

[소액대출의 업종별 점유율(약정금액 기준)]



[소액대출의 업종별 점유율(계좌수 기준)]

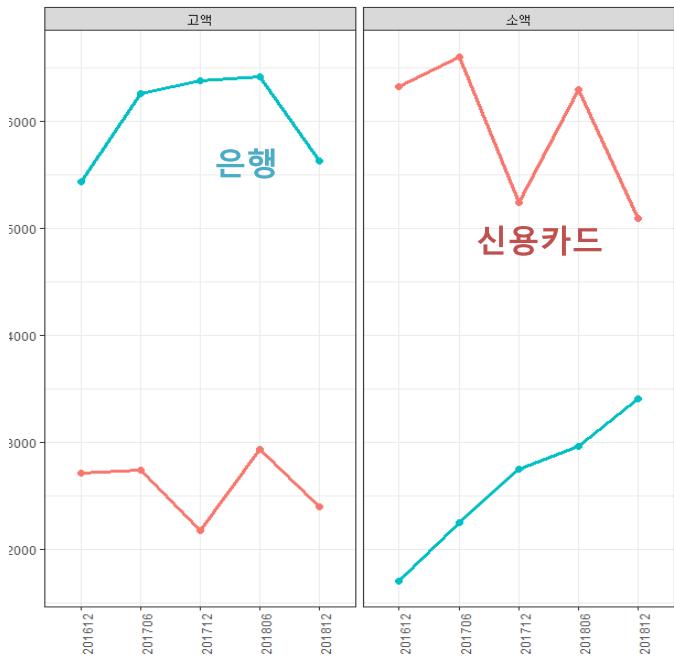


- ▶ 신용카드, 은행이 소액대출 시장의 대부분을 점유
- ▶ 인터넷전문은행 출범 이후 은행의 점유율 증가, 신용카드의 점유율 감소

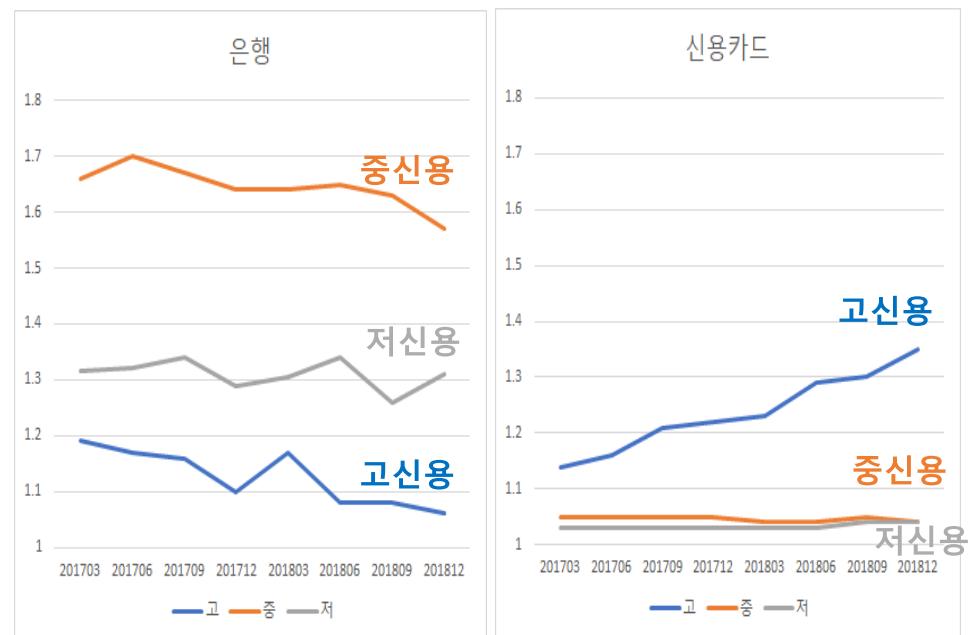
II. (1) 소액대출 시장의 현황

◆ 소액대출 시장의 경쟁

[업종별 신규계좌 개설 수 추이]



[업종별 소액대출 금리 추이]



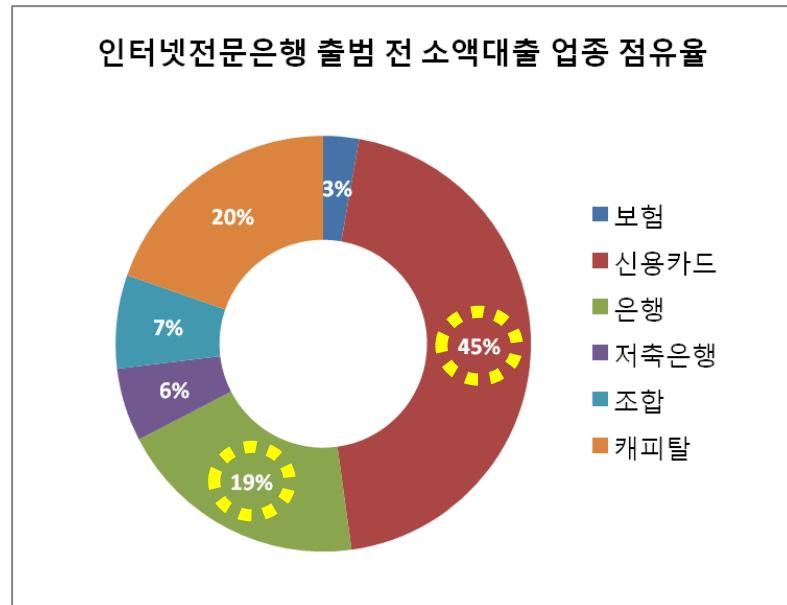
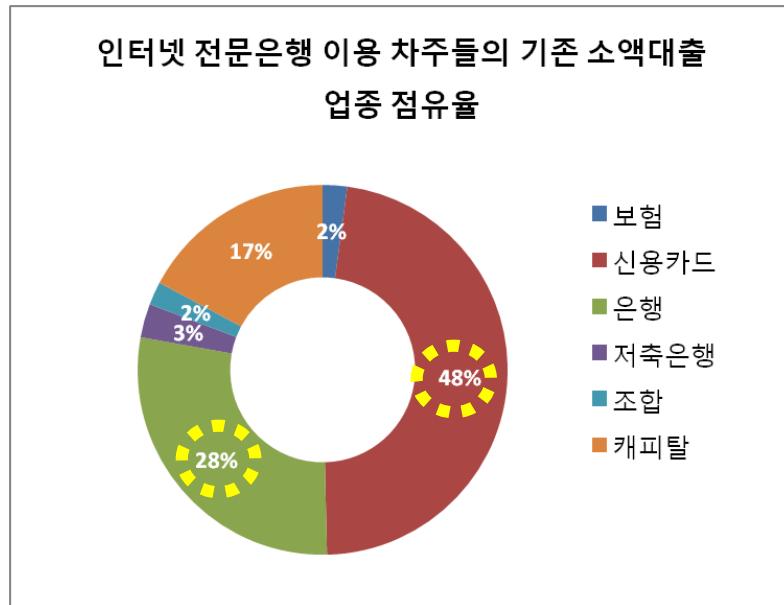
- ▶ 은행의 신규계좌는 소액대출 증가
- ▶ 신용카드의 신규계좌는 소액대출 감소

- ▶ 은행에서는 고신용자와 중신용자를 중심으로 소액대출 금리 감소
- ▶ 신용카드에서는 고신용자의 소액대출 금리가 증가하는 추세

⇒ 소액대출은 은행의 중신용자와 고신용자를 대상으로 특히 경쟁이 과열된 것으로 보임

II. (2) 인터넷전문은행 출범 후 유입된 고객

◆ 인터넷전문은행 이용 고객들의 기존 소액대출 업종 점유율



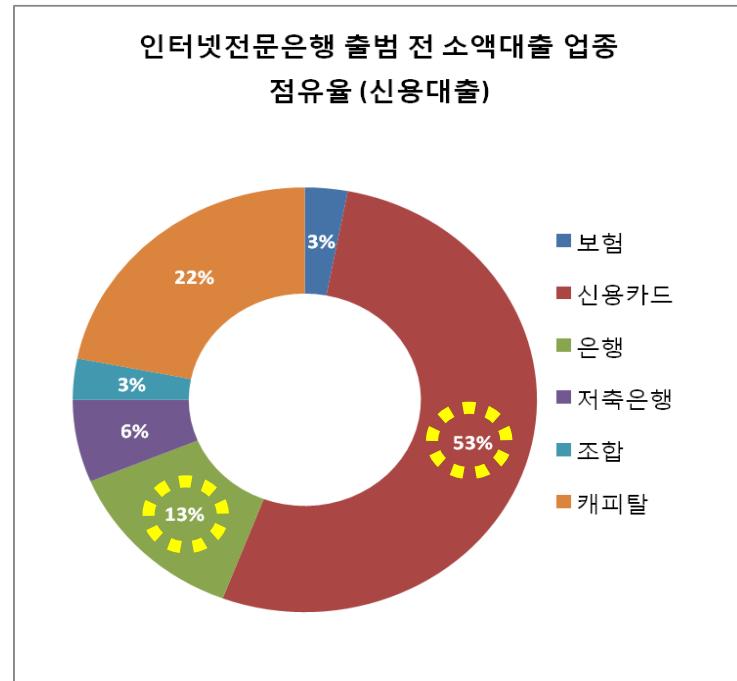
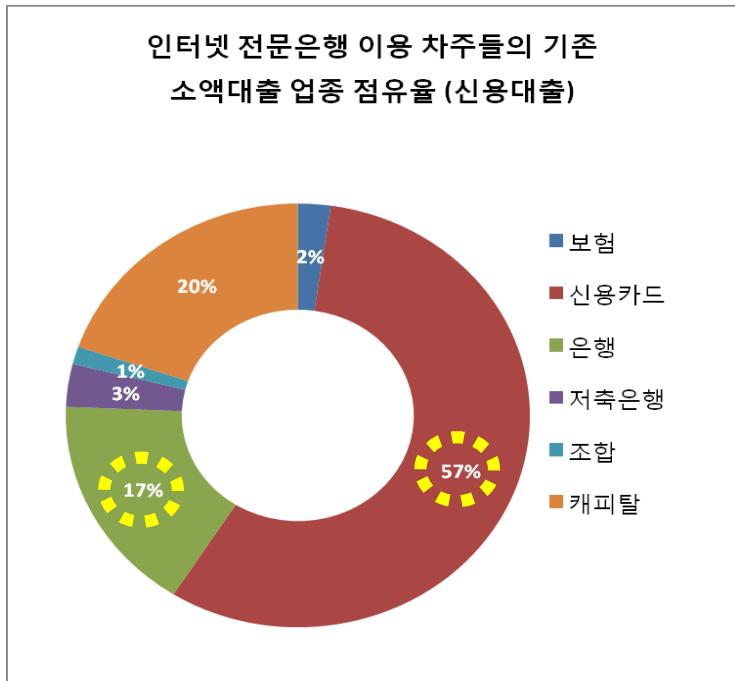
업종	유입비율*
보험	0.77
신용카드	1.05
은행	1.44
저축은행	0.52
조합	0.28
캐피탈	0.88

▶ 인터넷전문은행 소액대출 고객은 주로
은행과 신용카드 업종에서 유입된 것으로 보임

*유입비율: 기존 업종 점유율 대비 인터넷은행 이용 차주들의 과거 이용 업종 점유율
Ex) 신용카드 : 48%/45% = 1.05

II. (2) 인터넷전문은행 출범 후 유입된 고객

◆ 인터넷전문은행 이용 고객들의 기존 소액대출 업종 점유율



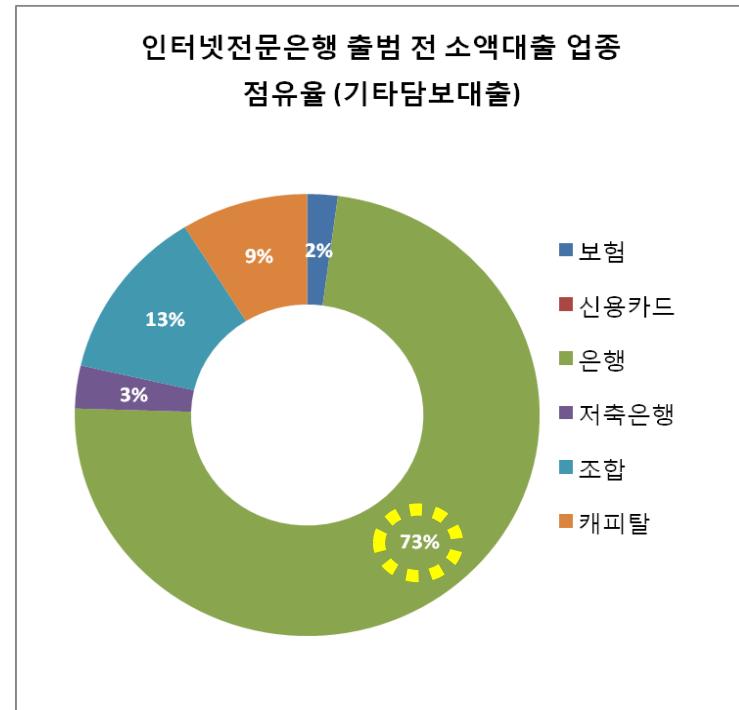
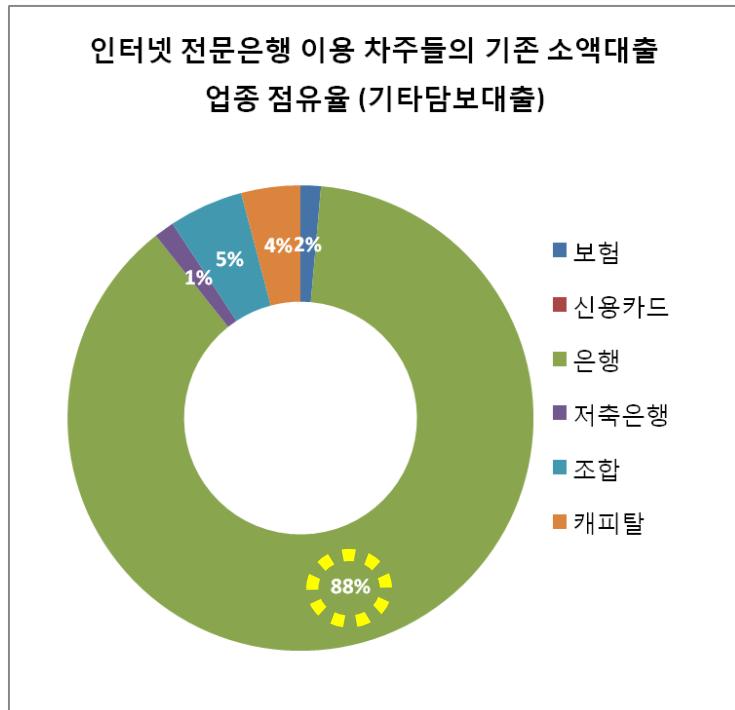
업종	유입비율*
보험	0.77
신용카드	1.07
은행	1.28
저축은행	0.52
조합	0.43
캐피탈	0.90

*유입비율: 기존 업종 점유율 대비 인터넷은행 이용 차주들의 과거 이용 업종 점유율
Ex) 신용카드 : 48%/45% = 1.05

▶ 인터넷전문은행의 소액 신용대출 고객은 주로 은행과 신용카드 업종에서 유입된 것으로 보임

II. (2) 인터넷전문은행 출범 후 유입된 고객

◆ 인터넷전문은행 이용 고객들의 기존 소액대출 업종 점유율



업종	유입비율*
보험	0.67
은행	1.20
저축은행	0.45
조합	0.41
캐피탈	0.47

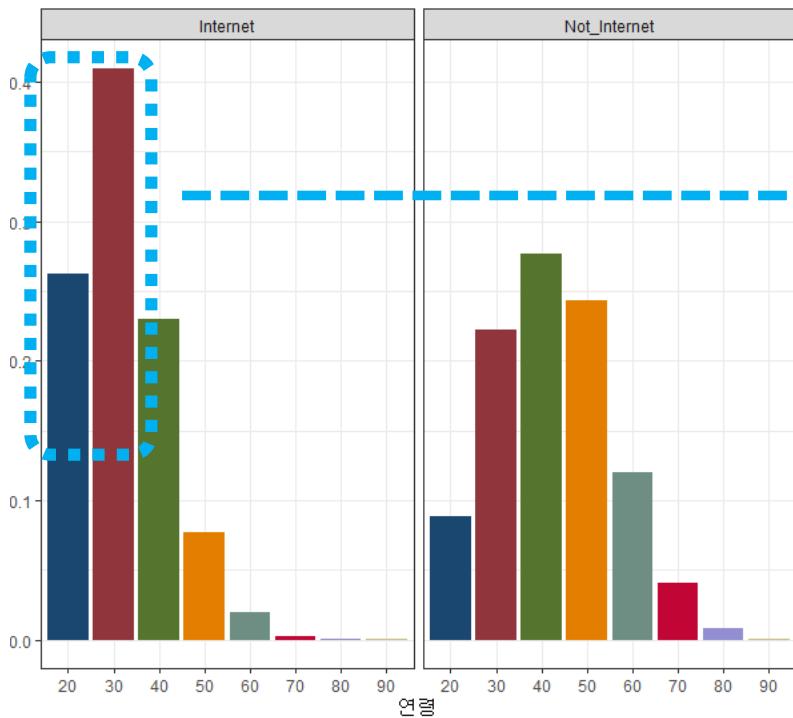
*유입비율: 기존 업종 점유율 대비 인터넷은행 이용 차주들의 과거 이용 업종 점유율
Ex) 신용카드 : 48%/45% = 1.05

▶ 인터넷전문은행의 소액 기타담보대출 고객은
주로 은행 업종에서 유입됨

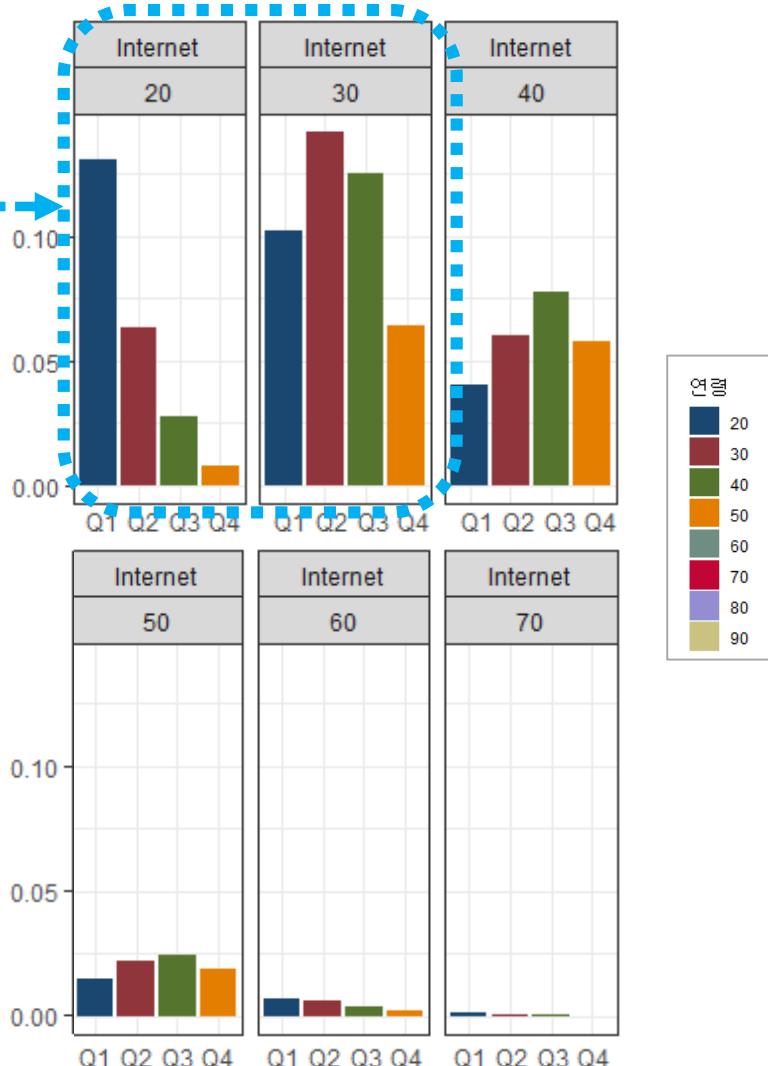
II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 이용하는 소액대출 고객의 특징

[인터넷전문은행 이용고객과 비이용고객의 연령대]



[인터넷전문은행 이용고객의 연령 및 소득분위]

