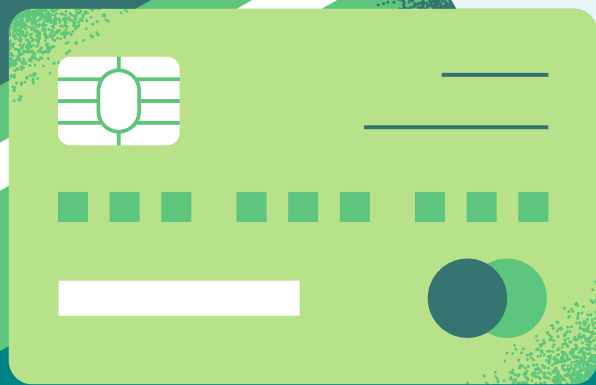


# 개인 고객 신용평가모형 모델링

중신용자 대출승인을 위한 잠재기준 제안  
(그레이존 공략 전략)

[1조: 다이기조]

구영지, 김재현, 김정민, 이한비, 임채영



A hand with black nail polish is pointing at a calendar. The calendar shows months from February to June. A US dollar bill is also visible, partially overlapping the calendar. The background is a light-colored wooden surface.

# 목차

01. 프로젝트 개요

02. 탐색적 데이터 분석(EDA)

---

03. 프로젝트 수행 절차 및 방법

04. 프로젝트 수행 결과

05. 자체 평가 의견

---

# 1. 프로젝트 개요

## 1.1 프로젝트 주제 및 선정 배경

 법률신문

### 금융위원회, 「AI 기반 신용평가모형 검증체계」 및 「금융분야 AI 보안 가이드라인」 마련

금융위원회는 2023년 4월 18일 신뢰받는 AI 활용 환경을 구축하기 위한 「AI 기반 신용평가모형 검증체계」와 「금융분야 AI 보안 가이드라인」을...

2023. 5. 2.

### 추진배경

AI 신용평가모형은 다양한 비금융·비정형 데이터를 평가항목으로 반영할 수 있어 **예측력·변별력이 우수**하다는 장점이 있으나, 복잡한 알고리즘을 사용하기 때문에 평가결과에 대한 **직관적인 해석과 설명이 다소 어렵다**는 한계도 존재합니다.

이에 금융기관은 AI신용평가모델에 XAI(설명가능한 AI·eXplainable AI)' 기능을 더해 앞으로도 유의미한지 결과를 신뢰할 수 있도록 하기 위해 노력하고 있습니다.



신용평가모형에서 설명가능한 요소 파악을 위한 **데이터 분석**의 필요성 존재

## 1.2 프로젝트 개요

### ● 프로젝트 구현 내용

- 정교한 신용평가모형으로  
개인 고객의 신용등급 분류
- 그레이 영역(Gray Area)  
대출 신청자의  
대출 승인 기준 명확화

### ● 훈련내용과의 관련성

- 교육 내용  
가설검정, 머신러닝을  
활용한 데이터 분석
- 하나은행 현직자 특강  
대출 규모 감소 현황 파악

### ● 활용 장비 및 재료

- 코드관리: github
- 일정관리: jira
- 데이터분석: python  
(pandas, sklearn 등)



## 1.3 프로젝트 구조

프로젝트 재정의 및 리서치  
결측치, 이상치에 대한 데이터 정제  
머신러닝 모델간 특징 비교

### 데이터 전처리

7/27~7/31

다양한 클러스터링 알고리즘과  
통계 분석을 통한 고객 그룹별 패턴 파악

### 클러스터링 및 통계분석

7/27~7/31

8/3~8/6  
인사이트 도출

클러스터링 정보를 바탕으로  
그레이존 고객 대출 심사 잠재기준 제안

프로젝트 기획 및 주제 선정  
프로젝트 정의서 작성  
탐색적 데이터 분석 실행

7/25~7/26  
사전 기획

데이터 전처리 및 모델링  
불균형 데이터를 해소하기 위한 SMOTE등의 모델링 진행  
잠재적 머신러닝 모델 선정

7/31~8/2  
모델링



# 1.3 프로젝트 구조

프로젝트 재정의 및 리서치  
결측치, 이상치에 대한 데이터 정제  
머신러닝 모델간 특징 비교

데이터 전처리

7/27~7/31

다양한 클러스터링 알고리즘과  
통계 분석을 통한 고객 그룹별 패턴 파악

클러스터링 및 통계분석

7/27~7/31

총 프로젝트 수행기간: 7/25 ~ 8/7 (약 2주)