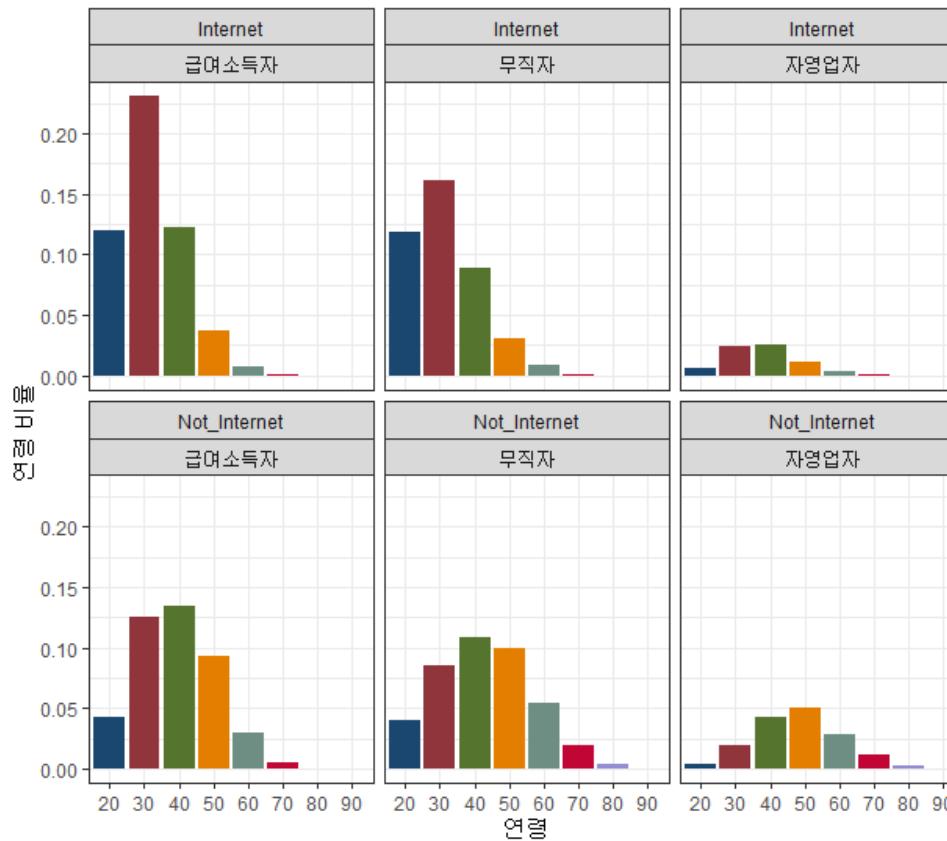


- ▶ 인터넷전문은행 소액대출 이용고객은 상대적으로 연령대가 낮은 편
- ▶ 특히 20대와 30대 저소득 차주의 비중이 상당히 높음

II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 이용하는 소액대출 고객의 특징

[인터넷 전문은행 이용고객과 비이용고객의 직업 및 연령대]

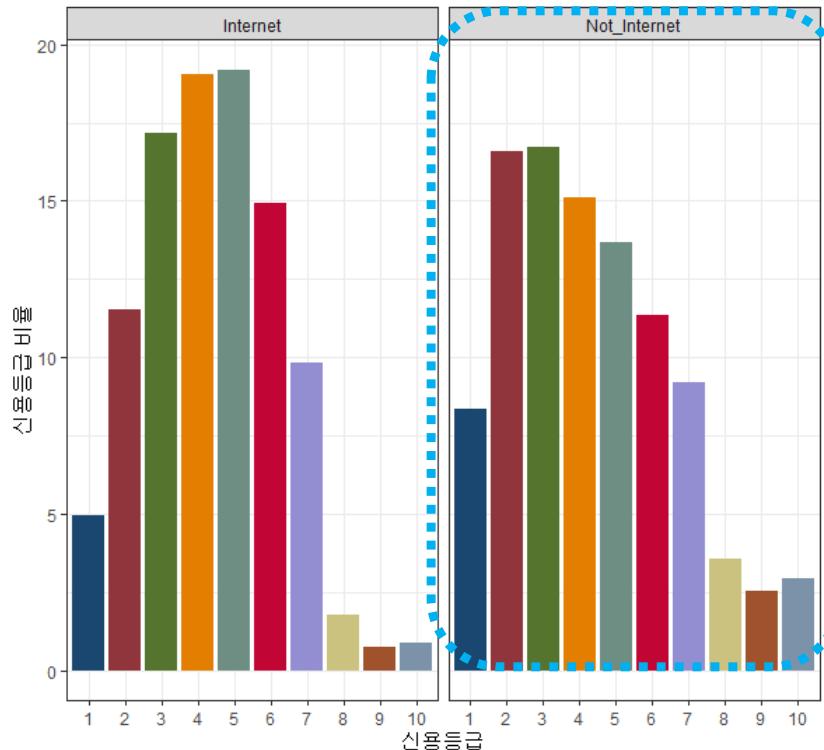


- ▶ 자영업자의 소액대출 비중이 다른 직업군에 비해 낮고, 인터넷전문은행 이용 고객은 특히 더 낮음
- ▶ 급여소득자와 무직자의 경우 인터넷전문은행 외 업종에서는 40대의 비중이 가장 높지만, 인터넷전문은행에서는 20, 30대의 비중이 40대의 비중과 비슷하거나 높음

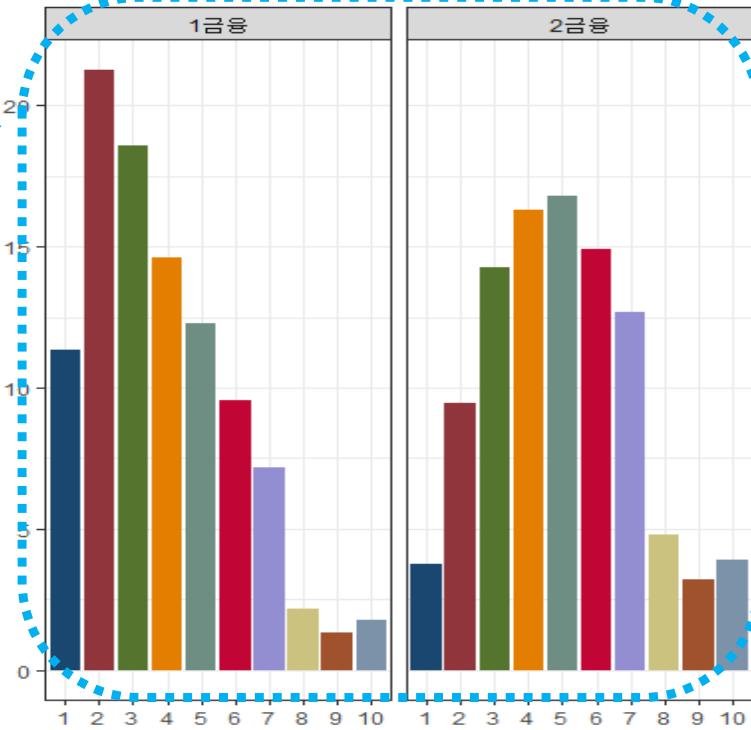
II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 이용하는 소액대출 고객의 특징

[인터넷전문은행 이용고객과 비이용고객의 신용등급]



[인터넷전문은행, 1금융, 2금융 고객의 신용등급]

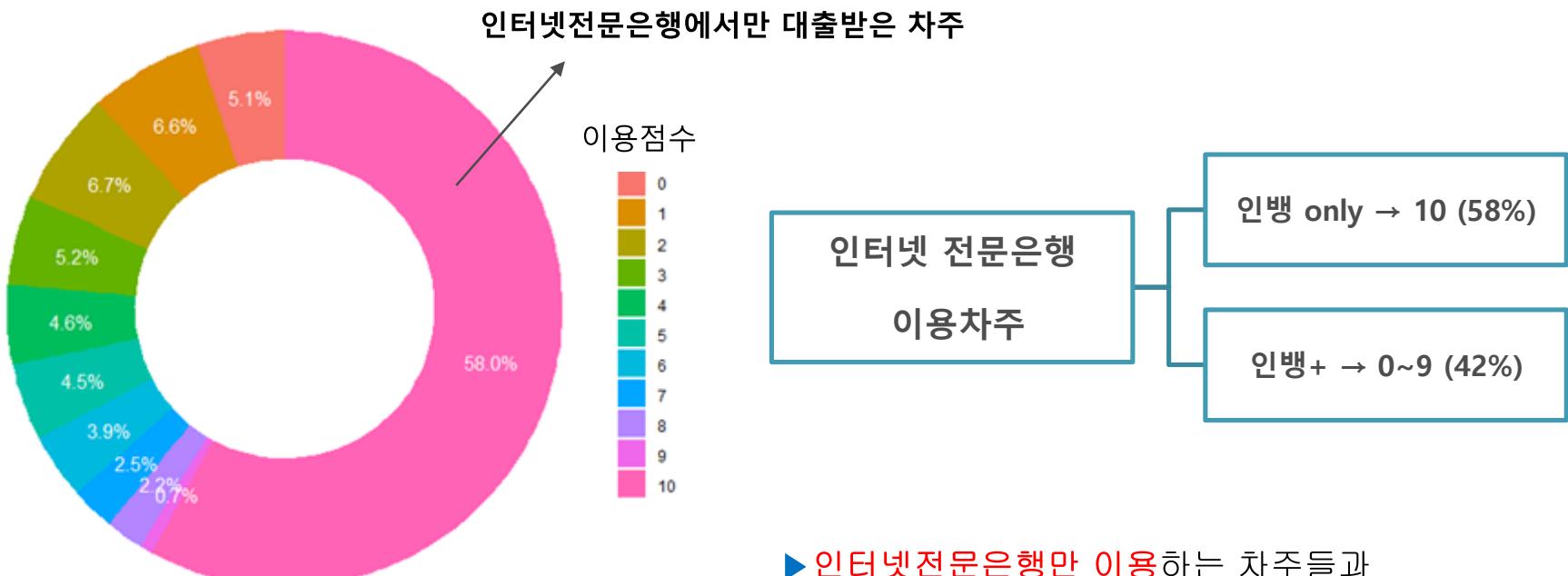


- ▶ 인터넷전문은행 이용차주들은 기존 1금융권 차주들에 비해서는 고신용자 비율이 낮지만 2금융권 차주들에 비해서는 고신용자 비율이 높음

II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 단독으로 이용하는 소액대출 고객의 특징

[인터넷전문은행 고객의 이용점수 구성비]



• 이용점수

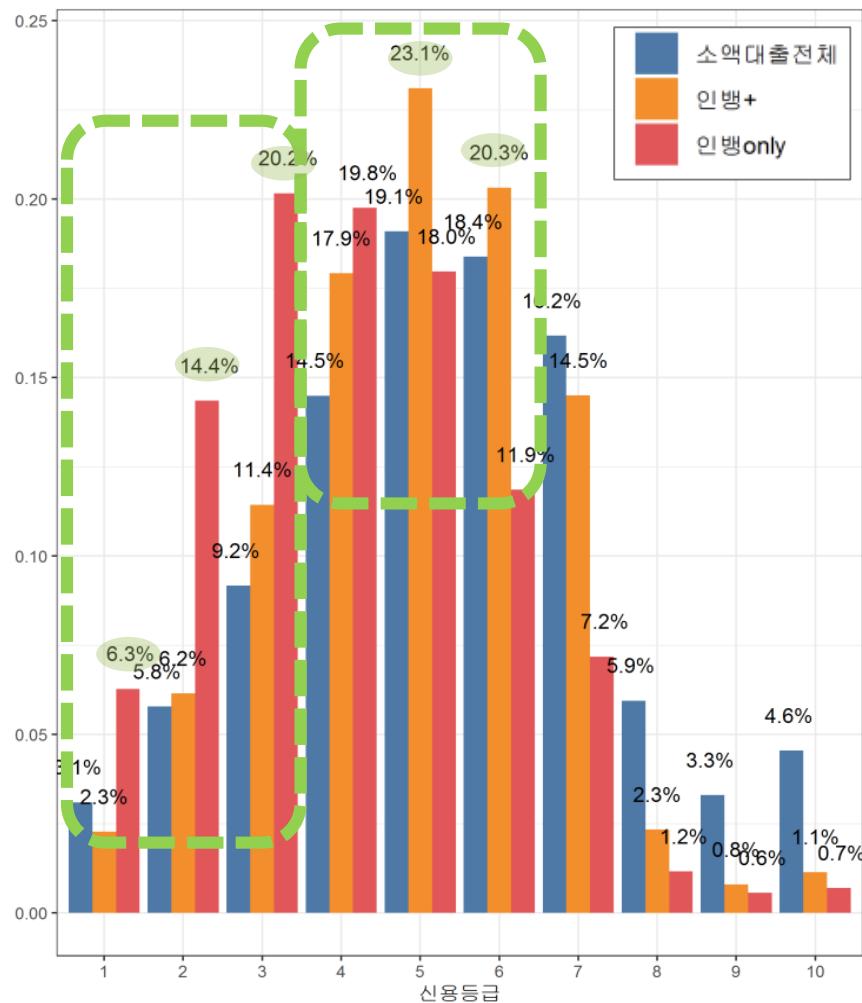
인터넷 전문은행 소액대출상품 가입 후 각 차주의
총 소액대출 약정금액 중 인터넷 전문은행 소액대출
약정금액이 차지하는 비중

	인터넷전문은행	인터넷전문은행 외	비율	점수
고객 A	50만원	50만원	0.5	5
고객 B	100만원	0만원	1	10

II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 단독으로 이용하는 소액대출 고객의 특징

[인터넷전문은행 소액대출 고객의 신용등급]

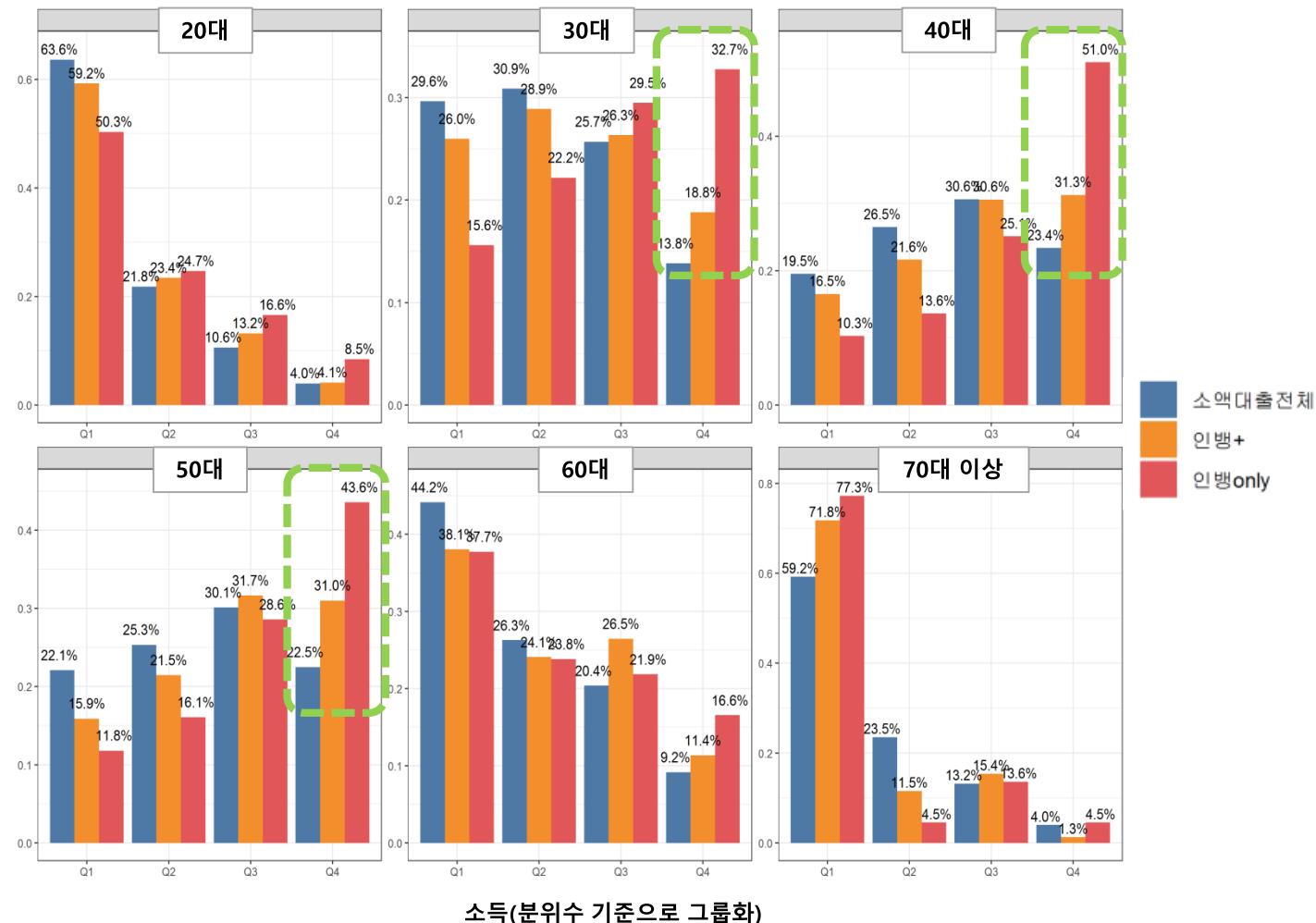


- ▶ 인터넷전문은행만 이용하는 고객은 고신용자 비중이 높음
- ▶ 인터넷전문은행 외에 다른 업종도 이용하는 고객은 중신용자 비중이 높음

II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 단독으로 이용하는 소액대출 고객의 특징

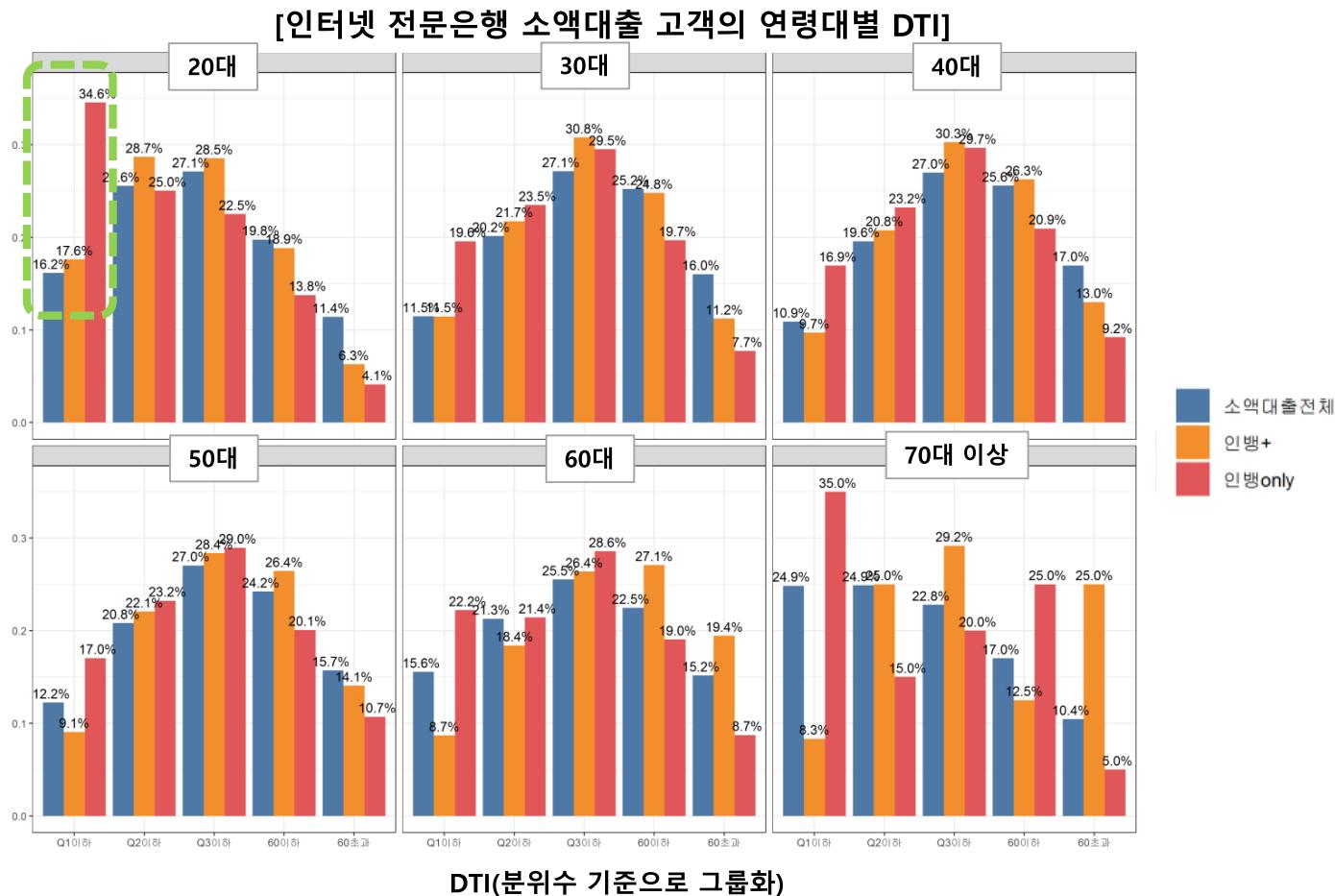
[인터넷전문은행 소액대출 고객의 연령대별 소득]



▶ 30-50대의 경우 인터넷전문은행만 이용하는 차주들이 다른 그룹에 비해 고소득자 비중이 높음

II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 단독으로 이용하는 소액대출 고객의 특징

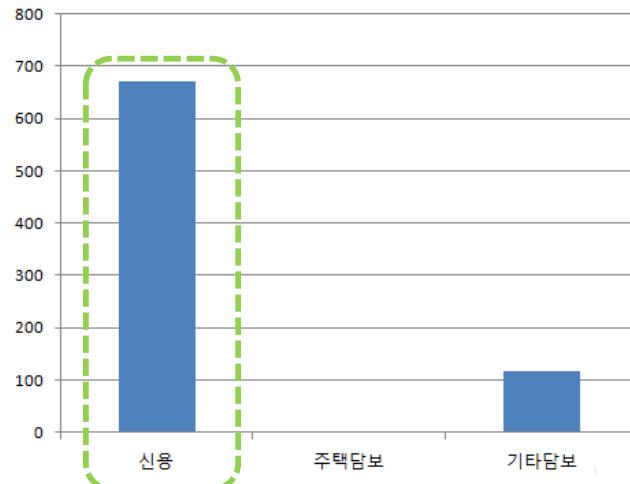


- ▶ 모든 연령대에서 인터넷 전문은행만 사용하는 차주들은 DTI가 낮은 그룹의 비율이 상대적으로 높음
- ▶ 특히 20대 인뱅only 고객의 경우 DTI가 낮은 그룹의 비율이 매우 높게 나타난 것으로 보임