7/25~7/26

사전 기획

프로젝트 기획 및 주제 선정  
프로젝트 정의서 작성  
탐색적 데이터 분석 실행

7/31~8/2

모델링

데이터 전처리 및 모델링  
불균형 데이터를 해소하기 위한 SMOTE등의 모델링 진행  
잠재적 머신러닝 모델 선정

8/3~8/6

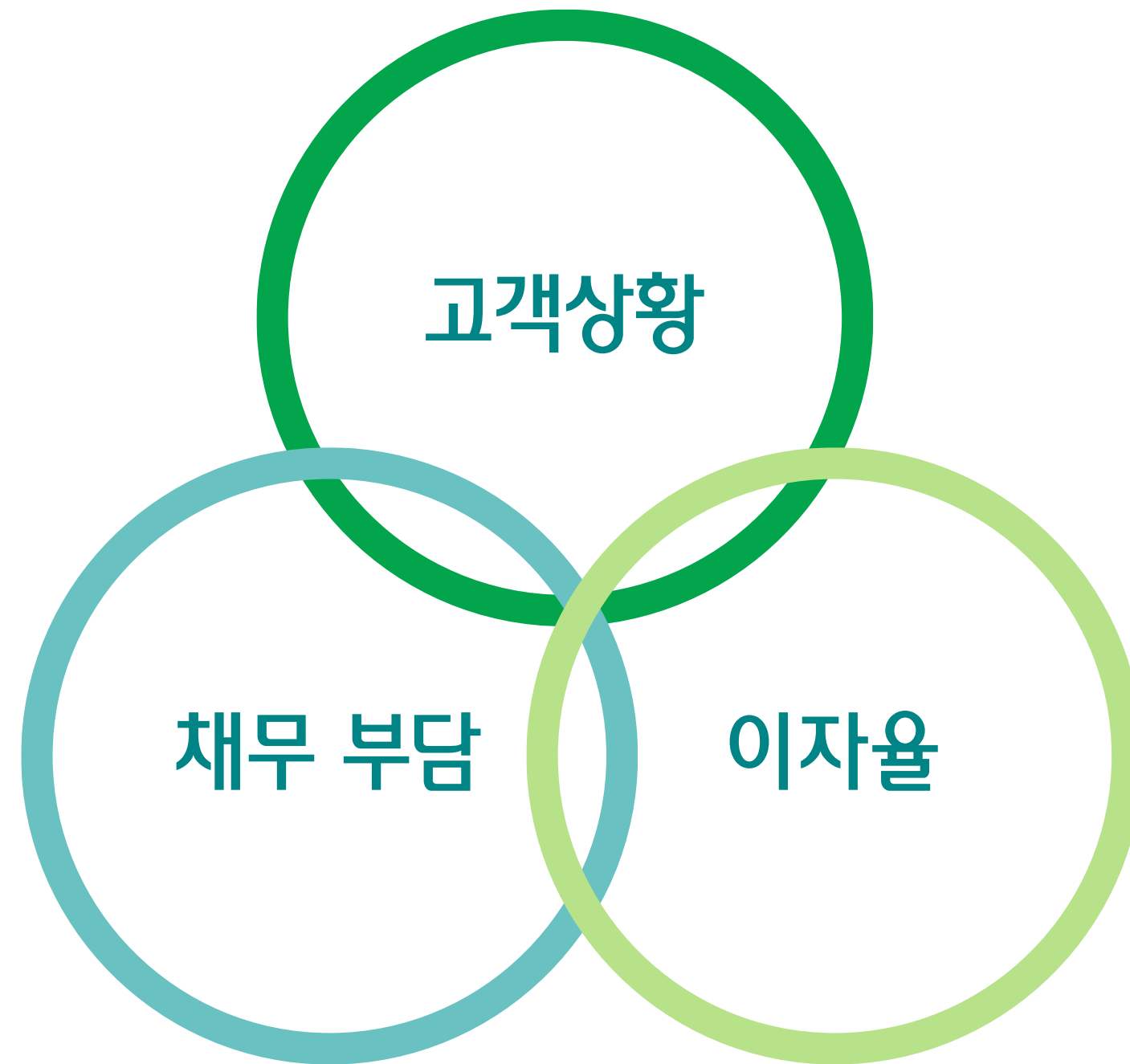
인사이트 도출

클러스터링 정보를 바탕으로  
그레이존 고객 대출 심사 잠재기준 제안