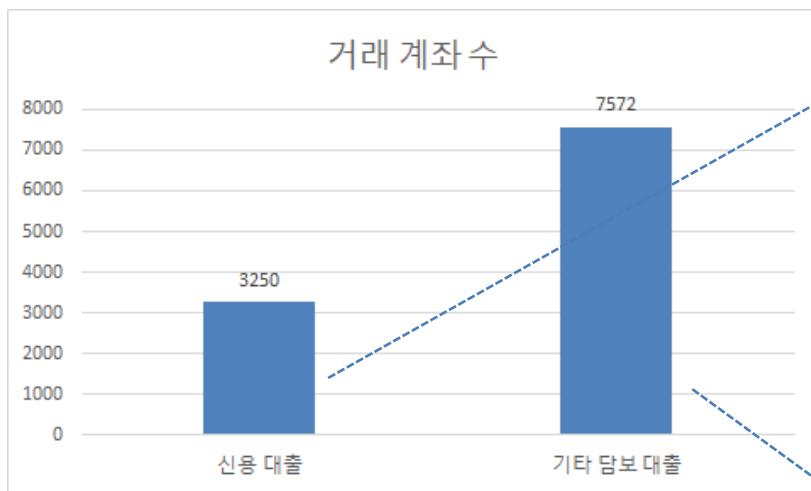
II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 단독으로 이용하는 소액대출 고객의 특징

[신용대출 차주들이 기존에 이용한 대출상품]

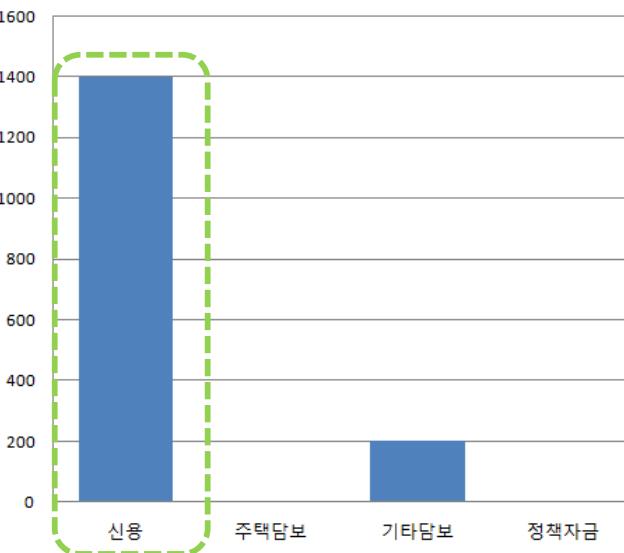


[인터넷전문은행 단독 이용 고객의 대출상품]



▶ 현재 신용대출을 이용하는 차주들과
기타담보대출을 이용하는 차주들 모두
과거에는 대부분 신용대출을 이용

[기타담보대출 차주들이 기존에 이용한 대출상품]



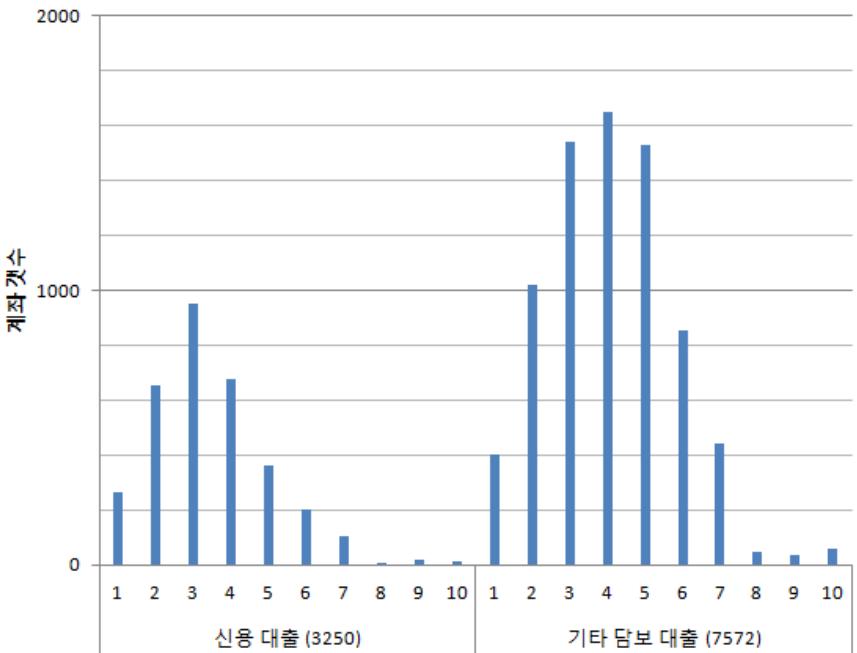
II. (3) 유입고객과 충성고객의 특징

◆ 인터넷전문은행을 단독으로 이용하는 소액대출 고객의 특징

[카카오 뱅크의 대출 상품 및 종류]

대출 상품 (종류)	한도(원)	금리(%) (2019.09.02 기준)
비상금 대출 (기타 담보)	300만	3.06 ~ 15.00
사잇돌 대출 (기타 담보)	2천만	3.96 ~ 9.80
개인사업자 대출 (기타 담보)	2천만	3.08 ~ 9.80
전월세 대출 (기타 담보)	2.22억	2.54 ~ 3.51
마이너스 통장 대출 (신용)	1.5억	2.78 ~ 6.19
신용 대출 (신용)	1.5억	2.48 ~ 6.27
중 신용 대출 (신용)	5천만	3.71 ~ 6.91

[인터넷전문은행 단독 이용 고객의 대출상품별 신용등급]



▶ 기타담보대출 이용 고객은 신용대출 이용 고객에 비해 상대적으로 **중신용자의 비중이 더 높음**

II. (3) 유입고객과 충성고객의 특징

◆ 충성고객의 특징

▶ 충성고객이란?

- 인터넷 전문은행으로 유입되지 않은 차주 중, 기존에 이용하던 거래기관만 지속적으로 이용하며 거래기관을 옮기지 않는 차주들을 의미
- 예시
- 기존에 신한은행과 국민카드를 이용하던 차주가 우리은행에서도 대출을 받음 → 충성고객 X
- 기존에 신한은행과 국민카드를 이용하던 차주가 은행대출을 갚아 국민카드 대출계좌만 남은 경우 → 충성고객 O

[인터넷 전문은행으로의 유입비율]

업종	유입비율*
보험	0.77
신용카드	1.05
은행	1.44
저축은행	0.52
조합	0.28
캐피탈	0.88

▶ 유입비율이 높은 은행과 신용카드 업종의 충성고객
을 중심으로 분석

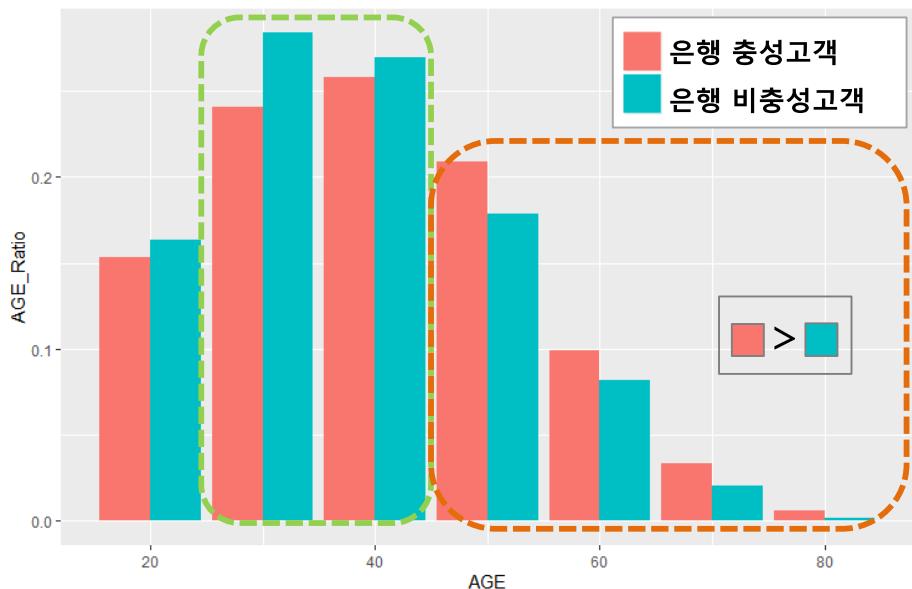
* 은행 충성고객 : 고객의 전체 소액대출 중 은행의 소액대출이 50% 이상

* 신용카드 충성고객 : 고객의 전체 소액대출 중 신용카드의 소액대출이 50% 이상

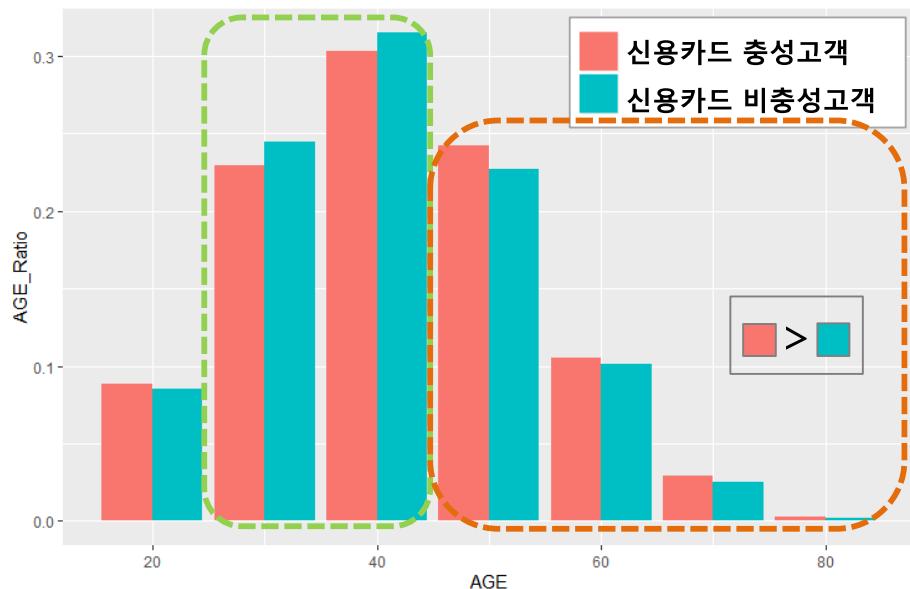
II. (3) 유입고객과 충성고객의 특징

◆ 충성고객의 특징

[은행의 충성고객 연령대]



[신용카드의 충성고객 연령대]

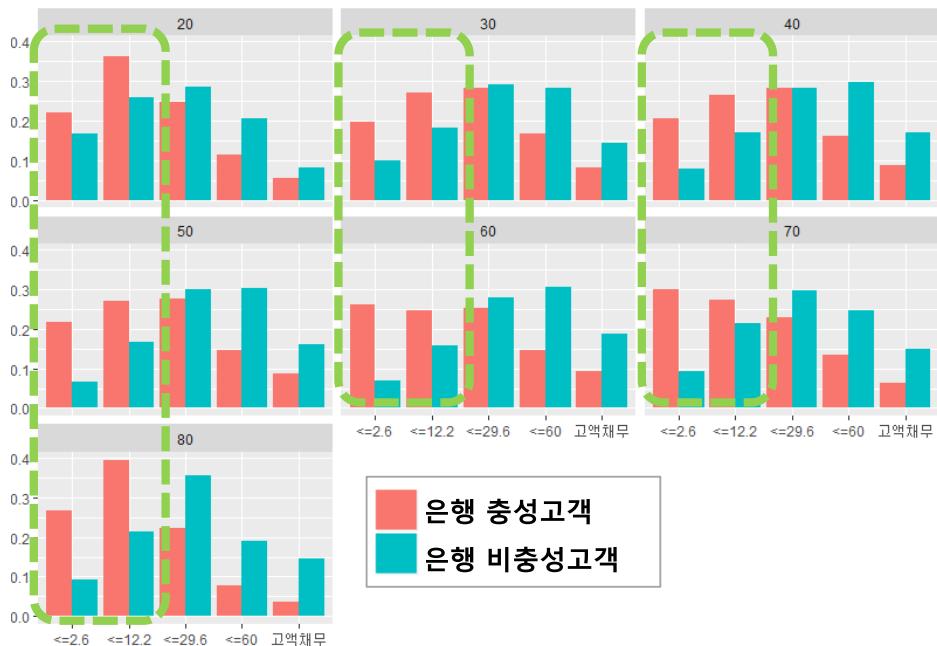


- ▶ 거래기관을 유지하는 고객은 그렇지 않은 고객에 비해 상대적으로 연령대가 더 높은 경향
- ▶ 카드사는 충성고객과 비충성고객의 연령 분포 차이가 매우 미미함

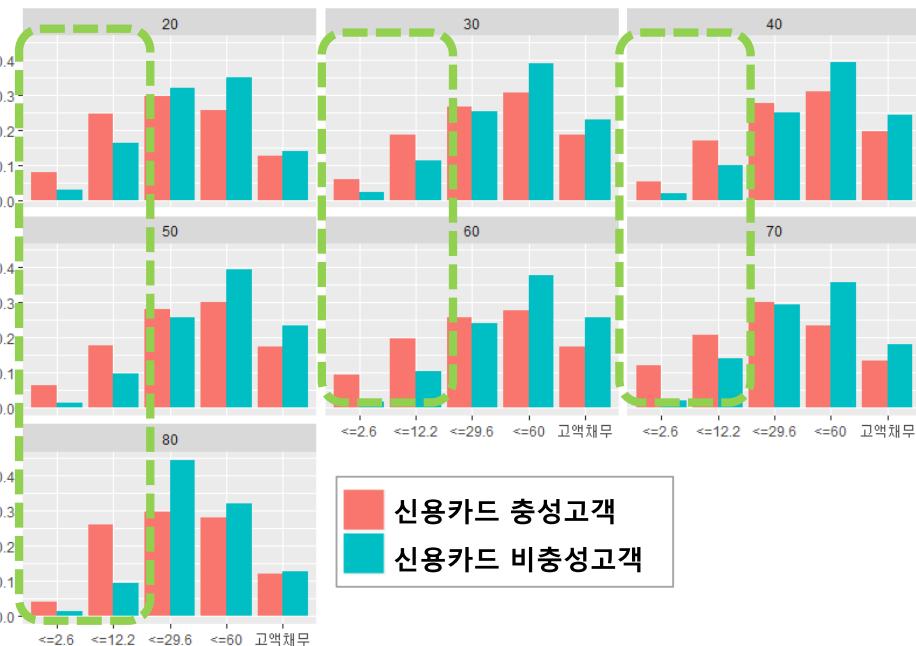
II. (3) 유입고객과 충성고객의 특징

◆ 충성고객의 특징

[은행의 충성고객 연령대별 DTI]



[신용카드의 충성고객 연령대별 DTI]



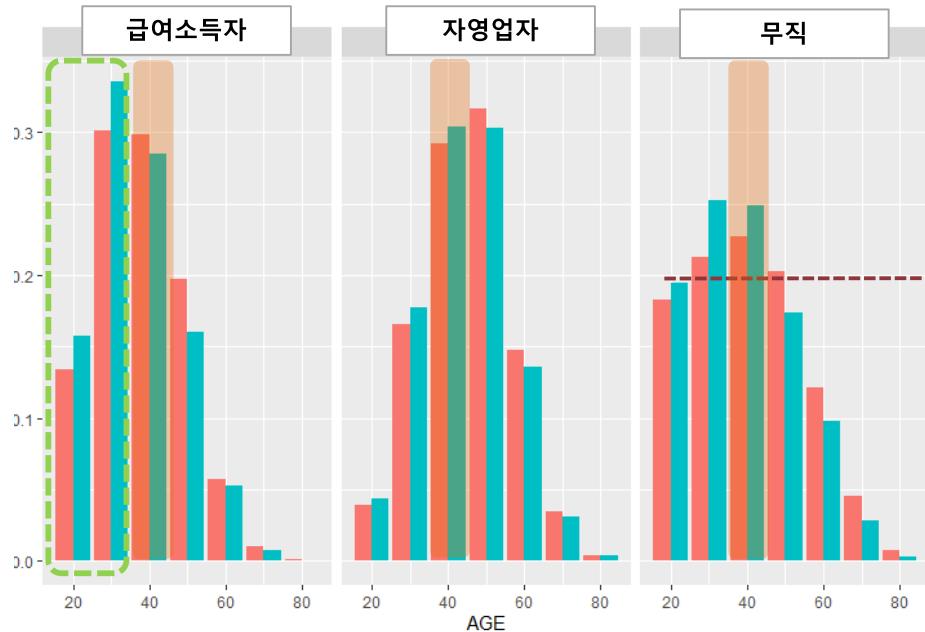
- ▶ 은행에서는 연령대가 높아질수록, 충성/비충성고객의 DTI 분포 차이가 커짐
- ▶ 은행 충성고객은 DTI가 더 낮은 쪽으로, 비충성고객은 DTI가 더 높은 쪽으로 차이 발생

- ▶ 카드사보다는 은행에 DTI가 낮은 고객들이 많고, **DTI가 낮은 고객일수록 거래기관을 유지하는 경향**
- ▶ DTI가 낮은 차주들은 대출을 추가적으로 더 받지 않는 것으로 유추 가능

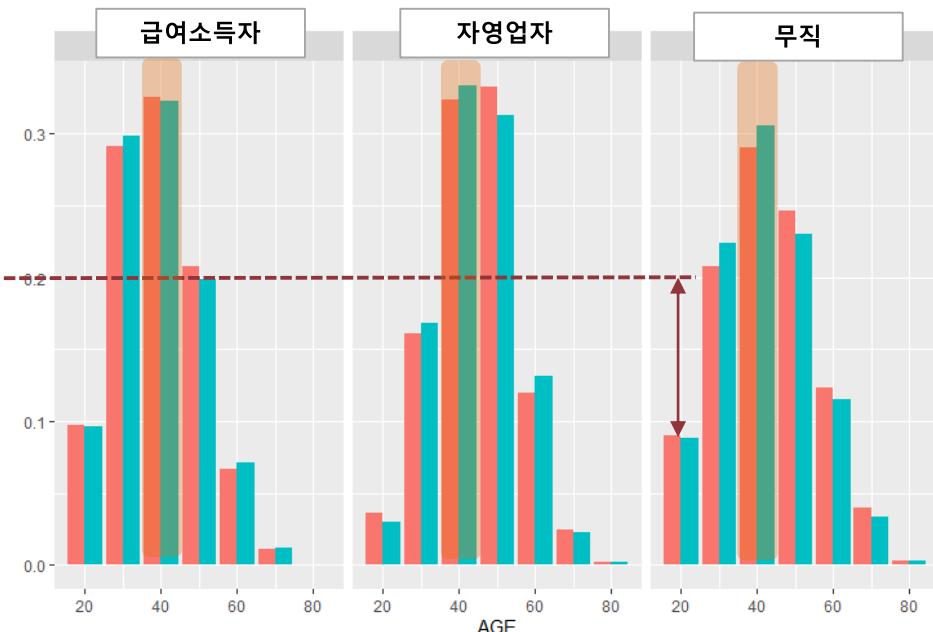
II. (3) 유입고객과 충성고객의 특징

◆ 충성고객의 특징

[은행의 충성고객 직업별 연령대]



[신용카드의 충성고객 직업별 연령대]



- ▶ 은행 이용고객 중 급여소득자는 젊을수록 거래기관을 옮기는 경향
- ▶ 40대 급여소득자의 경우 충성고객의 비중이 더 크고, 자영업자와 무직의 경우 비충성고객의 비중이 더 큼
- ▶ 무직 20대 차주들(학생 추정)의 경우, 카드사보다는 은행에서의 대출 비중이 더 큼

III. 요약 및 결론

◆ 요약

▶ 소액대출 시장 현황

- 인터넷전문은행 출범 이후 **1금융권**에서 **소액대출 경쟁이 심화**
- 인터넷전문은행 출범 이후 소액대출 시장에서 **은행 점유율 증가, 신용카드 점유율 감소**
- **인터넷전문은행**은 1금융권임에도 **소액대출 중심**으로 성장

▶ 인터넷전문은행 소액대출 이용고객 특징

- **은행, 신용카드** 업종에서 인터넷전문은행으로 유입이 많음
- 인터넷전문은행 비이용고객에 비해 **연령대가 낮은 편**
- **1금융권보다는 고신용자 비중이 작고, 2금융권보다는 고신용자 비중이 큼**
- 인터넷전문은행 비이용고객에 비해 **급여소득자의 비중이 크고, 자영업자의 비중은 작음**

III. 요약 및 결론

◆ 요약

▶ 인터넷전문은행을 단독으로 이용하는 소액대출 고객의 특징

- 고신용자 비중이 크고, 40~50대에서 고소득자 비중이 큼
- DTI가 낮은 고객들의 비중이 큼(특히 20대)
- 이용하는 대출상품은 대부분 기타담보대출이지만 과거이력은 대부분 신용대출
- 기타담보대출을 이용하는 고객은 중신용자 비중이 큼

▶ 인터넷 비유입고객 중 은행과 신용카드의 충성고객 특징

- 업종에 따른 충성고객 간 특성 차이는 발견할 수 없었음
- 카드보다는 은행에서 충성고객과 비충성고객 간 차이가 뚜렷하게 나타남
- 충성고객이 비충성고객에 비해 연령대는 높고, DTI는 낮은 편

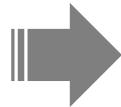
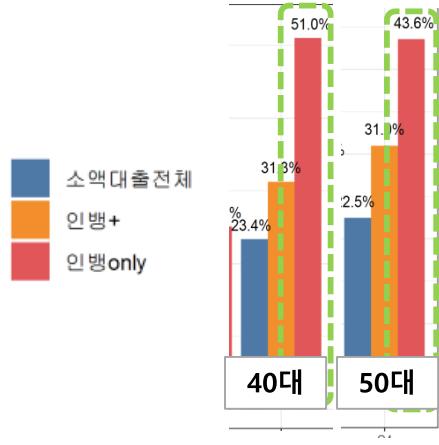
III. 요약 및 결론

◆ 결론

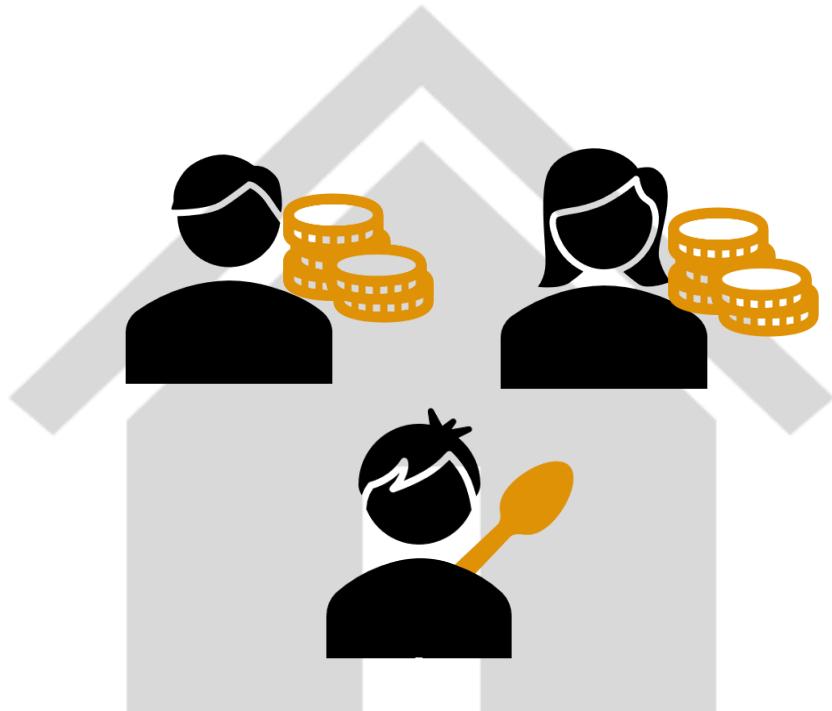
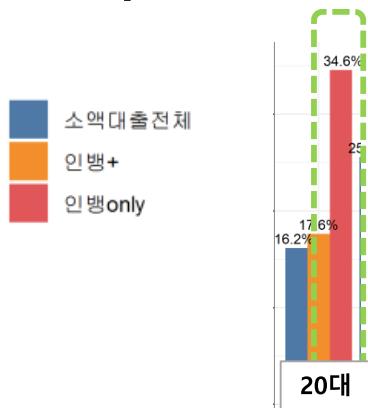
인터넷 전문은행?