## 2. 탐색적 데이터 분석 (EDA)



## 2.1 탐색적 데이터 분석(EDA)



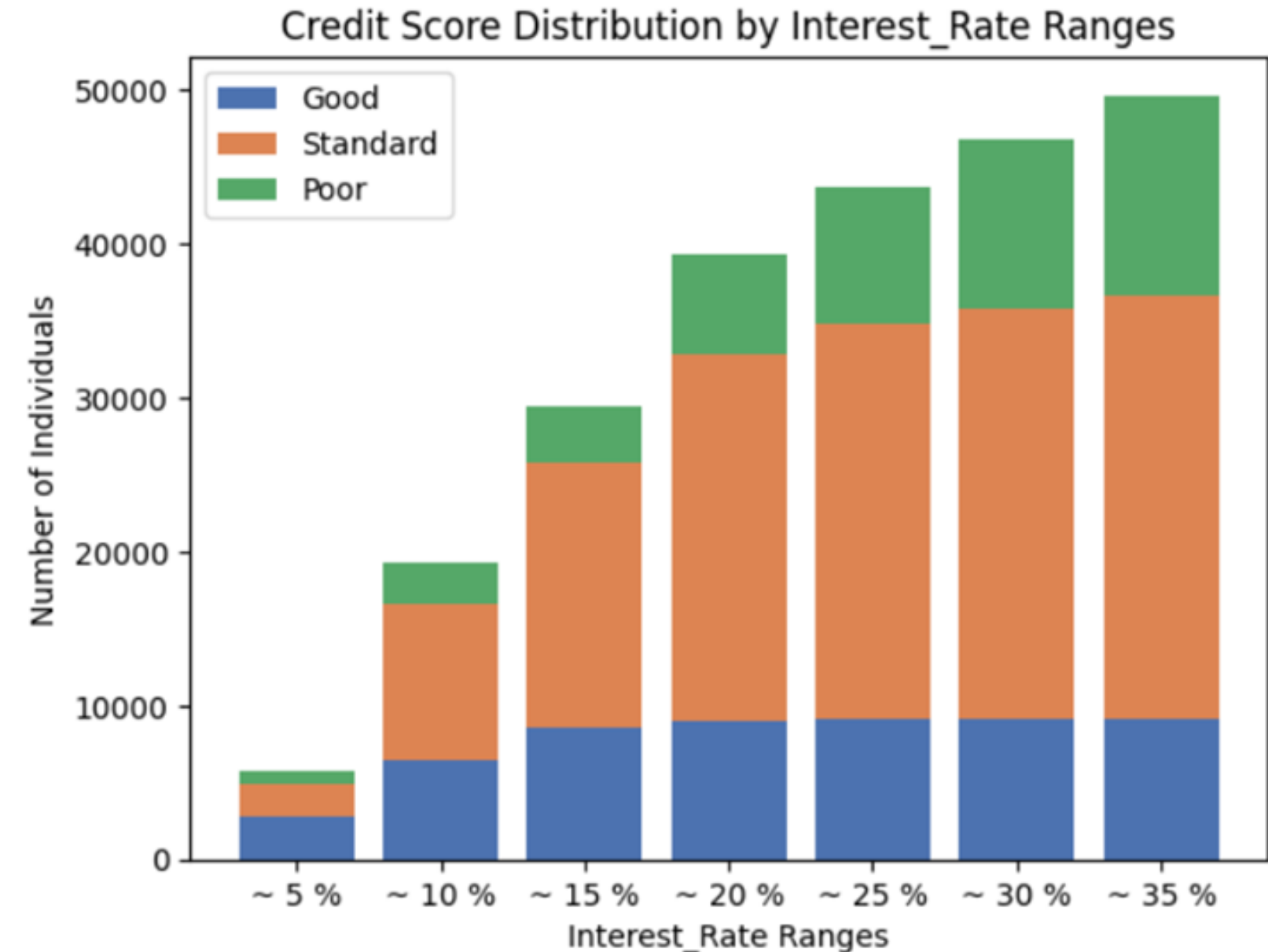
개인 대출 심사에 영향을 미치는 요인들을 기준으로 EDA를 진행

## 2.1 탐색적 데이터 분석(EDA) - target값 분석



신용등급 비율

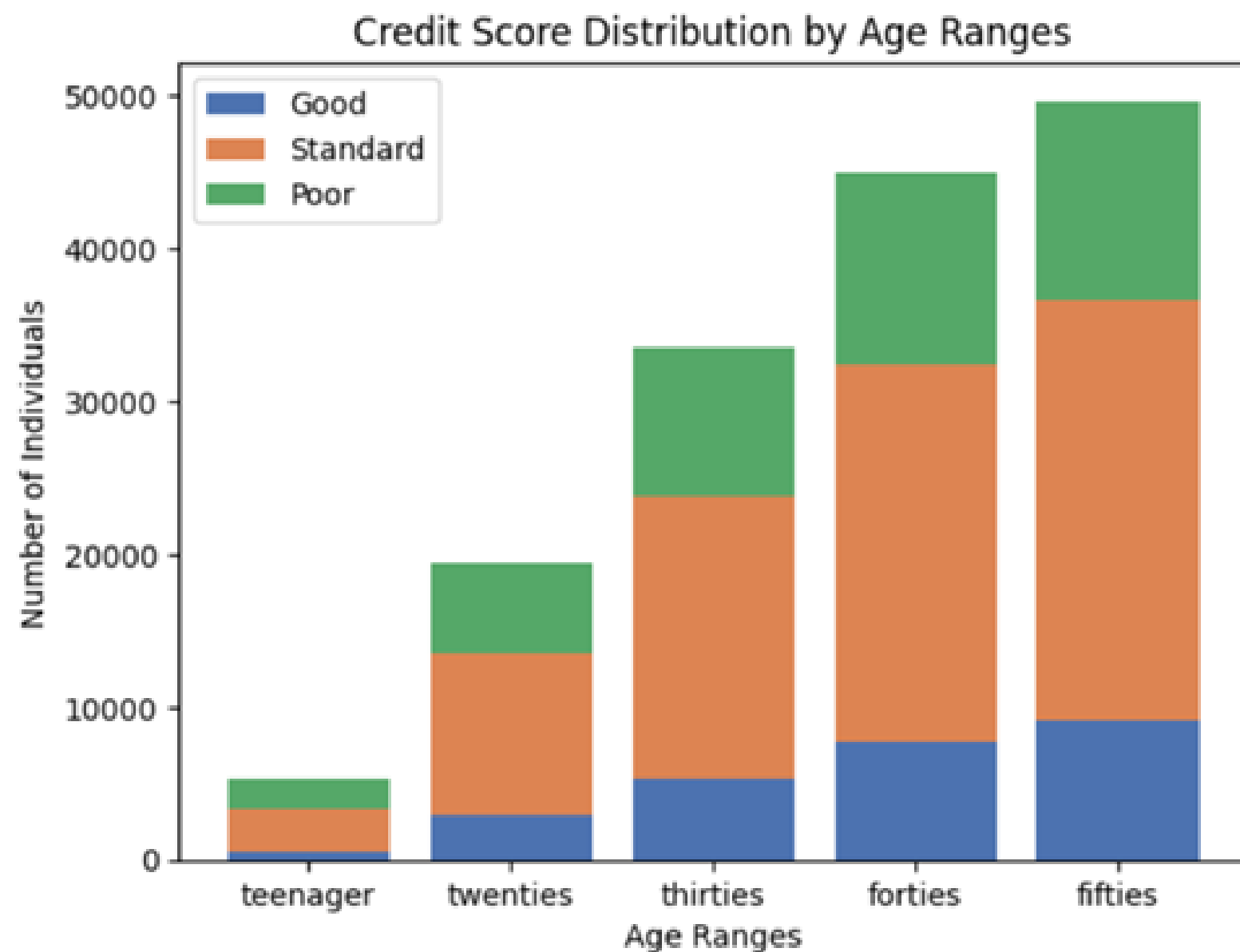
Target 값인 Credit Score의 비율을 확인한 결과  
2:5:3 으로 imbalanced data임을 파악



이자율과 신용등급

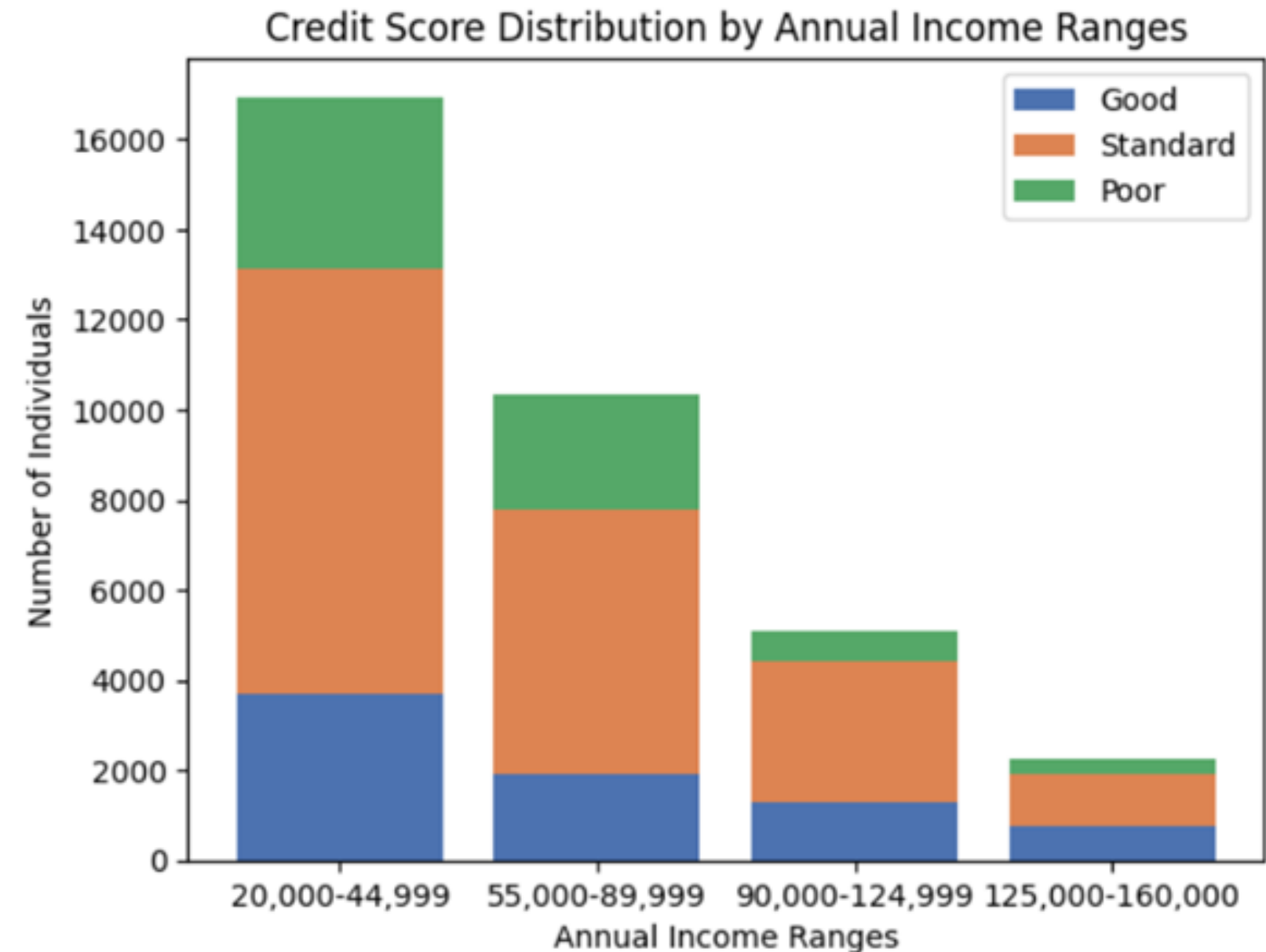
이자율이 클수록 저신용자 비율이 증가하고 있음

## 2.1 탐색적 데이터 분석(EDA) - target값 분석 / 고객 상황 = 고객 신상 정보



연령과 신용등급

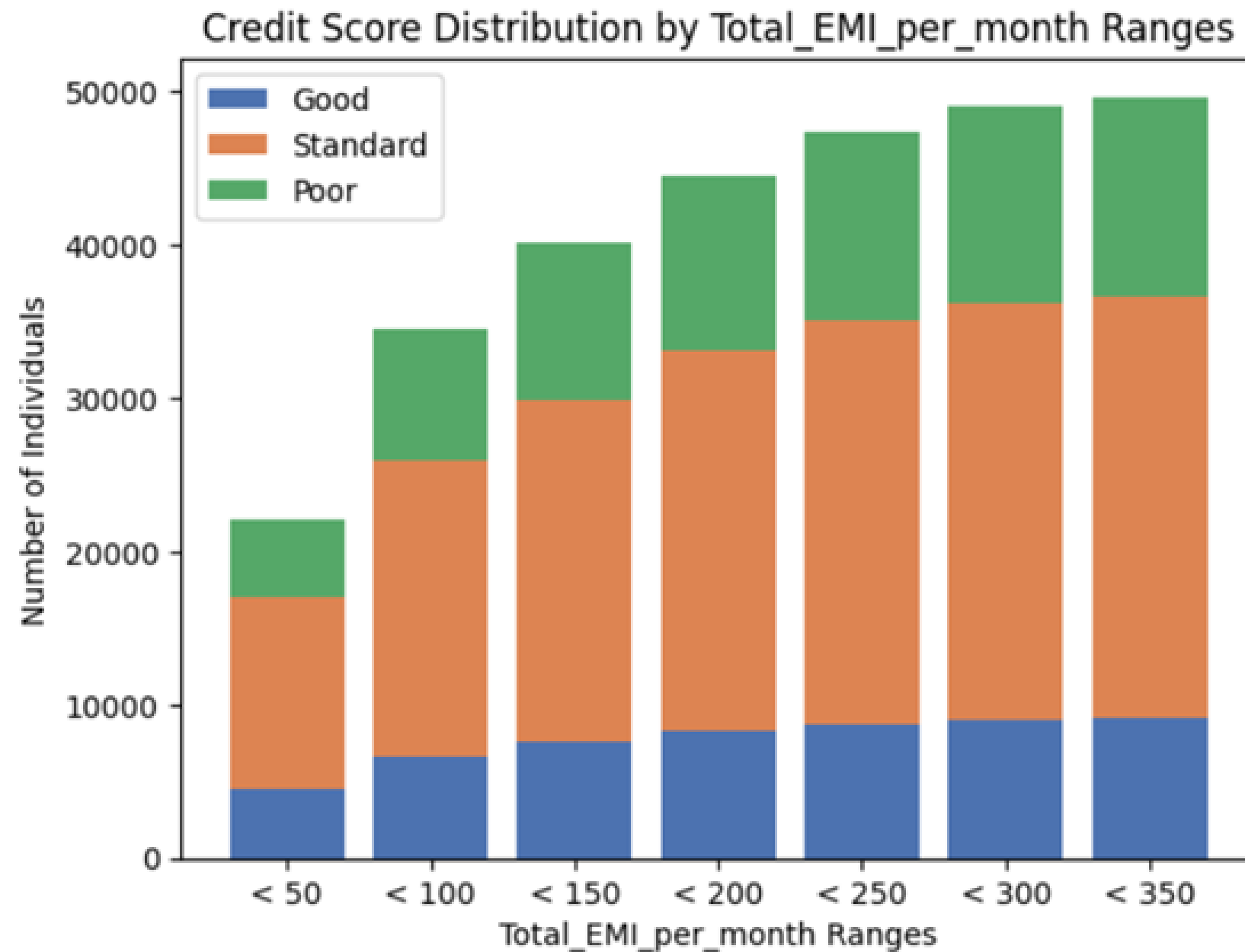
나이대별 신용등급 분포는 비슷하나,  
나이가 들수록 good과 poor 등급의 비율 ↑



연 수입과 신용등급

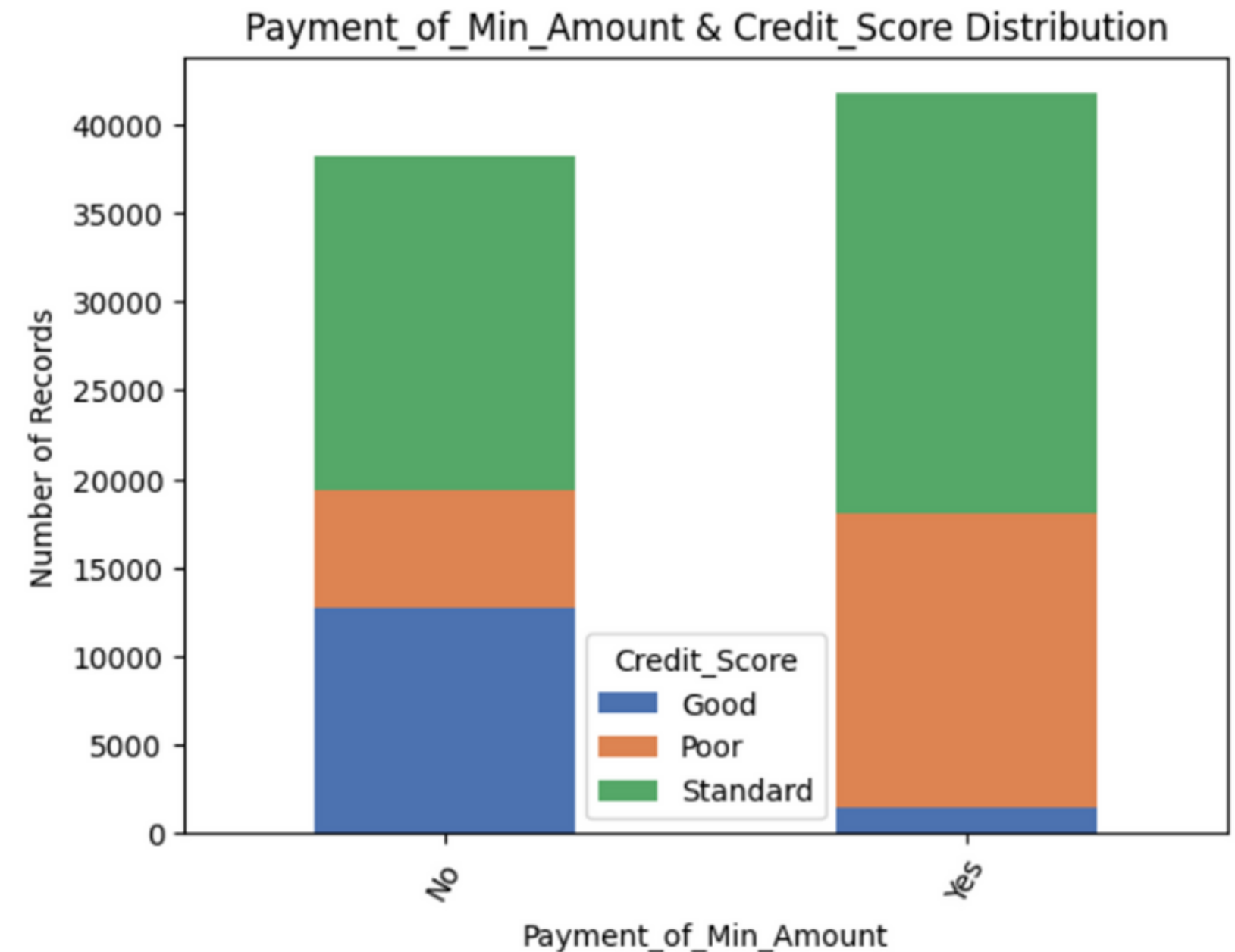
데이터 대다수가 낮은 수입  
대출 절벽에 몰릴 가능성이 높은 그레이존 고객多