1) 40~50대의 고소득 고객 대상 가족결합 마케팅

[소득 Q3이상 인터넷 소액대출 비중]



[DTI Q1이하 인터넷 소액대출 비중]



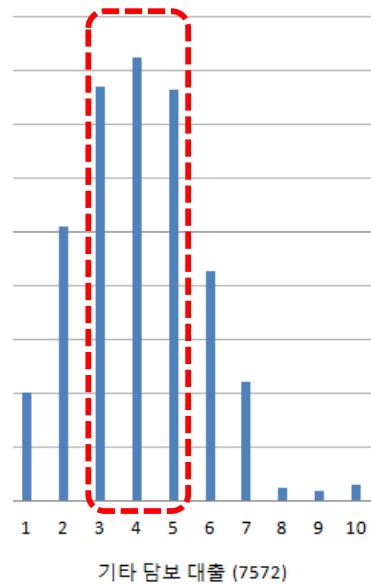
III. 요약 및 결론

◆ 결론

인터넷 전문은행?

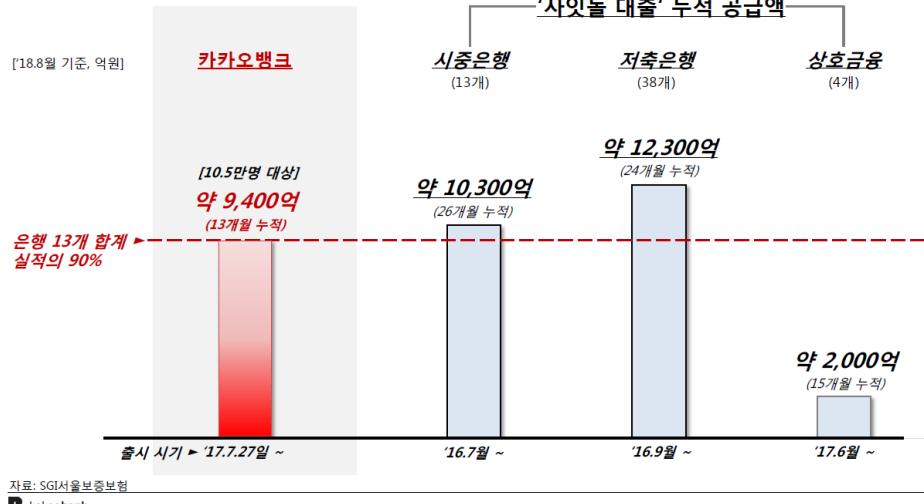
2) 중신용자 대상 중금리 대출 상품 홍보

[인터넷 전문은행 기타담보대출 이용 차주의
신용등급 분포]



카뱅의 SGI보증부 신용대출 상품 누적 공급 현황

'사잇돌 대출'과 실질적으로 동일한 보증부 중금리 대출 상품을 통해, 카뱅은 약 13개월간 시중 은행 사잇돌 대출 전체 실적의 90%에 해당하는 9,400억을 10.5만명에게 공급했음.



자료: SGI서울보증보험
kakaobank

3

대출 상품 (종류)	한도(원)	금리(%) (2019.09.02 기준)
비상금 대출 (기타 담보)	300만	3.06 ~ 15.00
사잇돌 대출 (기타 담보)	2천만	3.96 ~ 9.80
개인사업자 대출 (기타 담보)	2천만	3.08 ~ 9.80
전월세 대출 (기타 담보)	2.22억	2.54 ~ 3.51

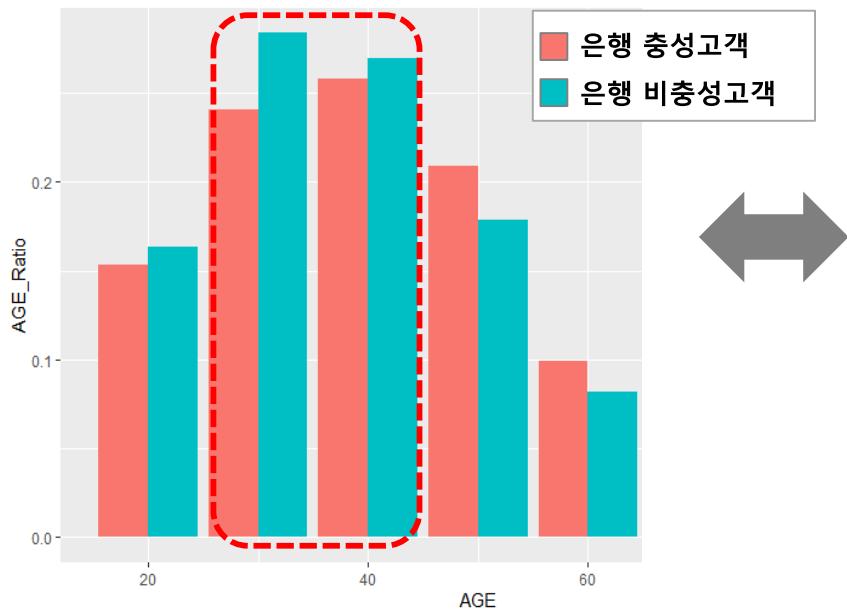
III. 요약 및 결론

◆ 결론

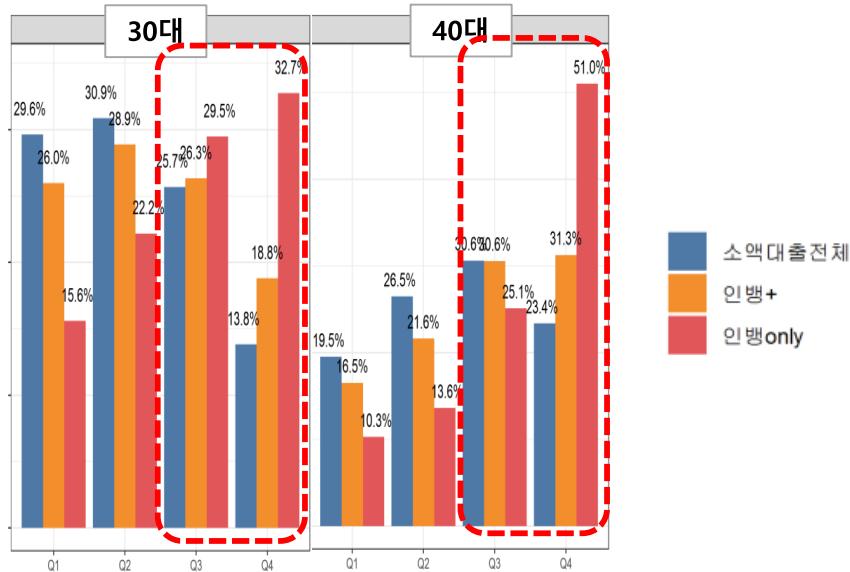
은행?

1) 잠재 유출 고객에 대해 관리 필요 (1)

[은행의 충성고객 연령대]



[인터넷 전문은행 고객의 연령대별 소득]



30-40대 고객의 경우 고소득 고객 위주로 인터넷 전문은행으로 유출 가능성 ↑
⇒ 잠재 유출 고객에 대해 관리 필요

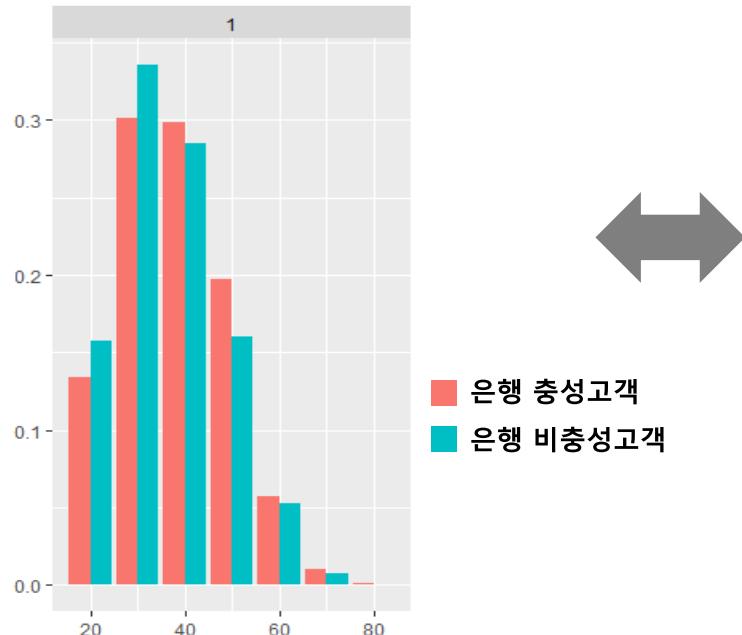
III. 요약 및 결론

◆ 결론

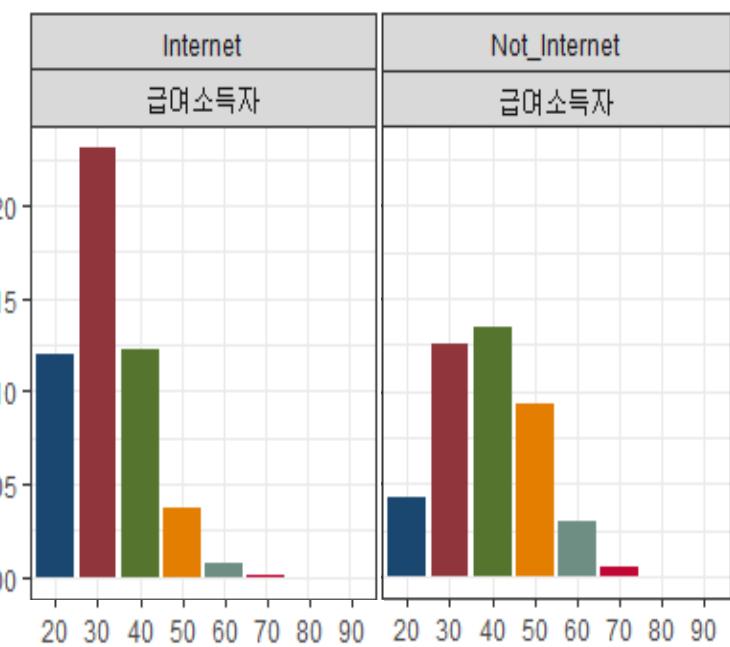
은행?

1) 잠재 유출 고객에 대해 관리 필요 (2)

[은행의 급여소득자 충성고객 연령대]



[인터넷/비인터넷 급여소득자 충성고객 연령대]



20-30대 급여소득자 고객의 경우 인터넷 전문은행으로 유출 가능성 ↑
⇒ 잠재 유출 고객에 대해 관리 필요

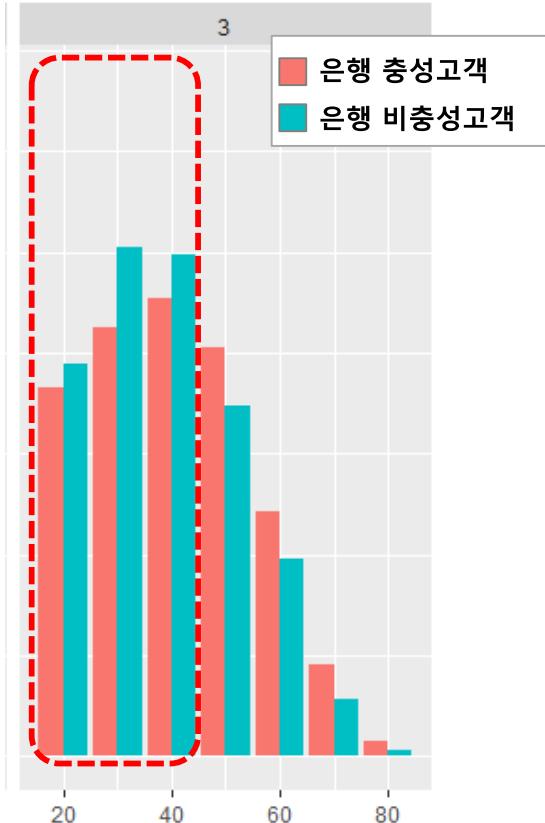
III. 요약 및 결론

◆ 결론

은행?

2) 제 2금융권과 제휴하여 수익창출 (2)

[은행의 무직 충성고객 연령대]



무직의 20~40대는 충성고객의 비중이 낮음

⇒ 제 1금융권에서 대출 거절 추측

⇒ **제 2금융권과 제휴하여 연계대출 실시**

(실제로, 카카오뱅크에서 시행 중)

2금융권, 카뱅과 연계대출 '맞손' 효과는

吕布 이봄 기자 | 0 승인 2019.05.08 16:02 | 댓글 0

카드·캐피탈·저축은행 5곳, 카뱅과 연계대출 시작
카뱅 모바일 앱에서 한 번에 한도·금리 조회 가능



그래픽=강세이 편집기자

III. 요약 및 결론

◆ 결론

신용카드?

1) 타 업권과 구별되는 경쟁력 강화 필요

- 충성/ 비 충성 고객 간 뚜렷한 차이점이 보이지 않음
- 카드대출의 가장 큰 장점인 편리성에 의해 고객이 유동적으로 업종 이용 추측
- 최근 인터넷전문은행, 1금융권에서도 비대면대출 상품

은행별 비대면 신용대출 상품

자료: 각 은행

은행	상품명	금리(%)	대상	한도(원)
국민은행	KB스타신용대출	3.26	직장인(12개월 이상 재직)	1억5000만
KEB 하나은행	하나원큐신용대출	2.792~3.292	소득 있는 국민 (직장인 이외 자영업자, 주부 포함)	2억2000만
신한은행	쏠편한직장인대출S	2.61~4.51	지정업체 직장인(1년 이상 재직)	2억
우리은행	우리주거래직장인대출	3.4~4	직장인(1년 이상 재직), 주거래 고객	8000만
카카오뱅크	카카오뱅크신용대출	2.798~6.566	직장인(6개월 이상 재직)	1억5000만

⇒ 타 업권과 구별되는 신용카드만의 경쟁력 강화 필요

- ① 기존 신용카드의 장점인 편리성 보강 및 홍보
- ② 편리성 외에 신용카드만의 장점 혹은 경쟁력 강화

감사합니다

개인신용평가 모델링 과제

2019.9.26

KCB 인턴 4조

목 차

I. 변수 소개

II. 주요 분석 내용

(1) 샘플링

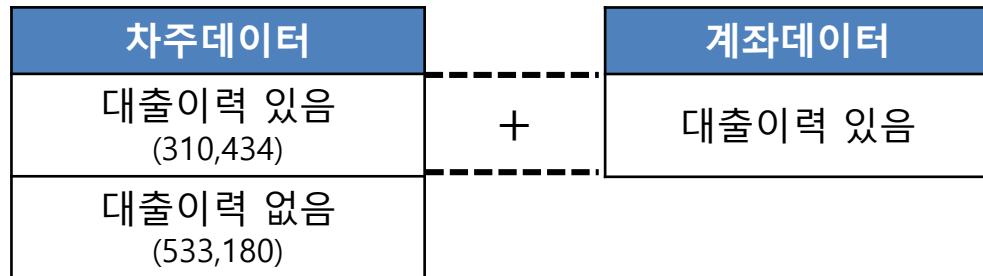
(2) 모델링

(3) 변별력 및 안정성 확인

(4) 추가 고려사항

III. 분석의의 및 발전방향

I. 변수 소개 – 데이터 구조



차주데이터	기준월	시점
Train + Validation	201609	
	201612	과거시점
	201703	
	201706	기준시점
Test	201709	
	201712	과거시점
	201803	
	201806	기준시점

→ 기불량자(해당시점 대출, 카드 연체자) 제외 (5,854)

→ 기불량자(해당시점 대출, 카드 연체자) 제외 (6,174)

I. 변수 소개 – 차주단위 데이터

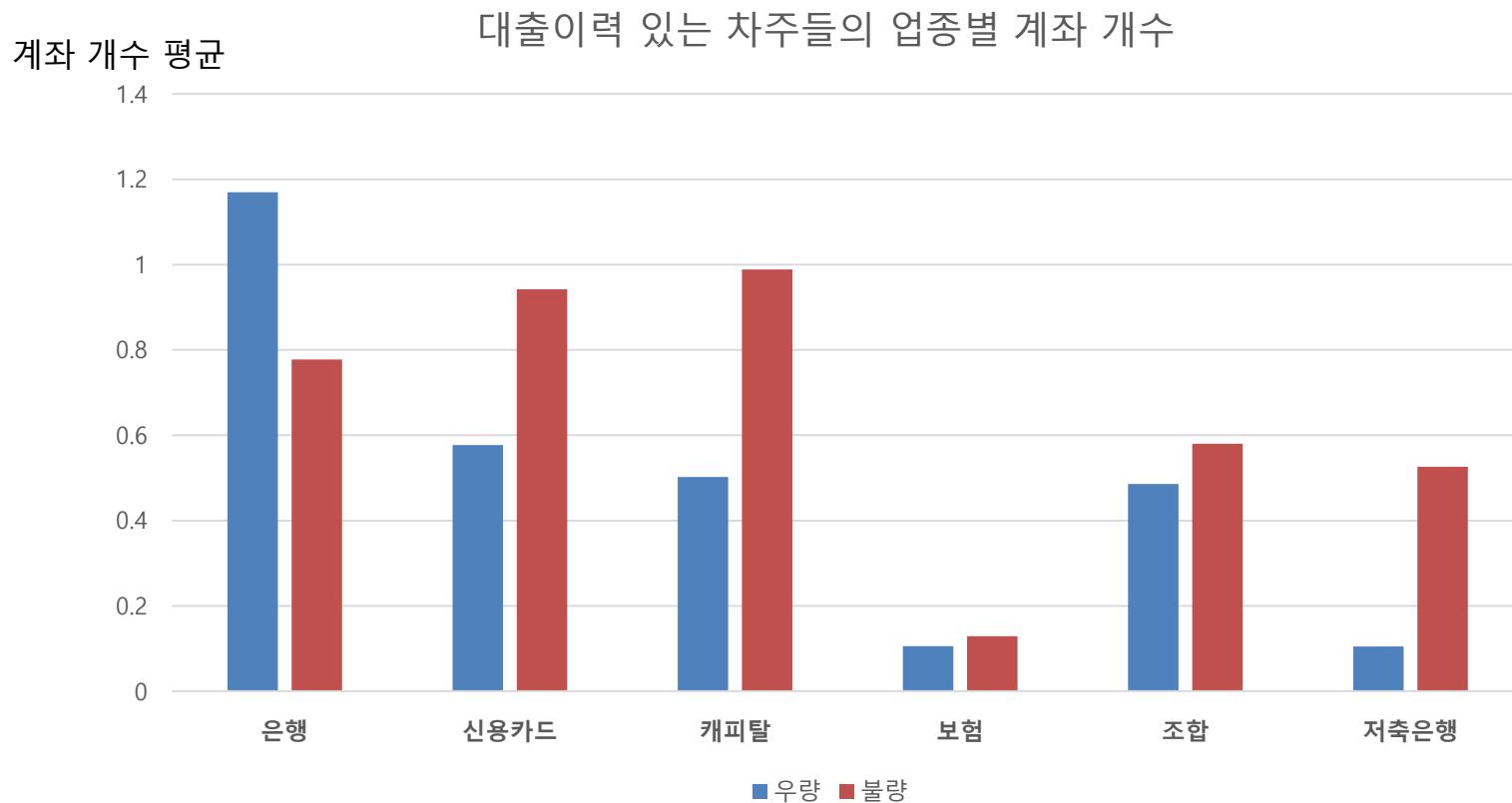
변수명	변수설명	변수유형
AGE	연령대(20대, 30대,..., 70대 이상)	범주형
INCOME	연 소득	수치형
DTI	총부채상환비율	수치형
JOB	직업(급여소득자, 자영업자, 무직, 정보 없음)	범주형
HOME	자택광역시(강원,...,충북)	범주형
SEX_CD	성별(남자, 여자)	범주형
LN_FLAG	대출보유여부(1:보유, 0:미보유)	범주형
CARD_FLAG	카드보유여부(1:보유, 0:미보유)	범주형
CNT_CARD_USE	이용 신용카드 기관수	수치형
TOT_USE_AMT	신용카드 총 이용금액	수치형
DLQ_N1YE_FLAG	우·불량여부(1:불량(21,389), 0:우량(822,225)), TARGET	범주형