## 2.1 탐색적 데이터 분석(EDA) - target값 분석 / 채무 부담 = 채무 상환 의지



월별 상환금액과 신용등급

월별 상환금액별 신용등급 분포는 거의 균일



리볼빙 여부와 신용등급

리볼빙 yes : poor 등급 비율 ↑, good 등급 비율 ↓  
=> 신용등급 관리가 어려운 사람들이 주로 사용

## 2.1 탐색적 데이터 분석(EDA)

### 요약

#### 이자율

- 개인고객의 **과거 신용등급**이 반영된 지표
- 이자율이 높을수록 중저신용자 다수 분포

#### 고객 상황

- 연령, 직업, 연 수입 등 고객의 **현재 경제적 지표**
- 대다수가 저소득자이며 각 분포는 균일

#### 채무 부담

- 월별 상환금 액수 자체는 신용등급이 균일하게 분포
- **리볼빙**을 사용 시에 **저신용자 다수 분포**



과거 신용등급이 반영된 **이자율** 뿐만 아니라 **채무 상환 의지**도 신용등급 산정에 큰 영향

-> 다양한 관점에서의 **상관관계 분석**, **클러스터링**, **통계 분석**을 통한 요인들 간 관계 확인 필요

-> 데이터 분석 결과를 토대로 추가적 데이터 인사이트 도출

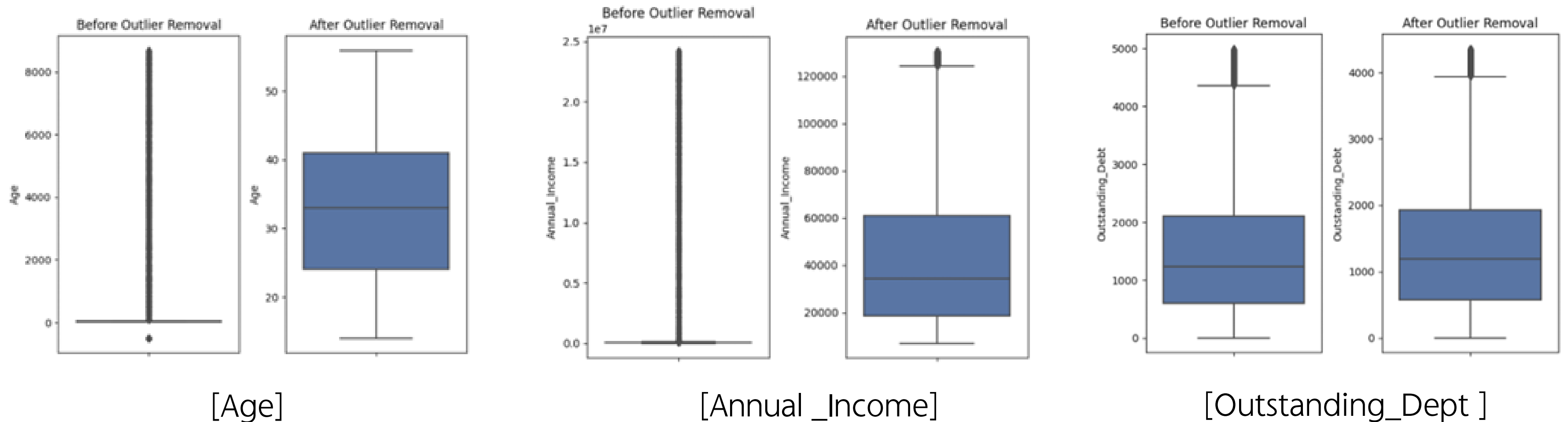
# **3 프로젝트 수행 절차 및 방법**

## **3.1 데이터 전처리**

## 3.1.1 이상치 처리

원본 데이터 : Age, Annual\_Income, Outstanding\_Dept 등

- 앞선 EDA를 통해 이상치가 의심되는 데이터(numerical values)를 확인함
- 데이터에 대한 이상치를 자세히 확인하기 위해 Boxplot EDA를 진행



이상치 처리 방법 => IQR 방식을 사용하여 이상치를 제거하는 방법 선택



### 3.1.2 결측치 처리

수치형 변수

customer\_ID로 그룹화 & 그룹별 중앙값으로 대체

Monthly_Inhand_Salary	
1824.843333	7266.686667
	7266.686667
	7266.686667
1824.843333	7266.686667
	4766.68
1824.843333	4766.68
1824.843333	
3037.986667	4766.68
3037.986667	4766.68

예시) Monthly\_Inhand\_Salary

-결측치 개수: 11303개

-결측치 처리: customer\_id로 그룹화 → 중앙값 대체

범주형 변수

customer\_ID로 그룹화 & 그룹별 최빈값으로 대체  
최빈값이 없는 경우 중앙값으로 대체

Payment_Behaviour	
High_spent_Small_value_payments	High_spent_Large_value_payments
Low_spent_Large_value_payments	!@9#%8
Low_spent_Medium_value_payments	Low_spent_Medium_value_payments
Low_spent_Small_value_payments	High_spent_Large_value_payments
High_spent_Medium_value_payments	Low_spent_Medium_value_payments
!@9#%8	!@9#%8
Low_spent_Small_value_payments	High_spent_Small_value_payments
High_spent_Medium_value_payments	High_spent_Large_value_payments
Low_spent_Small_value_payments	High_spent_Medium_value_payments
High_spent_Large_value_payments	Low_spent_Medium_value_payments
High_spent_Large_value_payments	Low_spent_Small_value_payments
	Low_spent_Small_value_payments