I. 변수 소개 – 파생변수

변수명	변수설명	변수유형
JBLN0000	최근 1년 간 비 은행 비 주택담보대출 이용 대출 계좌 수	수치형
JBLN0002	최근 1년 간 비 은행 주택담보대출 이용 대출 계좌 수	수치형
JBLN0100	최근 1년 간 은행 비 주택담보대출 이용 대출 계좌 수	수치형
JBLN0102	최근 1년 간 은행 주택담보대출 이용 대출 계좌 수	수치형
MRTY_LEFT	기준시점 유효한 계좌들의 만기까지 남은 기간	수치형
INCOME_SD	최근 1년 간 연 소득 표준편차	수치형
CARD_MEAN_CNT	최근 1년 간 신용카드 평균 이용기관 수(0, 1이상으로 구분)	범주형
INCOME_CARD	최근 1년 간 월 소득 대비 신용카드 평균 이용금액	수치형
INCOME_DELAY_AMOUNT	최근 1년 간 연소득대비 연체잔액	수치형
CARD_DELAY_SCORE	카드연체점수. 연체시기가 기준시점에 가까울수록 높은 가중치 부여	수치형
ACCT_DELAY_SCORE	대출연체점수. 연체시기가 기준시점에 가까울수록 높은 가중치 부여	수치형

I. 변수 소개 – 파생변수 (JBLN0000 ~ JBLN0102)

1. 최근 1년간 업종별 대출 계좌 개수 (integer)



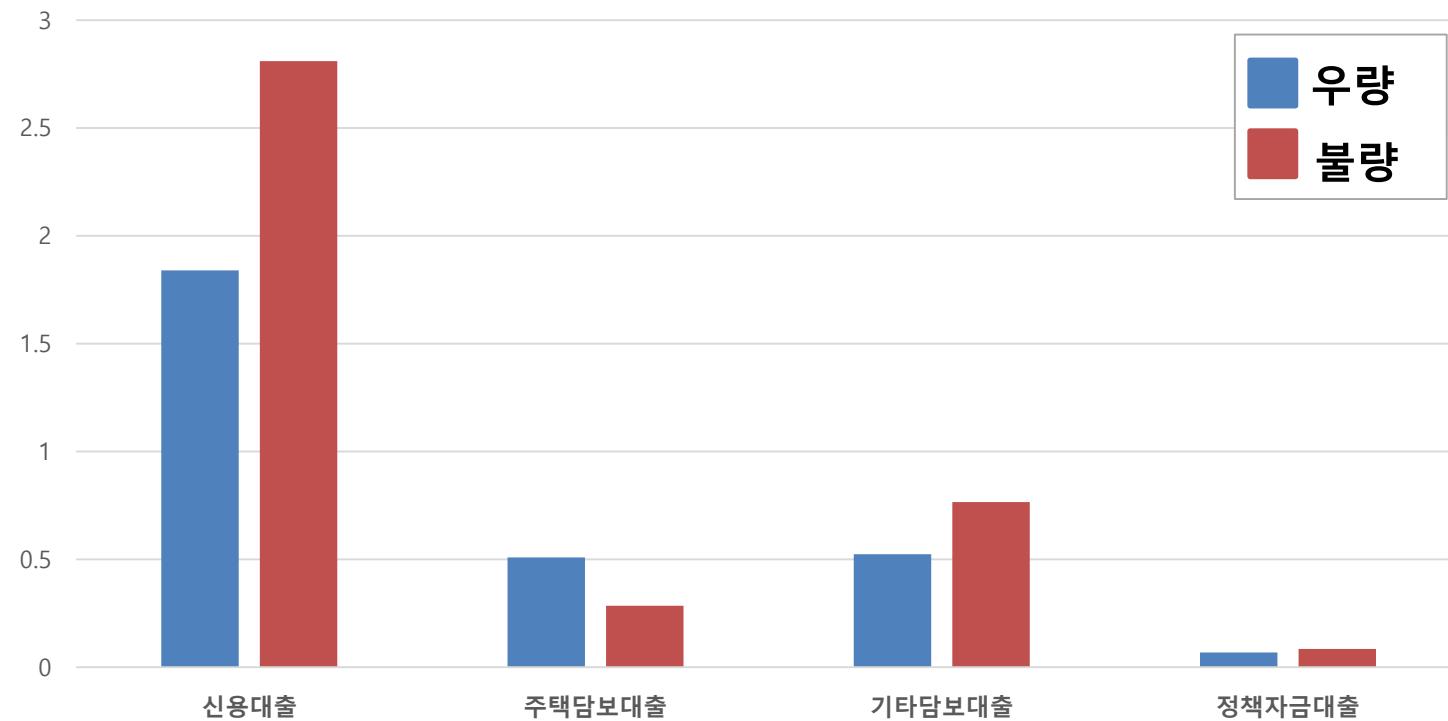
- 은행의 경우에서만 우량 차주 계좌 수 > 불량 차주 계좌 수
- 업종을 은행과 비 은행으로 통합

I. 변수 소개 – 파생변수 (JBLN0000 ~ JBLN0102)

2. 최근 1년간 상품별 대출 계좌 개수 (integer)

계좌 개수 평균

대출이력 있는 차주들의 상품별 계좌 개수



- 주택담보대출의 경우에는만 **우량 차주 계좌 수 > 불량 차주 계좌 수**
- 주택담보대출/비주택담보대출로 변수 통합

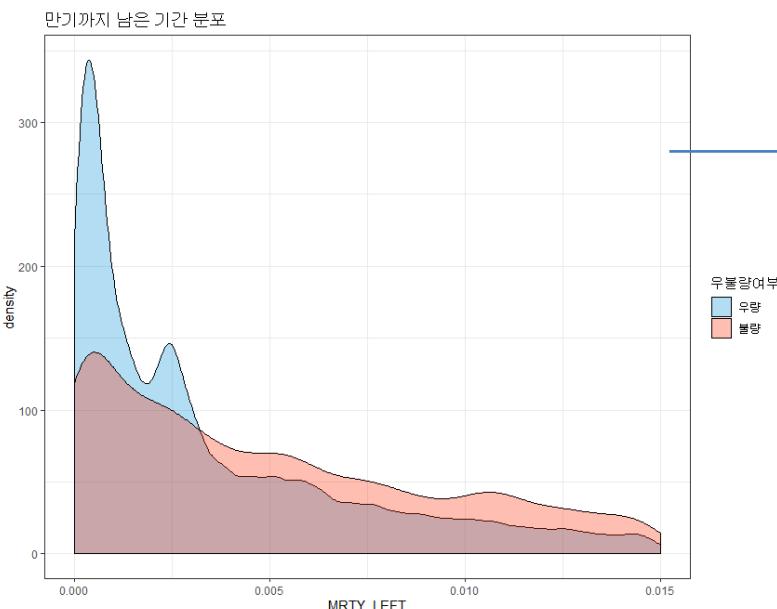
I. 변수 소개 – 파생변수 (MRTY_LEFT)

3. 만기까지 남은 기간 (numeric)

$$MRTY_LEFT_i := \sum (\text{업종별 만기까지 남은 기간} \times \text{업종별 불량률})$$

각 계좌의 만기까지 남은 기간(분기) 계산

- 상품별로 남은 기간 표준화: 신용대출에 비해 주택담보대출이 대출기간이 매우 긴 편
- 불량률이 높은 업종의 대출 계좌 만기가 많이 남아있을수록 MRTY_LEFT가 큼



▪ 불량 차주일수록 MRTY_LEFT
가 전반적으로 높음

I. 변수 소개 – 파생변수 (INCOME_SD)

4. 최근 1년 간 연 소득 표준편차

- 연 소득의 변동이 클수록 불량 확률이 증가할 것으로 추측

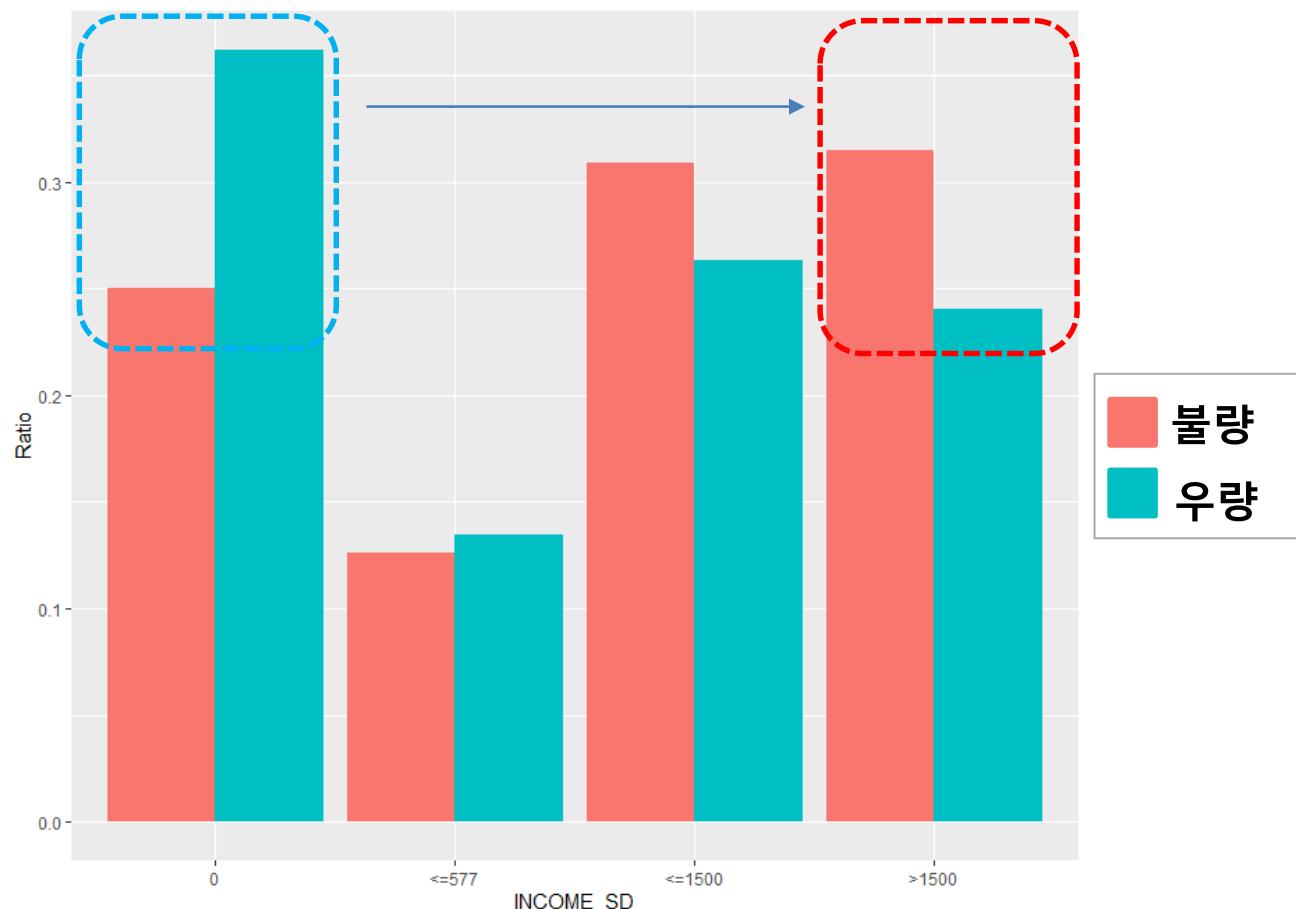
e.g.) 차주ID: 120009530632

BS_YR_MON	INCOME	DLQ_N1YE_FLAG
201609	22,639,720	0
201612	2,253,700	1
201703	90,000	1
201706	829,630	1
201709	1,590,690	1
201712	77,000	0
201803	75,000	0
201806	80,000	0

I. 변수 소개 – 파생변수 (INCOME_SD)

4. 최근 1년 간 연 소득 표준편차

- 확인해본 결과, 연 소득 표준편차가 클 때 불량률이 높다

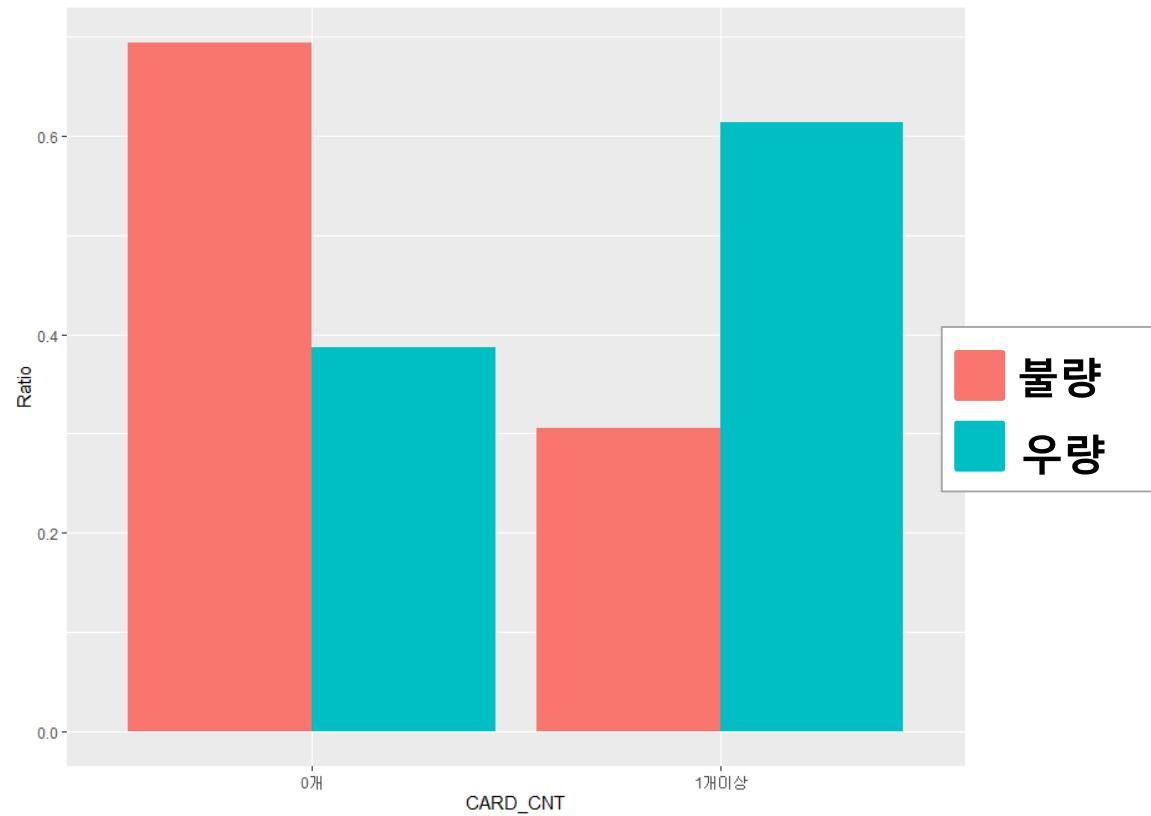


I. 변수 소개 – 파생변수 (CARD_MEAN_CNT)

5. 최근 1년 간 신용카드 평균 이용 기관 수 (factor)

- 이용 기관 수를 0개와 1개 이상으로 구분: 신용카드 이용자가 비이용자에 비해 우량 비율이 높음

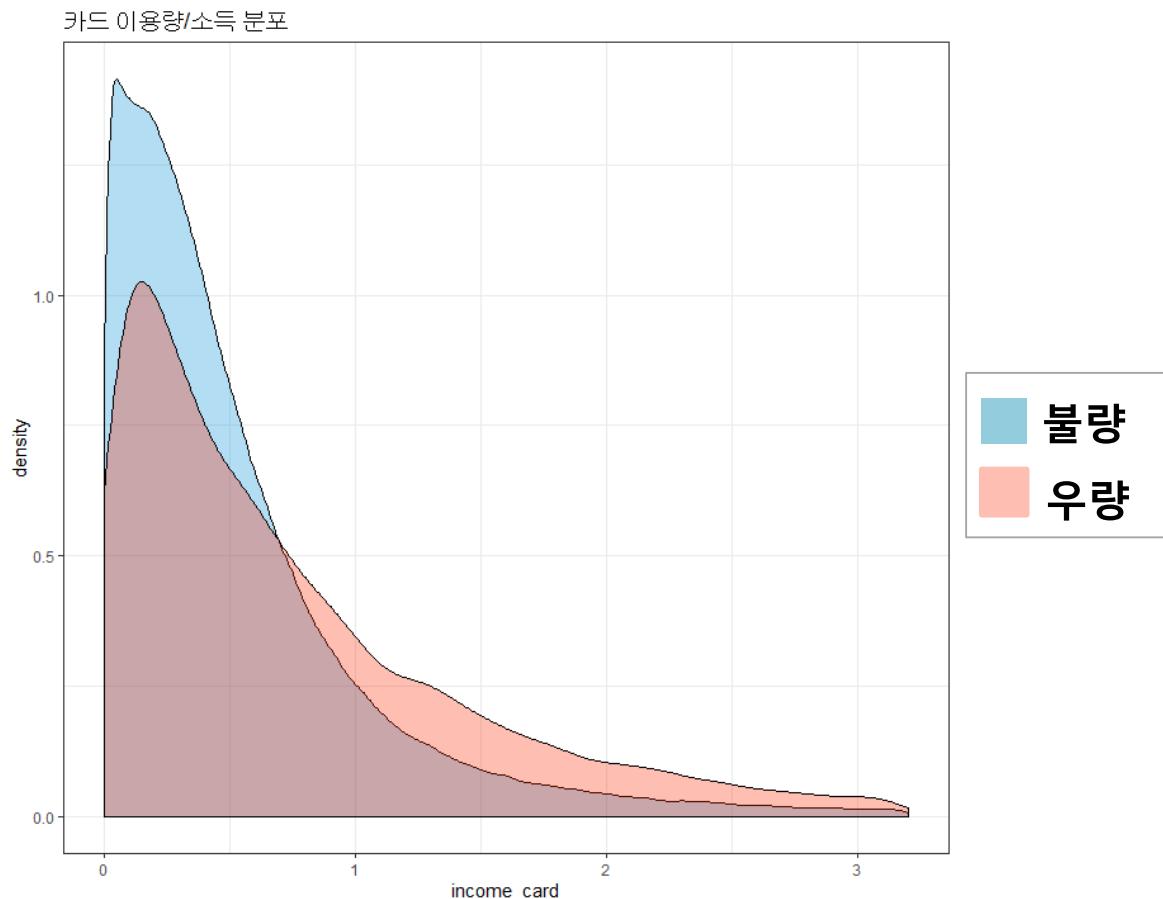
Summary of CARD_CNT	
MIN	0.0
1 st Quartile	0.0
Median	1.0
3 rd Quartile	1.166
MAX	11.250



I. 변수 소개 – 파생변수 (INCOME_CARD)

6. 최근 1년 간 소득 대비 신용카드 평균 이용금액 (numeric)

$$\text{INCOME_CARD}_i := \frac{1}{n} \sum \frac{\text{월별 신용카드 이용금액}}{\text{월별 소득}}$$



I. 변수 소개 – 파생변수 (INCOME_DELAY_AMOUNT)

7. 최근 1년 간 소득 대비 연체 잔액 평균(number)

$$\text{INCOME_DELAY_AMOUNT}_i := \frac{1}{n} \sum \frac{\text{연체잔액}}{\text{연소득}}$$

- 연소득 대비 총 연체잔액이 클수록 불량률이 높음



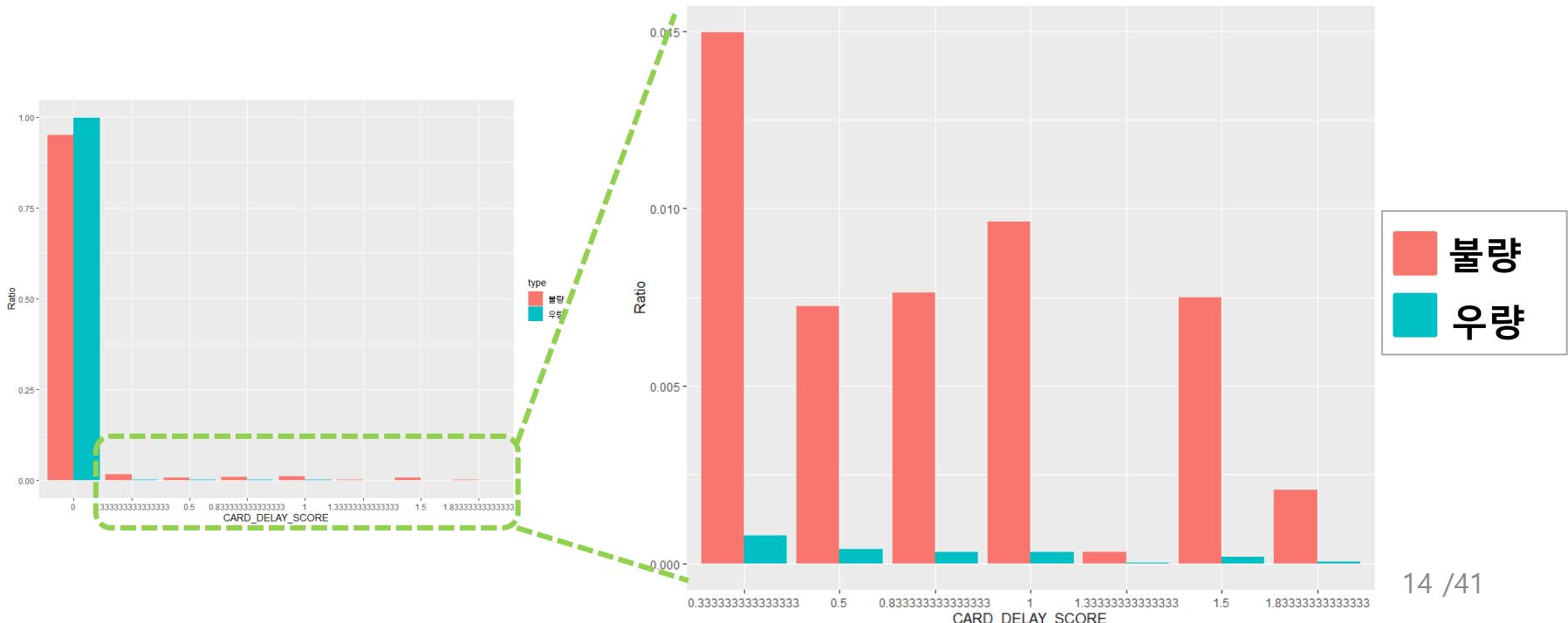
I. 변수 소개 – 파생변수 (CARD_DELAY_SCORE)

8. 신용카드 연체 점수 (number)

$$\text{CARD_DELAY_SCORE}_i := \sum \frac{1}{\text{기준시점 분기} - \text{연체시점 분기}}$$

- 과거 1년 동안의 연체 여부를 이용해 점수 계산
- 연체시기가 기준시점에 가까울수록 높은 가중치를 부여

Ex) 201706 기준 201609 연체, 201703 연체 $\rightarrow \text{CARD_DELAY_SCORE} = 1/3 + 1/1 = 1.3333333$



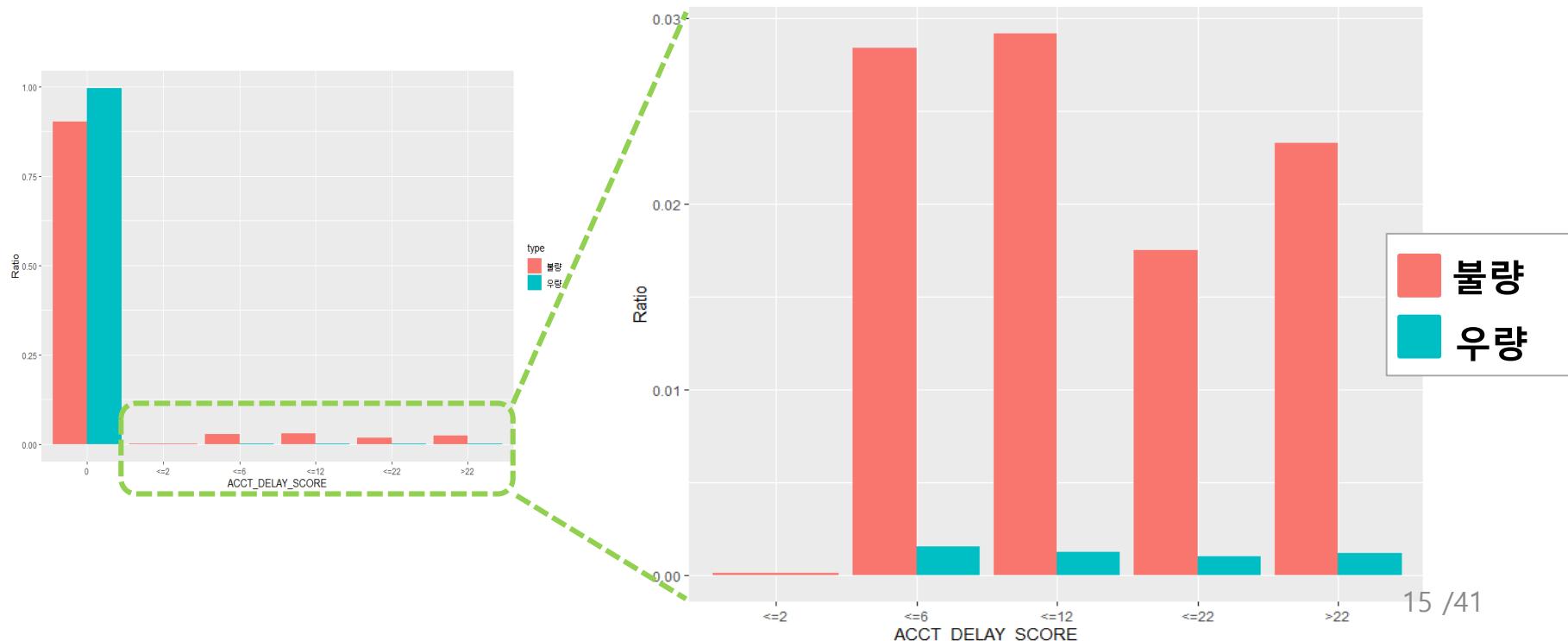
I. 변수 소개 – 파생변수 (ACCT_DELAY_SCORE)

9. 대출 연체 점수 (number)

$$ACCT_DELAY_SCORE_i := \sum \frac{1}{\text{기준시점 분기} - \text{연체시점 분기}}$$

- 과거 1년 동안의 연체 여부를 이용해 점수 계산
- 연체시기가 기준시점에 가까울수록 높은 가중치를 부여

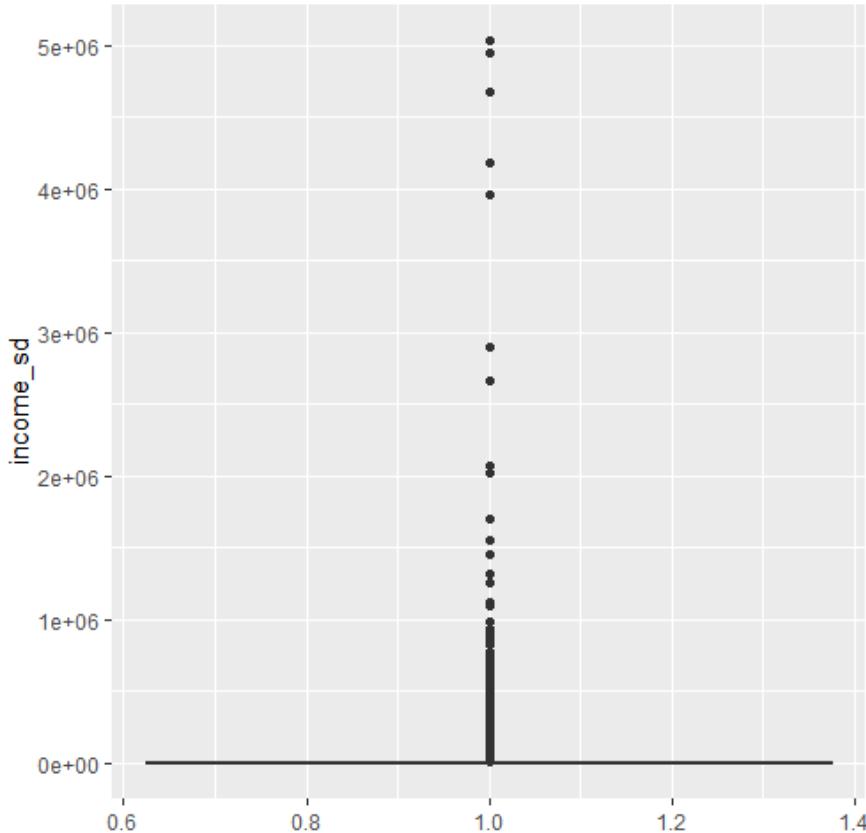
Ex) 201706 기준 201703 연체 + 201606 연체 $\rightarrow ACCT_DELAY_SCORE = 1/3 + 1/1 = 1.3333333$



I. 변수 소개 – 전처리 (표준화)

Robust Normalization

- INCOME_SD, DTI 등 변수들의 경우, 극단적으로 높은 값들이 존재.
- 극단적인 값에 Robust할 수 있도록, Robust Normalization 방법을 사용.



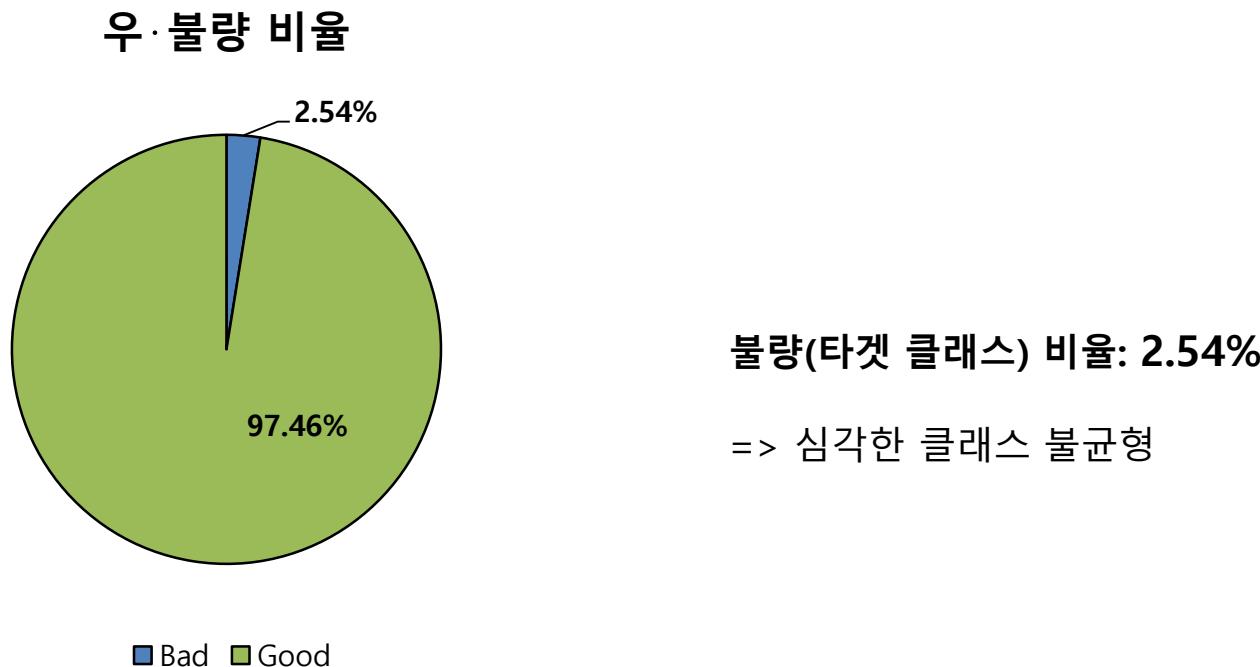
<Outlier 예시>

Robust Normalization

$$x_{norm, i} = \frac{x_i - q_2}{q_3 - q_1}$$

- q_i : i 번째 quartile.

II-(1). 샘플링 – 클래스 불균형 문제



- 문제점

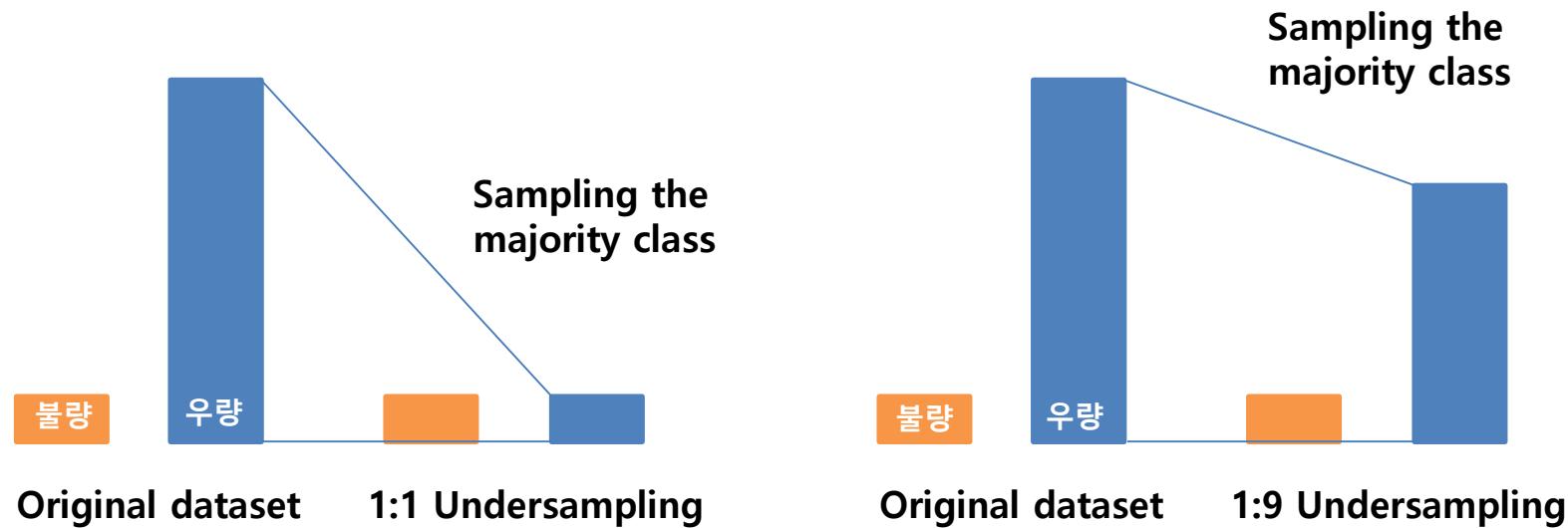
- 다수 클래스(우량)을 잘못 예측하는 것보다 소수 클래스(불량)을 잘못 예측하는 것의 비용이 더 큼

- 대처 방법

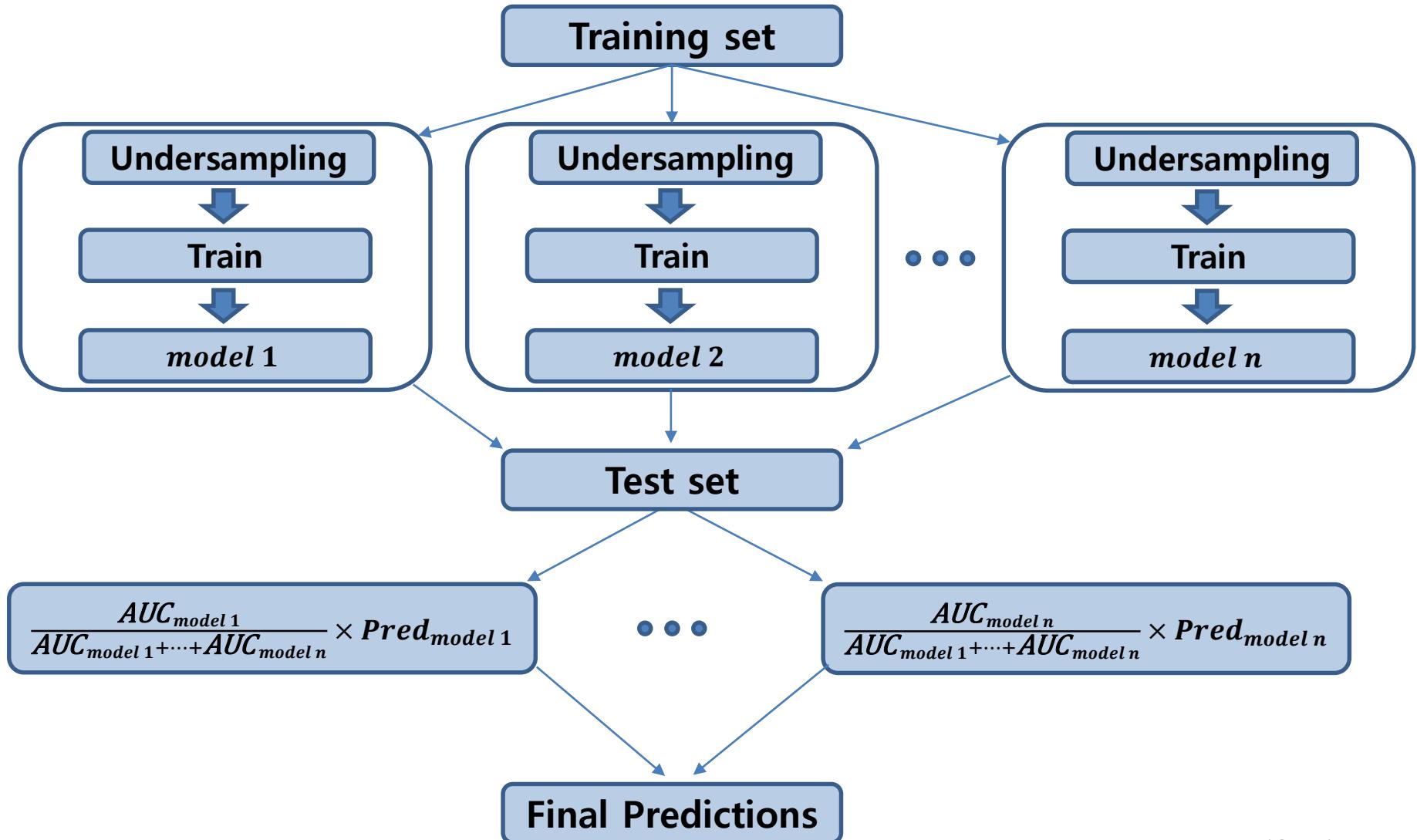
- Undersampling, Undersampling + Ensemble, Balance cascade, Oversampling 등

1. Undersampling

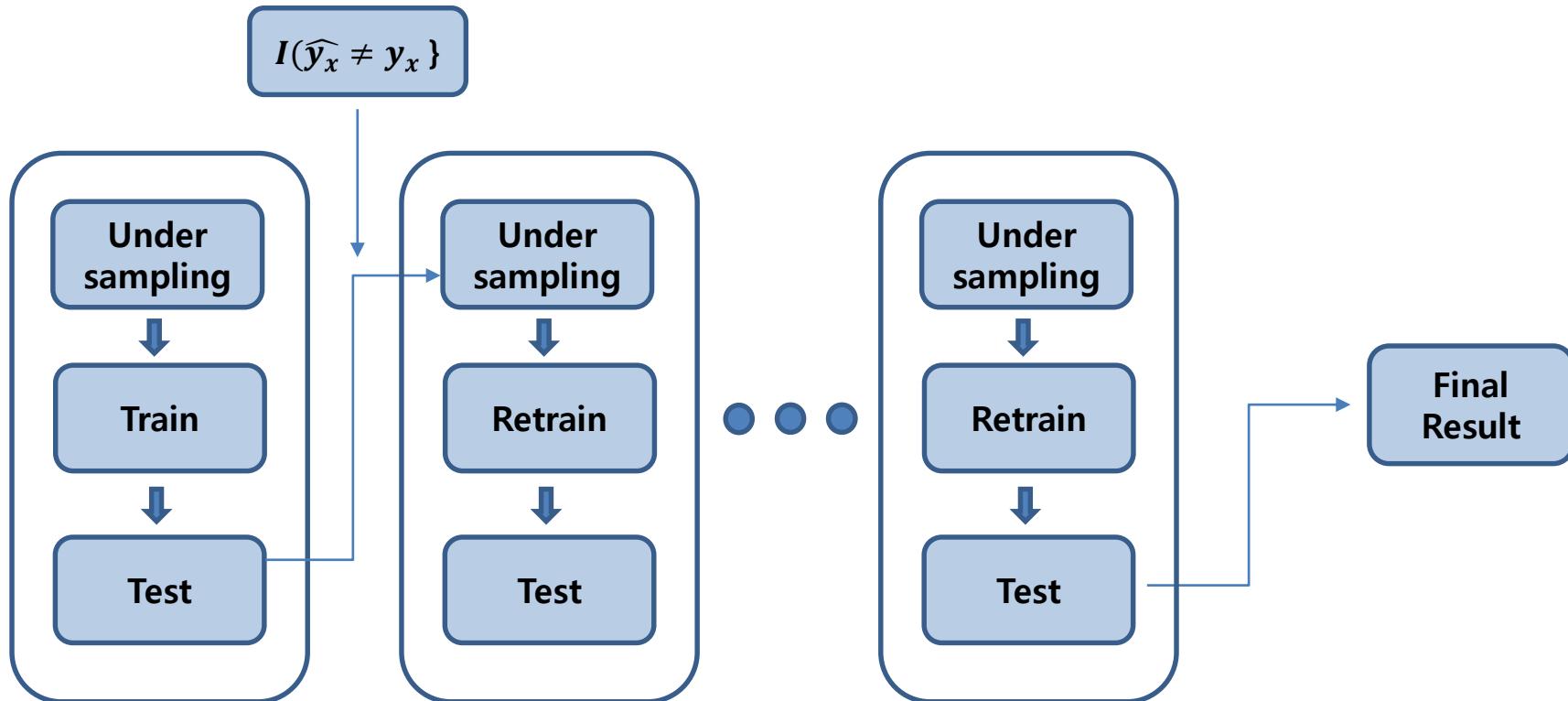
- 다수 클래스의 데이터를 랜덤 샘플링
- 1:1, 1:9 비율로 undersampling