예시) Payment\_Behaviour

- “!@9#%8” : NaN 값으로 변환 → 결측치 제거
- 결측치 처리: customer\_id로 그룹화 → 최빈값 대체



'\_', '#F%\$D@\*&8' 등의 의미없는 데이터 처리 방법  
모두 NaN값으로 대체 후 결측치 처리 방법 적용

### 3.1.3 그 외 처리방법

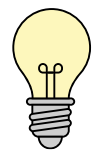
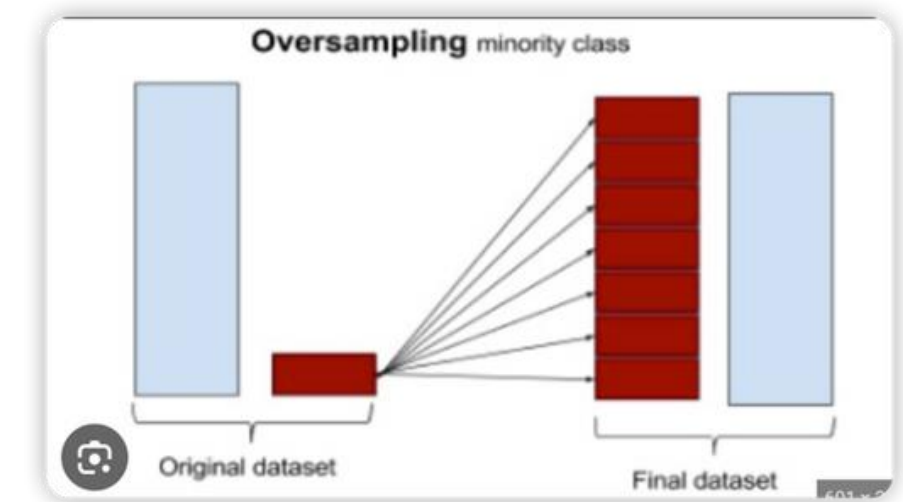
#### 💡 인코딩 (범주형 변수만)

Occupation, Credit\_Mix, Payment\_of\_Min\_Amount, Payment\_Behaviour, Credit\_Score

Species (범주형 변수)		인코더 값
Good	라벨 인코딩 ➡	0
Poor		1
Standard		2

#### 💡 불균형 데이터샘플링 : SMOTE 기법

- target 데이터 비율이 불균형한 경우, 오버샘플링을 수행하여 불균형 해소
- 가장 많은 비중을 차지하던 Standard 클래스에 맞춰져 총 데이터의 수 증가
- 오버샘플링 적용 전 후 성능에 차이 X



데이터를 왜곡시키는 **오버샘플링 기법은 적용하지 않고** 모델을 학습

# **3. 프로젝트 수행 절차 및 방법**

## **3.2 분류 모델링**

## 3.2.1 머신러닝 모델링

*[What is Pycaret?]*

	Description	Value
0	Session id	123
1	Target	Credit_Score
2	Target type	Multiclass
3	Original data shape	(50503, 23)
4	Transformed data shape	(50503, 23)
5	Transformed train set shape	(35352, 23)
6	Transformed test set shape	(15151, 23)
7	Numeric features	22
8	Preprocess	True
9	Imputation type	simple
10	Numeric imputation	mean
11	Categorical imputation	mode
12	Fold Generator	StratifiedKFold
13	Fold Number	10
14	CPU Jobs	-1
15	Use GPU	False
16	Log Experiment	False
17	Experiment Name	clf-default-name
18	USI	5d24

- AutoML 적용해주는 라이브러리
- XGB, LightBGM, ExtraTree 등의 알고리즘 성능을 비교해서 최적 모델을 선택하고 최적화하는 과정을 효율적으로 처리

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
et	Extra Trees Classifier	0.7996	0.9041	0.7996	0.7993	0.7993	0.6591	0.6593	4.4480
rf	Random Forest Classifier	0.7922	0.9011	0.7922	0.7924	0.7922	0.6482	0.6483	4.6110
lightgbm	Light Gradient Boosting Machine	0.7501	0.8778	0.7501	0.7531	0.7508	0.5790	0.5794	1.3680
gbc	Gradient Boosting Classifier	0.7279	0.8576	0.7279	0.7344	0.7295	0.5441	0.5451	19.7330
dt	Decision Tree Classifier	0.6936	0.7352	0.6936	0.6937	0.6936	0.4816	0.4816	0.6350
qda	Quadratic Discriminant Analysis	0.6770	0.8009	0.6770	0.7152	0.6832	0.4884	0.5000	0.4600
ada	Ada Boost Classifier	0.6712	0.7311	0.6712	0.6751	0.6696	0.4353	0.4374	1.6560
lda	Linear Discriminant Analysis	0.6645	0.7931	0.6645	0.6671	0.6605	0.4178	0.4213	0.4550
knn	K Neighbors Classifier	0.6502	0.7777	0.6502	0.6484	0.6492	0.4041	0.4042	1.0180
ridge	Ridge Classifier	0.6370	0.0000	0.6370	0.6349	0.6201	0.3320	0.3456	0.4120
nb	Naive Bayes	0.6337	0.7617	0.6337	0.6911	0.6429	0.4342	0.4514	0.3890
lr	Logistic Regression	0.6051	0.7149	0.6051	0.5831	0.5576	0.2337	0.2597	4.7040
dummy	Dummy Classifier	0.5534	0.5000	0.5534	0.3063	0.3943	0.0000	0.0000	0.3850
svm	SVM - Linear Kernel	0.5248	0.0000	0.5248	0.4858	0.4394	0.1168	0.1578	1.7270

Extra Tree

## 3.2.1 머신러닝 모델링

### *[What is **Extra Trees**?]*

- Random Forest 보다 좀 더 극단적으로 random하게 만든 모델
- 3배 정도 연산이 빠르고 bias와 variance를 낮출 수 있음

### *[Why **Extra Tree Model**?]*

- Random Forest 모델처럼 최적의 분할을 계산하지 않고 무작위로 분할하기 때문에 속도가 빠름
- overfitting 문제를 회피하여 모델 정확도를 향상시킴



( 튜닝 여부, 오버샘플링 여부 등 경우의 수를 고려  
여러 개의 Extra Tree 모델의 성능을 비교 )