2. Undersampling + Ensemble

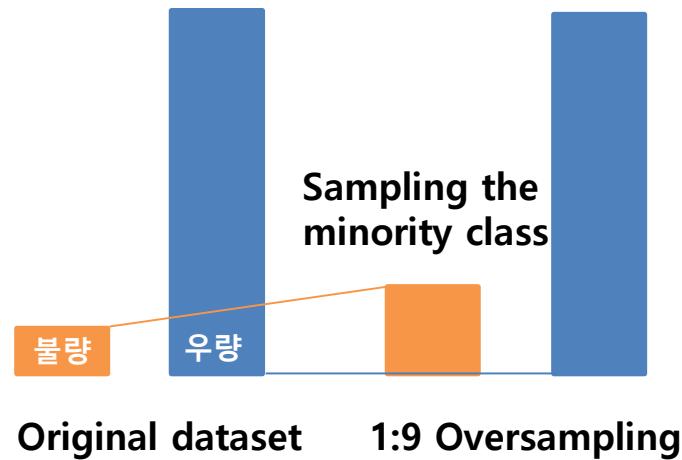
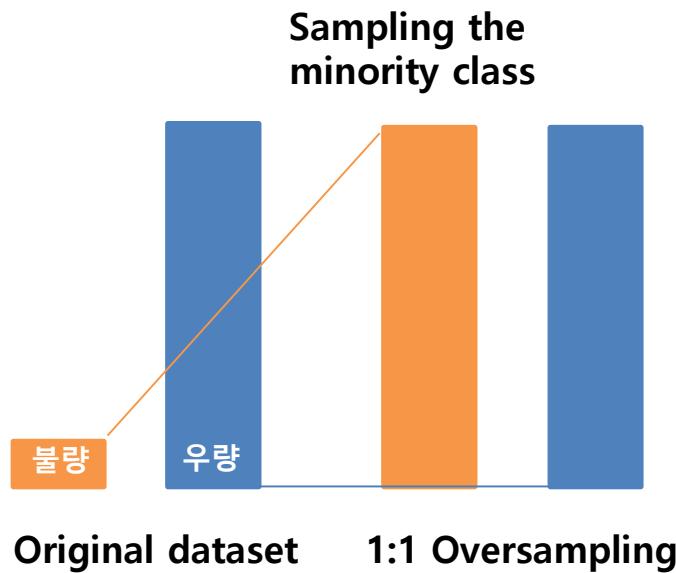


3. Balance cascade

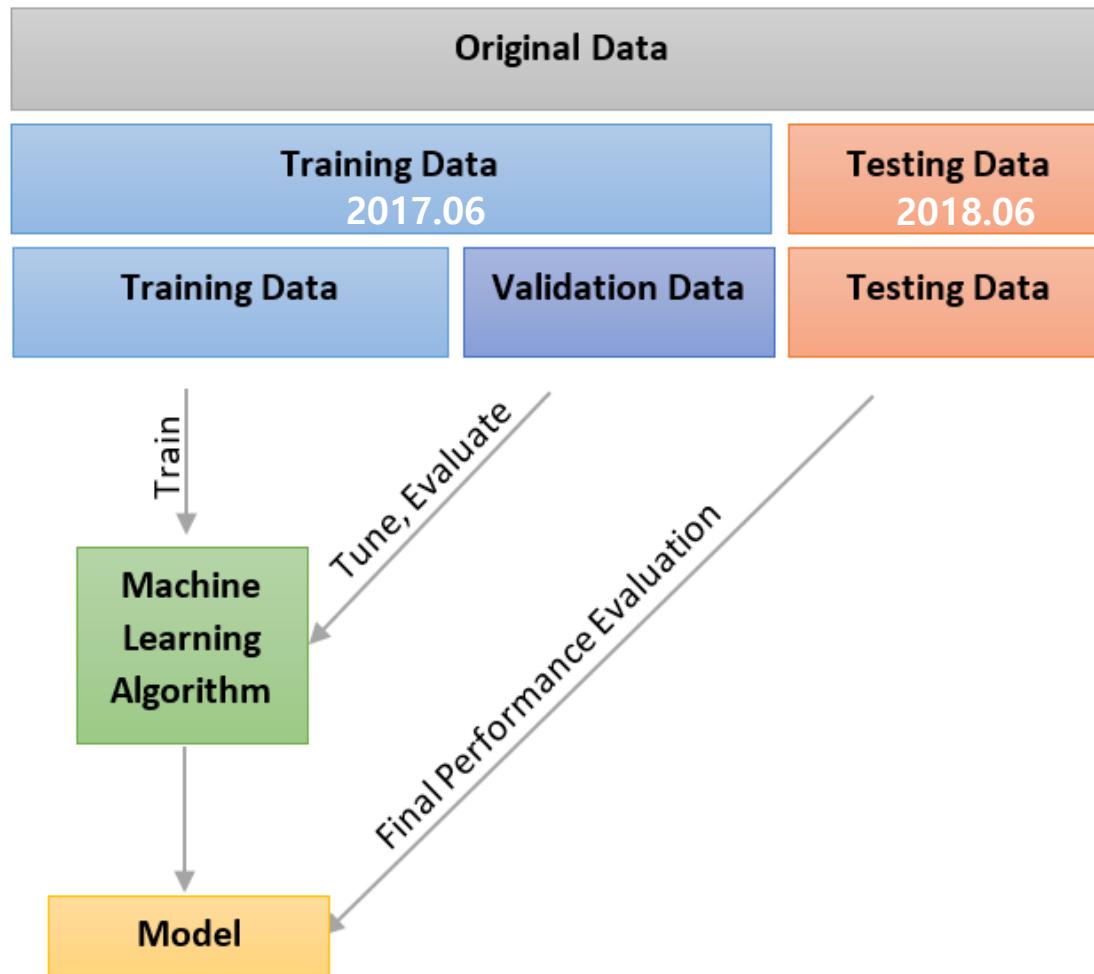


4. Oversampling

- Undersampling은 데이터의 일부를 버리는 것이므로 정보의 손실 발생
- 소수 클래스의 데이터를 반복 추출
- 1:1, 1:9 비율로 oversampling



II-(2). 모델링 – 데이터 분할

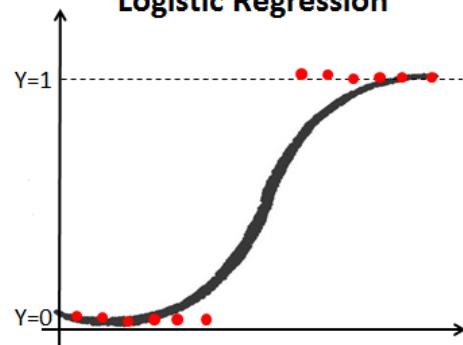


II-(2). 모델링 – 사용 모형

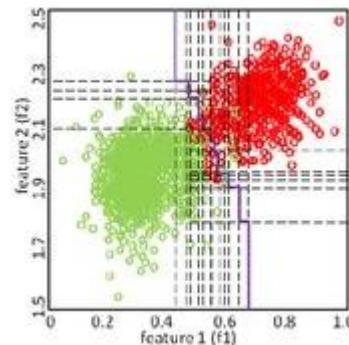
로지스틱 회귀

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Logistic Regression



랜덤 포레스트



XGBoost

Cache awareness and out-of-core computing

Regularization for avoiding overfitting

Tree pruning using depth-first approach

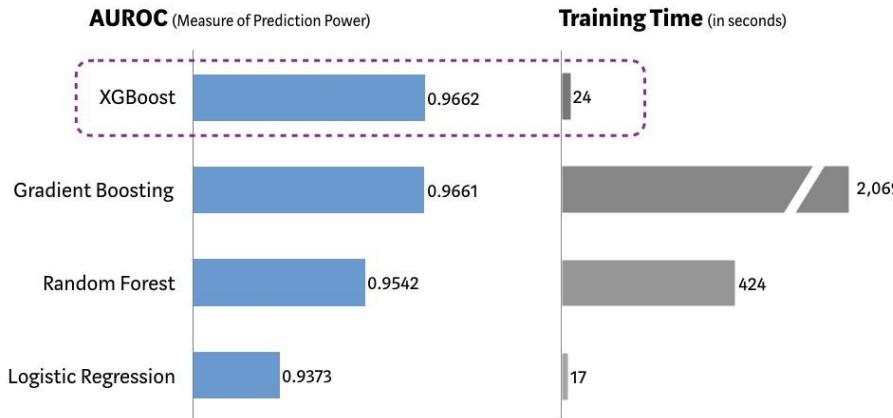
Parallelized tree building

XGBoost

Efficient handling of missing data

In-built cross-validation capability

Performance Comparison using SKLearn's 'Make_Classification' Dataset
(5 Fold Cross Validation, 1MM randomly generated data sample, 20 features)



II-(3). 모형 변별력 확인

Positive class = 불량

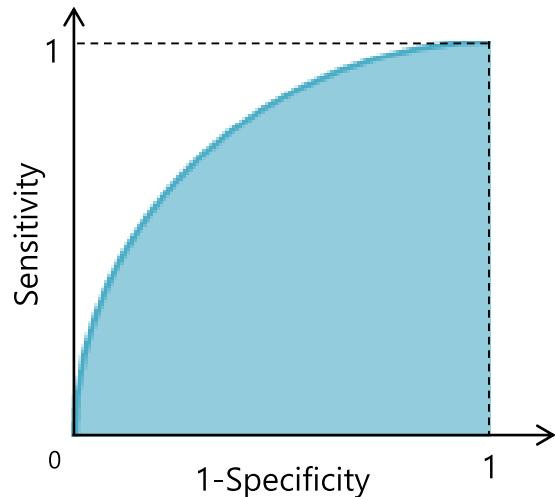
	실제 불량	실제 우량
예측 불량	TP	FP
예측 우량	FN	TN

$$Sensitivity = Recall = \frac{TP}{TP + FN}$$

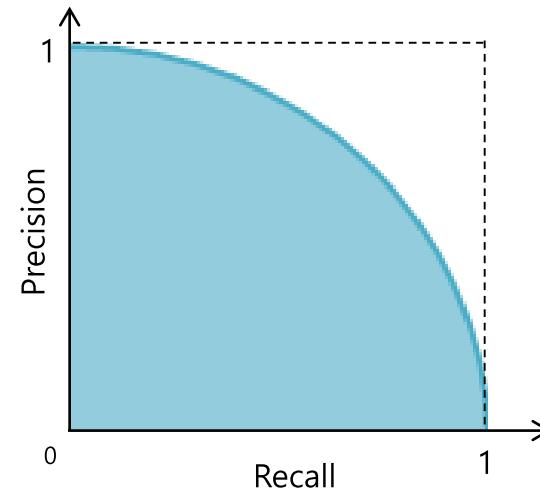
$$Specificity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

- **ROC-AUC:** ROC 곡선의 아래 면적

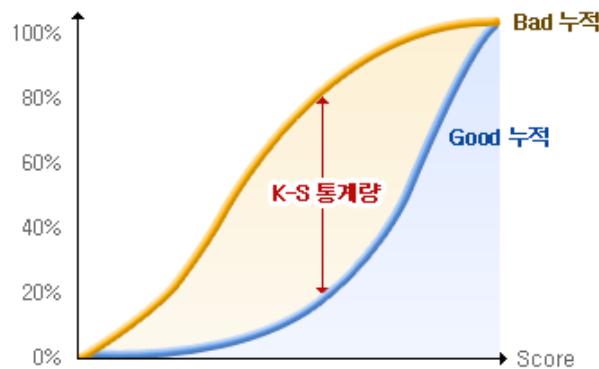


- **PR-AUC:** Precision-Recall 곡선의 아래 면적

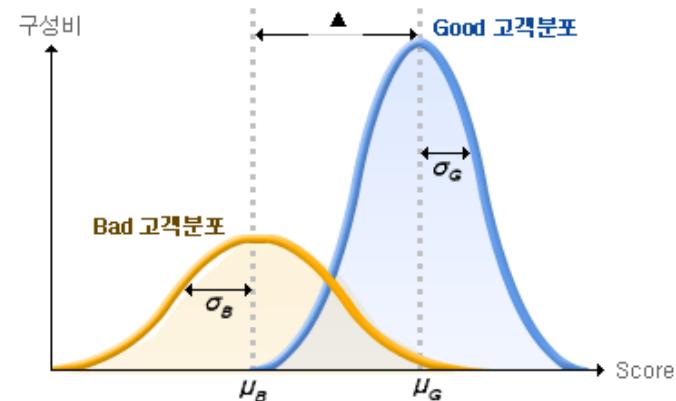


II-(3). 모형 변별력 확인

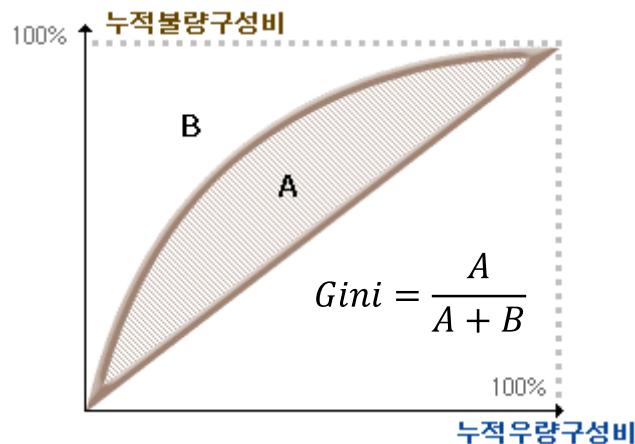
- K-S Statistics



- Divergence



- GINI index



성능지표	적정 기준치
K-S(Kolmogorov-Smirnov)	50 이상
Divergence	1 이상
Gini	0.6 이상

[출처] 올크레딧

II-(3). 모형 변별력 확인

Baseline (Validation)				
	AUC	PRAUC	KS	GINI
LR	0.765	0.146	38.24	0.530
RF	0.797	0.176	44.53	0.593
XGB	0.829	0.206	50.38	0.658

Without Sampling (Validation)				
	AUC	PRAUC	KS	GINI
LR	0.794	0.188	44.29	0.587
RF	0.833	0.251	50.90	0.665
XGB	0.849	0.259	53.04	0.697

Undersampling 1:1 (Validation)				
	AUC	PRAUC	KS	GINI
LR	0.806	0.185	46.80	0.613
RF	0.841	0.198	52.30	0.682
XGB	0.846	0.242	52.70	0.692

Undersampling 1:9 (Validation)				
	AUC	PRAUC	KS	GINI
LR	0.799	0.189	45.31	0.600
RF	0.844	0.254	52.87	0.688
XGB	0.849	0.257	53.01	0.698

Undersampling 1:9 Ensemble (Validation)				
	AUC	PRAUC	KS	GINI
LR	0.800	0.192	45.53	0.599
RF	0.845	0.255	52.94	0.691
XGB	0.849	0.261	52.92	0.698

Undersampling 1:9 Balance Cascade (Validation)				
	AUC	PRAUC	KS	GINI
LR	0.758	0.187	37.54	0.516
RF	0.805	0.239	46.06	0.611
XGB	0.822	0.255	48.43	0.645

Oversampling 1:1 (Validation)				
	AUC	PRAUC	KS	GINI
XGB	0.824	0.198	49.44	0.641

Oversampling 1:9 (Validation)				
	AUC	PRAUC	KS	GINI
XGB	0.824	0.210	49.42	0.640

II-(3). 모형 변별력 확인

〈표 III-3〉

국내외의 차별방지 입법례

국 가	관련법규	차별금지 항목
한 국	국가인권위법	성별, 종교, 장애, 나이, 사회적 신분, 출신 지역(출생지), 등록 기준지, 성년이 되기 전의 주된 거주지 등), 출신 국가, 출신 민족, 용모 등 신체조건, 기혼·미혼·별거·이혼·사별·재혼·사실 혼 등 혼인 여부, 임신 또는 출산, 가족 형태 또는 가족 상황, 인종, 피부색, 사상 또는 정치적 의견, 형의 효력이 실효된 전과(前科), 성적(性的) 지향, 학력, 병력(病歷) 등

<KIF 한국금융연구원 금융동향: 분석과 전망 2012년 10월호 제 3부 주요 금융 이슈>

< 민감정보(성별, 나이, 거주지) 제외한 1:9 undersampling 모형>

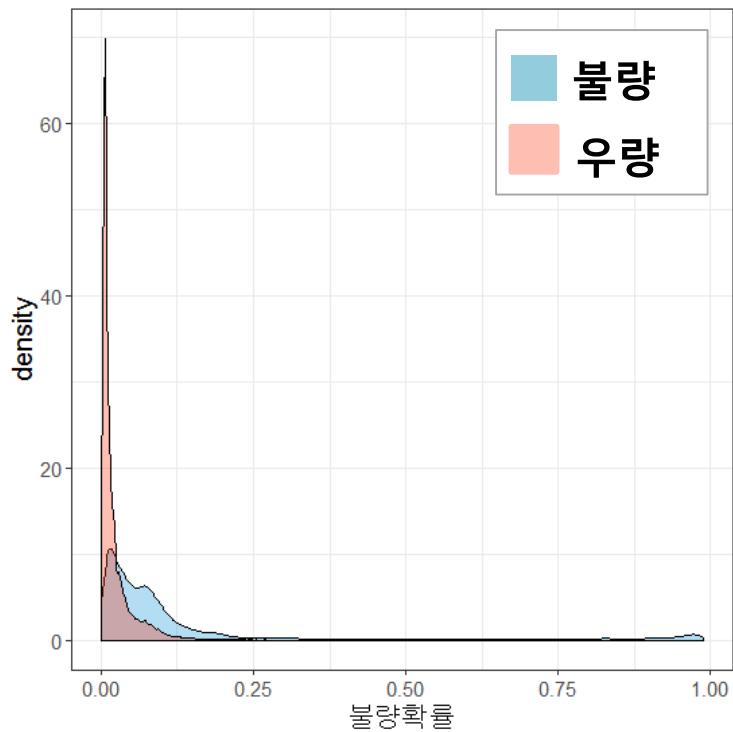
Undersampling 1:9 (Val)			
	AUC	PRAUC	KS
LR	0.799	0.189	45.31
RF	0.844	0.254	52.87
XGB	0.849	0.257	53.01
			0.698



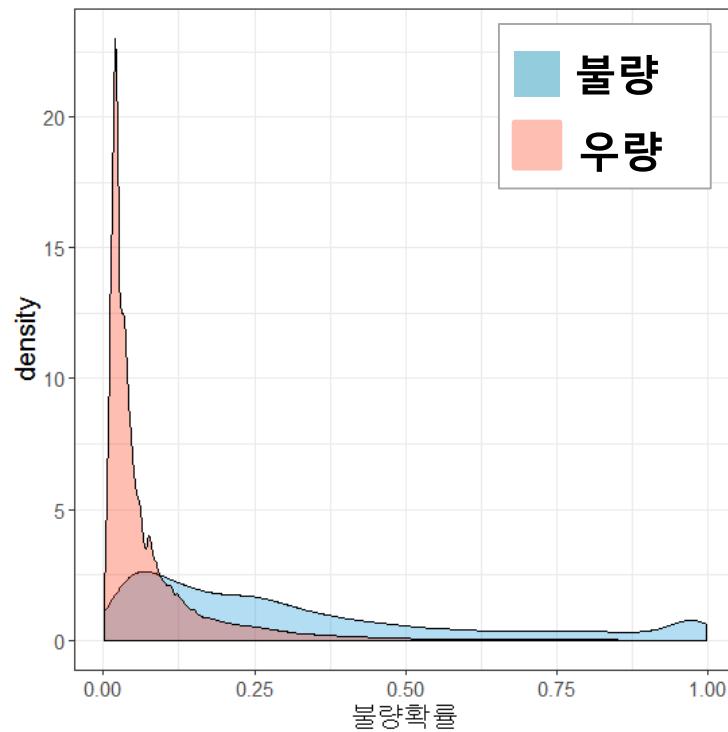
Undersampling 1:9 (Val) - 민감정보 제외				
	AUC	PRAUC	KS	GINI
LR	0.776	0.184	43.38	0.553
RF	0.828	0.247	49.99	0.657
XGB	0.832	0.250	50.54	0.665