**튜닝과 오버샘플링을 하지 않은 Extra Tree 모델**이 성능이 가장 좋게 나와, 최종 모델로 설정

## 3.2.1 머신러닝 모델링

### *[Data Split]*

전처리를 완료한 train.csv 데이터를 분할하여 모델 훈련 및 성능평가에 사용

### *[Train validation split]*



+

5 fold 교차검증 수행



분류 성능 평가 결과, **정확도는 0.8154(약 0.82)**



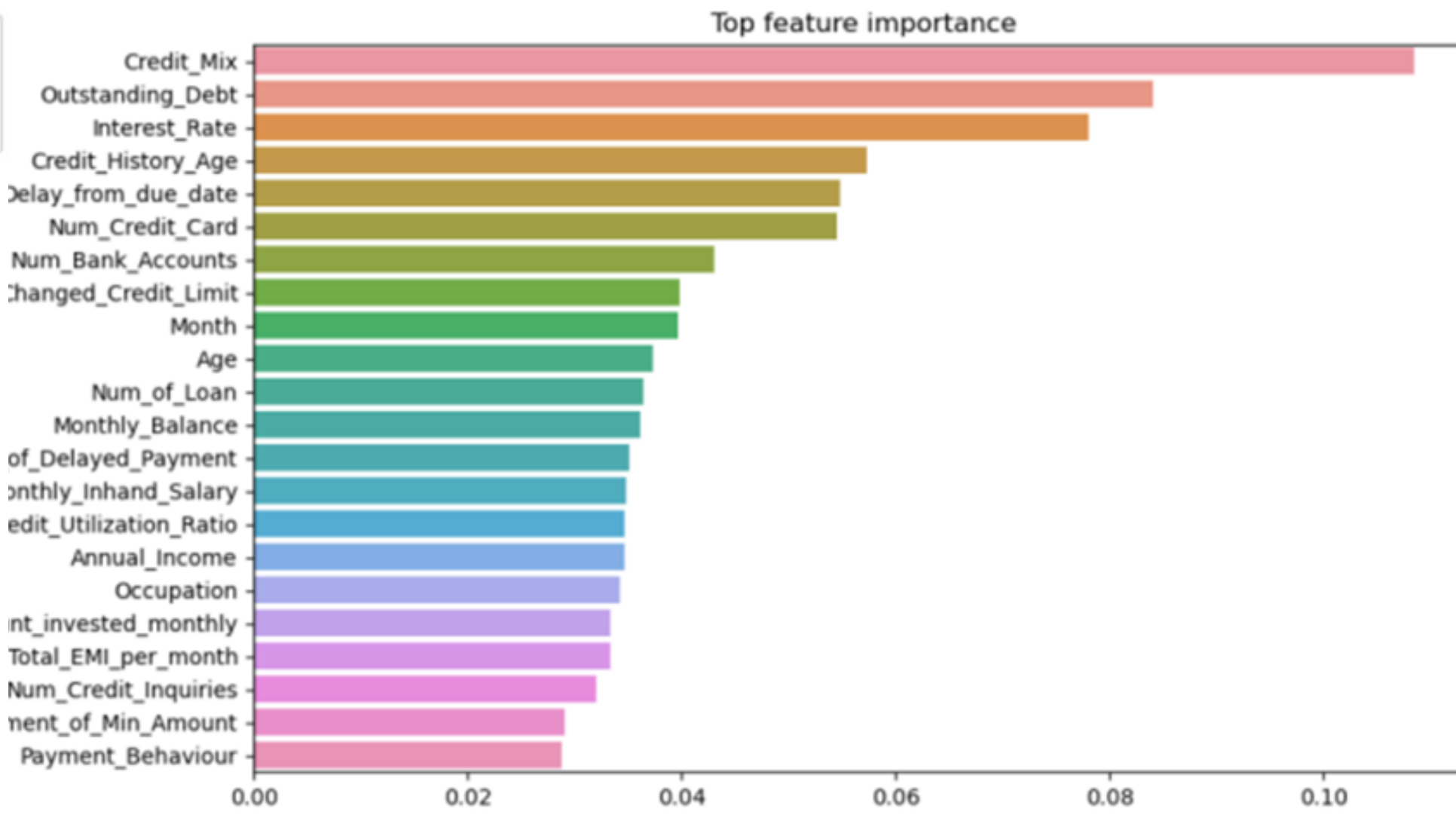
# 3.2.2 모델 시각화

```
from sklearn.metrics import classification_report
print(classification_report(y_train_sub, y_train_sub_pred))
print(classification_report(y_val, y_val_pred))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	7445
1	1.00	1.00	1.00	10597
2	1.00	1.00	1.00	22360
accuracy			1.00	40402
macro avg	1.00	1.00	1.00	40402
weighted avg	1.00	1.00	1.00	40402

	precision	recall	f1-score	support
0	0.77	0.76	0.77	1861
1	0.81	0.80	0.80	2650
2	0.83	0.84	0.84	5590
accuracy			0.82	10101
macro avg	0.80	0.80	0.80	10101
weighted avg	0.82	0.82	0.82	10101



Feature Importance를 확인한 결과,  
**credit mix, outstanding debt, interest rate** 요소가 높게 다루어지는 것을 볼 수 있음  
일반적 상식과 달리 age, occupation 등  
개인의 현재 경제 지표를 나타내는 변수들의 중요도는 크게 두드러지는 특성을 보이지 않음



## 4. 프로젝트 수행 결과

### 추가 데이터 인사이트 도출

## 4.1 모델을 활용한 추가 데이터 분석

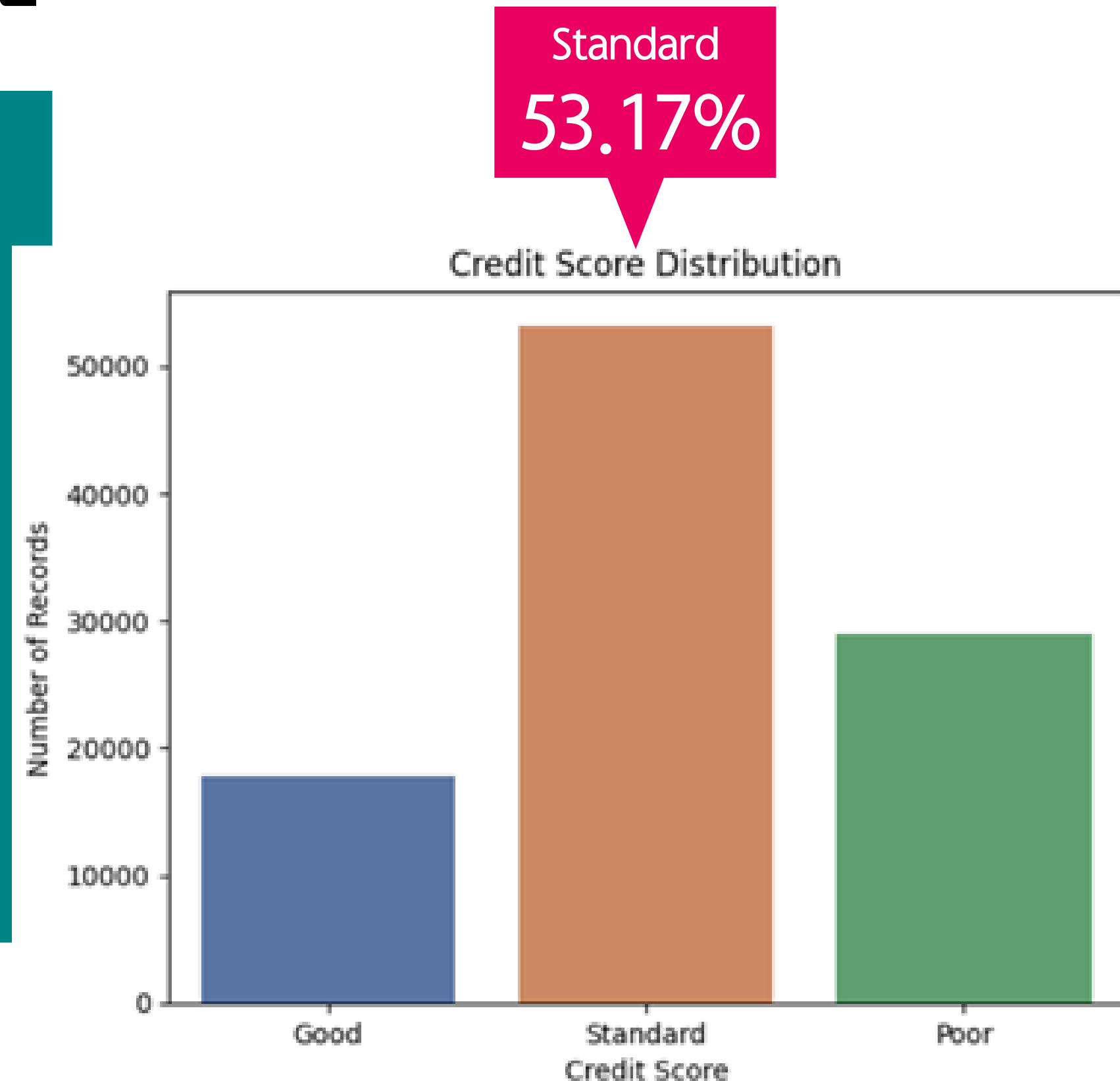
"그레이 영역 고객의 대출 승인을 위한 잠재 기준 제안"

그레이 영역이란?

어느 영역에 속하는지 불분명한 중간지대