II-(3). 모형 변별력 확인

< Baseline 모형 >



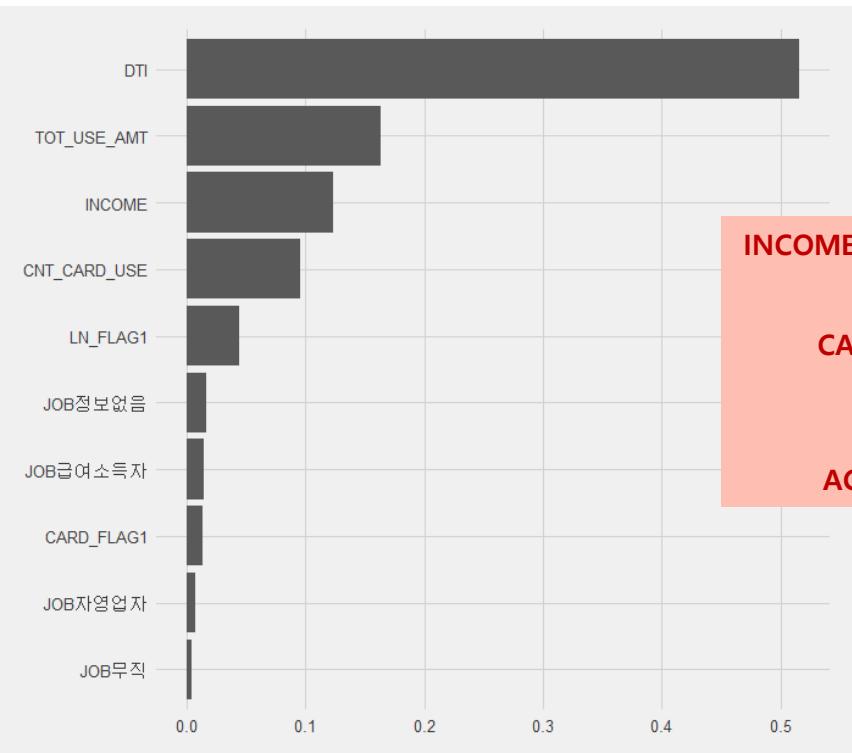
< 최종 모형 >



	Baseline (Test)	최종모형 (Test)
KS	50.5396	53.6529
Divergence	0.47	1.36

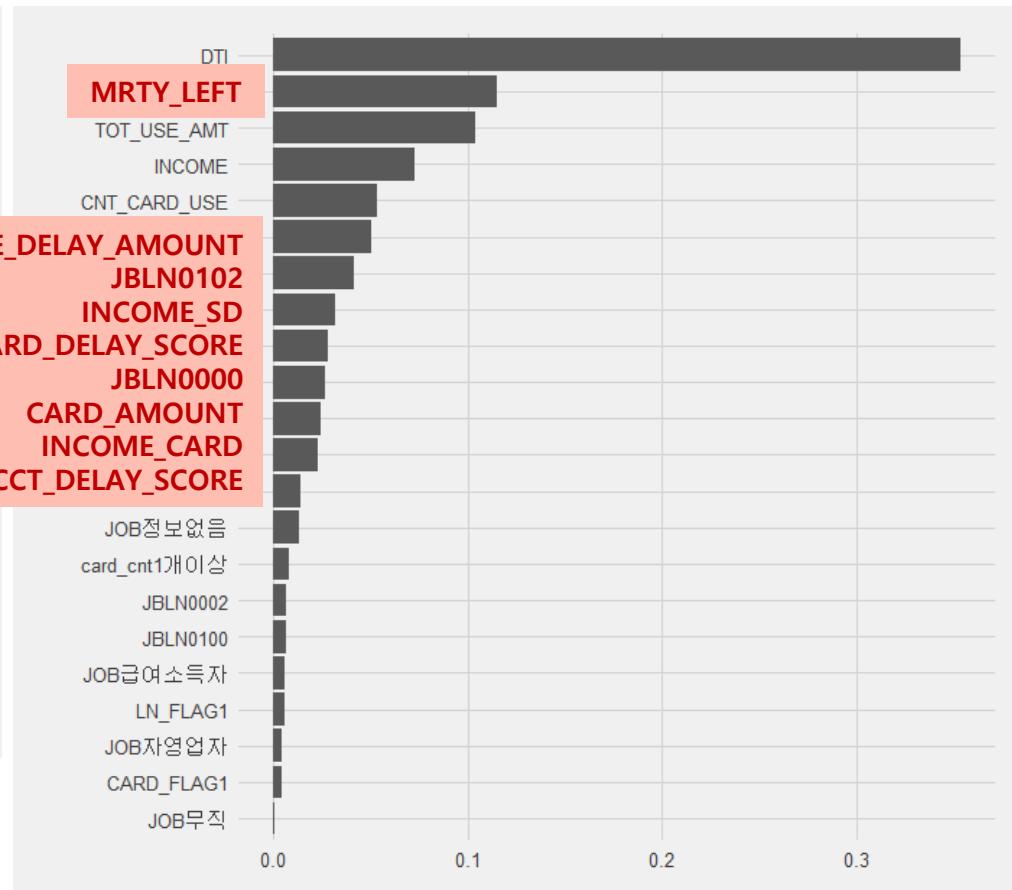
II-(3). 모형 변별력 확인 - 변수 중요도

<Baseline 모형>



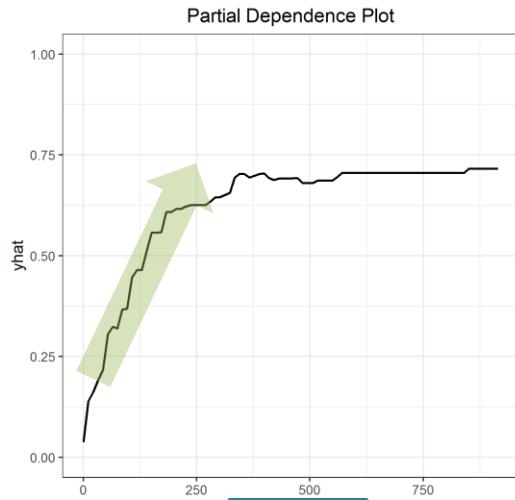
INCOME_DELAY_AMOUNT
JBLN0102
INCOME_SD
CARD_DELAY_SCORE
JBLN0000
CARD_AMOUNT
INCOME_CARD
ACCT_DELAY_SCORE

<최종 모형>

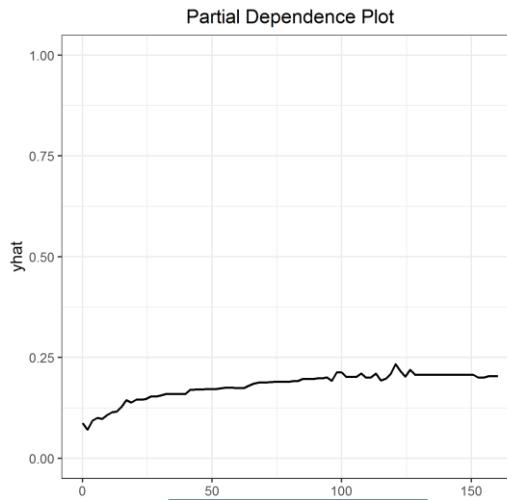


- 추가한 파생변수들의 변수 중요도가 대부분 높게 나타남

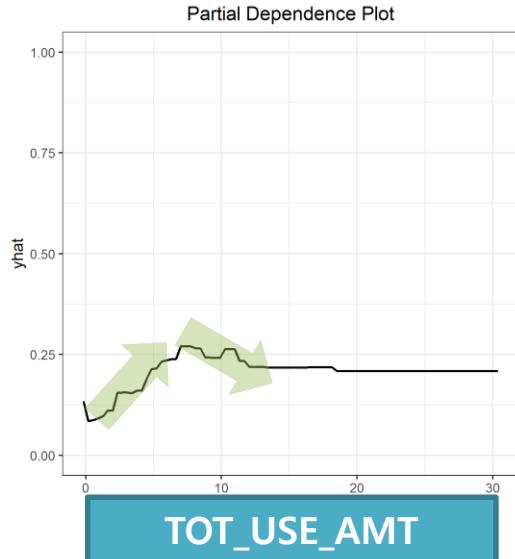
II-(3). 모형 변별력 확인 – Partial Dependence Plot



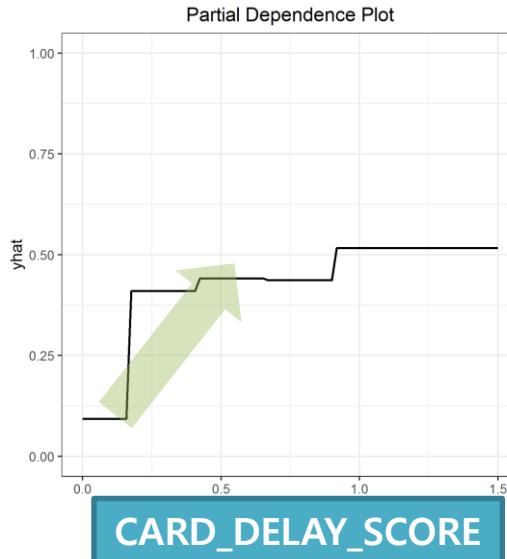
DTI



MRTY_LEFT



TOT_USE_AMT



CARD_DELAY_SCORE

- 모형의 평균적인 예측결과 파악
- 변수의 비선형적 패턴, 특히 예측값이 증가하다 감소하는 패턴도 파악 가능
- 각 변수가 우·불량 여부에 어떤 방향으로 영향을 미치는지 파악 가능

II-(3). 모형 안정성 확인

- PSI(Population Stability Index)

$$PSI = D_{KL}(\hat{q}(x) \mid \hat{p}(x)) + D_{KL}(\hat{p}(x) \mid \hat{q}(x))$$

$$\text{where } D_{KL}(q(x) \mid p(x)) = E_p \left(\ln \left(\frac{p(X)}{q(X)} \right) \right) = \sum_{i=1}^B p(x_i) \ln \frac{p(x_i)}{q(x_i)}$$

$\hat{p}(x)$: Validation set에서 각 cutoff 구간에 속하는 비율

$\hat{q}(x)$: Test set에서 각 cutoff구간에 속하는 비율

[출처] Yurdakul, Bilal, "Statistical Properties of Population Stability Index" (2018). Dissertations. 3208.

⇒ Validation set과 Test set에서 **두 확률분포의 차이**를 계산

	Baseline	최종모델
PSI	0.000505 <<<<0.1	0.000622 <<<<0.1

II-(4). 추가 고려사항

(1) 대출상품에 따른 모델 변별력

- 신용대출 보유자와 담보대출 보유자 모두 Baseline 모형보다 **최종모형의 성능이 더 좋았음**

Baseline (Test) - 민감정보 제외				
	AUC	PRAUC	KS	GINI
신용대출	0.7786	0.1999	40.7136	0.5572
담보대출	0.7715	0.1453	40.4249	0.543



Undersampling 1:9 (Test) - 민감정보 제외				
	AUC	PRAUC	KS	GINI
신용대출	0.8193	0.2617	47.5050	0.6385
담보대출	0.8302	0.2423	50.5845	0.6605

II-(4). 추가 고려사항

(2) 업종에 따른 모델 변별력

- 은행권의 경우에는 결과가 매우 좋은 편
- **캐피탈, 저축은행**의 경우에 은행권에 비해 모형의 변별력이 낮음

Baseline (Test) - 민감정보 제외				
	AUC	PRAUC	KS	GINI
은행	0.7881	0.1194	43.8720	0.5762
저축은행 +캐피탈	0.7068	0.2397	28.7530	0.4136



Undersampling 1:9 (Test) - 민감정보 제외				
	AUC	PRAUC	KS	GINI
은행	0.8423	0.2009	53.2428	0.6846
저축은행 +캐피탈	0.7581	0.2986	36.8929	0.5162

II-(4). 추가 고려사항

(2) 업종에 따른 모델 변별력

- 개선 방안

INCOME_DELAY_AMOUNT 변수를 계산할 때 **업종별 불량률**을 곱해서 같은
연체 금액이더라도 **캐피탈, 저축은행의 연체금액에 대한 높은 가중치**를 줌

$$\text{INCOME_DELAY_AMOUNT}_i := \sum \left(\frac{\text{업종별 연체잔액}}{\text{연소득}} \times \text{업종별 불량률} \right)$$

II-(4). 추가 고려사항

(2) 업종에 따른 모델 변별력

- 파생변수(INCOME_DELAY_AMOUNT) 수정한 결과 (전체 차주)

Undersampling 1:9 (Test)				
	AUC	PRAUC	KS	
XGB	0.848	0.267	53.65	0.696

Undersampling 1:9 (Test) – weighted mean 적용				
	AUC	PRAUC	KS	
XGB	0.826	0.222	50.29	0.650



- 파생변수(INCOME_DELAY_AMOUNT) 수정한 결과 (저축은행 + 캐피탈)

Undersampling 1:9 (Test)				
	AUC	PRAUC	KS	
XGB	0.7581	0.2986	36.8929	0.5162

Undersampling 1:9 (Test) – 저축은행+캐피탈				
	AUC	PRAUC	KS	
XGB	0.7653	0.2401	39.3614	0.5259



- 전체 차주에 대한 모형의 변별력은 나빠지지만 저축은행과 캐피탈 업종에 대한 변별력은 좋아짐

III-(1). 분석의의

- 분석 의의 ①
- 최종 모형인 1:9 Undersampling 은 다른 샘플링 기법에 비해 학습에 소요되는 시간이 적으면서 성능은 좋게 나타남

Undersampling 1:9 (Validation)				
	AUC	PRAUC	KS	GINI
LR	0.799	0.189	45.31	0.600
RF	0.844	0.254	52.87	0.688
XGB	0.849	0.257	53.01	0.698

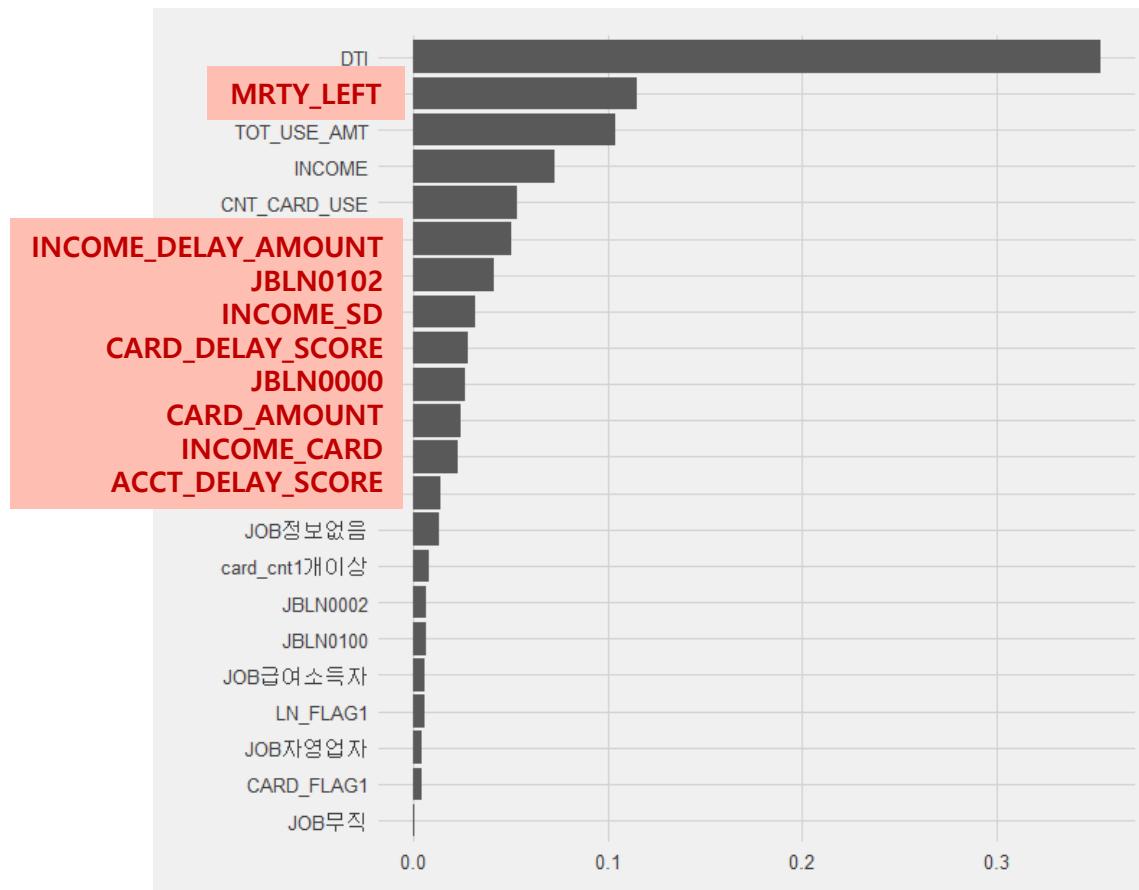
Undersampling 1:9 Ensemble (Validation)			
	AUC	PRAUC	KS
LR	0.800	0.192	45.53
RF	0.845	0.255	52.94
XGB	0.849	0.261	52.92

Undersampling 1:9 Balance Cascade (Validation)				
	AUC	PRAUC	KS	GINI
LR	0.758	0.187	37.54	0.516
RF	0.805	0.239	46.06	0.611
XGB	0.822	0.255	48.43	0.645

III-(1). 분석의의

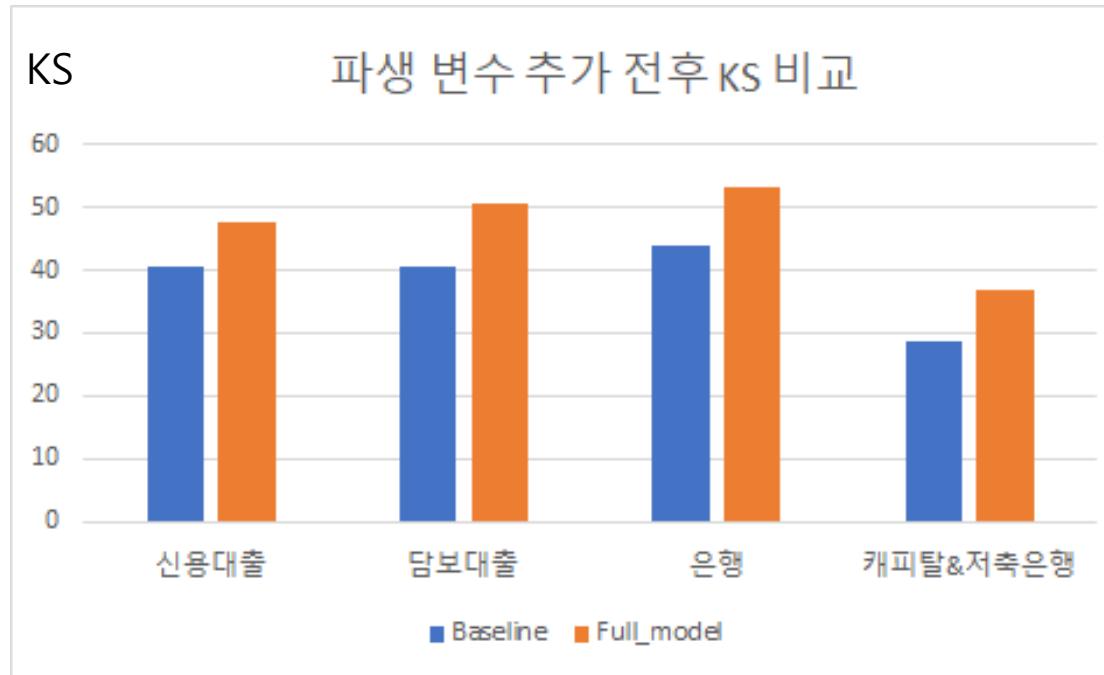
- 분석 의의 ②
- 추가한 파생변수들의 변수 중요도가 대부분 높게 나타남

변수중요도



III-(1). 분석의의

- 분석 의의 ③
- 모든 업종 및 상품에 효과적인 파생변수 생성



- 전체 모집단 외 특정 집단 대상으로도 변별력 있도록 모형 구성

III-(2). 발전방향

(1) 대출정보가 없는 차주에 의해 변별력이 낮아지는 현상

- 파생변수는 대부분 대출정보를 활용한 것인데 **대출정보가 없는 사람들은 값이 모두 0으로 들어가는 문제**
- 대출정보 보유 여부로 차주를 나누어 최종모형의 변별력을 확인했을 때, **대출정보가 없는 경우 대출정보가 있는 경우에 비해 변별력이 매우 낮음**
- 대출정보가 있는 차주와 없는 차주 각각에 대해 개별 모형을 만들어 예측한 결과도 크게 다르지 않았음

Undersampling 1:9 (Test) – 전체모형에서 예측				
	AUC	PRAUC	KS	GINI
대출정보 O	0.8304	0.2486	50.3447	0.6548
대출정보 X	0.7666	0.1230	41.3202	0.5459

Undersampling 1:9 (Test) – 개별 모형에서 예측				
	AUC	PRAUC	KS	GINI
대출정보 O	0.8291	0.2491	50.2979	0.6522
대출정보 X	0.7891	0.1239	42.9453	0.5718

III-(2). 발전방향

(1) 대출정보가 없는 차주에 의해 변별력이 낮아지는 현상

- 비금융 정보 활용
- 최종모형에서 연체관련 변수의 중요도가 높게 나온 것을 고려해볼 때 통신요금, 공공요금 등의 연체기록을 이용하여 파생변수를 생성하면 모형의 변별력을 높일 수 있을 것으로 기대됨

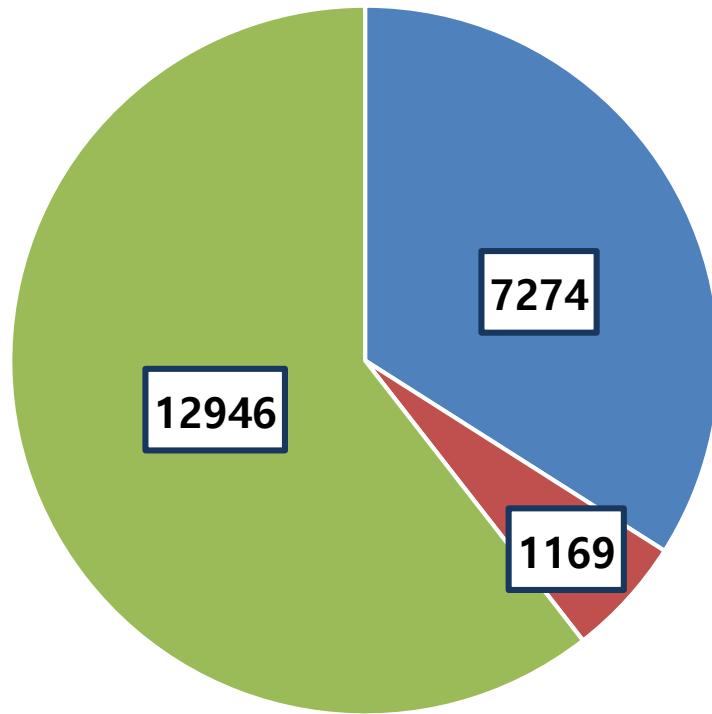
정보명	활용기간	평가내용	등록방법
소득금액증명	등록일로부터 18개월	신용여력	증명서 발급 후 등록
건강보험	등록일로부터 18개월	신용여력, 신용성향(약속이행)	증명서 발급 후 등록
국민연금	등록일로부터 18개월	신용여력, 신용성향(약속이행)	증명서 발급 후 등록
통신요금	등록일로부터 6개월	신용성향 (약속이행)	납부확인서 발급 후 등록
공공요금	등록일로부터 6개월	신용성향 (약속이행)	납부확인서 발급 후 등록
납부내역증명 (납세사실증명)	등록일로부터 18개월	신용성향 (약속이행)	증명서 발급 후 등록

[출처] 올크레딧

III-(2). 발전방향

(2) 제 3 연체자(대출/카드 연체 기록이 없는 불량 차주)

불량차주 구성비



- 연체관련 정보는 한국신용정보원, 개인 신용정보사의 채무불이행 정보, 기타 연체 공공기록정보와 금융질서문란 정보 등을 포함

[출처] 올크레딧

■ 대출계좌O+연체
■ 제 3연체자

■ 대출계좌X+카드연체

감사합니다