-> 신용등급이 Good or Poor가 아닌  
애매한 **Standard** 등급의 고객들을 의미

-> 이들 중 대출 승인 가능성이 높은  
**잠재고객을 발굴(분류)**하는 군집화 모델 생성(Model2)



# 4.1 모델을 활용한 추가 데이터 분석 - 기획의도

주요 은행별 가계예대마진 추이 단위:%

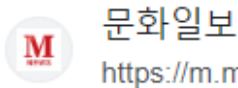
은행	구분	2022.07	2022.08	2022.09	2022.10	2022.11	2022.12
 하나은행	가계대출금리	4.12	4.33	4.76	5.06	5.28	5.35
	저축성수신금리	3.08	3.21	3.58	4.07	4.57	4.50
	가계예대금리차	1.04	1.12	1.18	0.99	0.71	0.85

대출규모의 감소로 최근 시중은행의  
주요 수익원인 예대마진율의 연속 감소

인터넷 뱅킹의  
공격적인 그레이존 고객 확보 전략

정확한 등급 산정으로  
채무불이행 리스크 최소화

신용등급 이외의 잠재 기준을 제시하여  
그레이존 고객확보 및 수익 확대

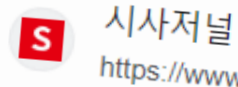


문화일보

<https://m.munhwa.com> > mnews > view

## 인증서 없어도 저신용자에 모바일 대출 ... 금융사가 놓친 ' ...

2020. 11. 25. — 자산관리, 개인 간(P2P) 금융 등 새로운 영역 개척도 본궐도다. 이승건 비바리퍼블리카 대표는 “금융의 슈퍼앱’을 만들겠다”고 선언했다. 토스는 토스 ...



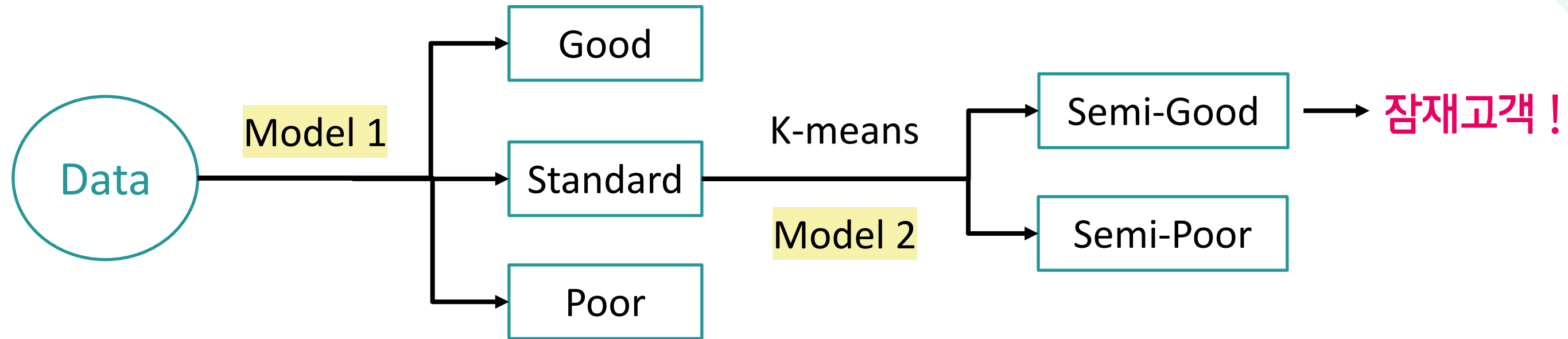
시사저널

<https://www.sisajournal.com> > news > articleView

## [카카오 제국] '그레이존' 순식간에 먹어치우며 금융 역사 바뀌 ...

2020. 6. 1. — 법령이나 질서가 새로운 경제 환경에 발맞추지 못해 영역 구분이 불분명해진 상황이나 집단 역시 그레이존이라 지칭한다. 이 그레이존을 해소하고 ...

## 4.2 중신용자 클러스터링



*[중신용자를  $k$ -means clustering 으로 다시 이진분류]*

- 신용등급 분류 모델(M1 모델)에서 중신용자로 분류된 고객 데이터를  $k$ -means clustering 하여 중신용자 중에서도 신용이 좋은 “중-고신용자”와 중신용자 중에서도 신용이 안 좋은 “중-저신용자”로 분류(M2 모델)

*[why  $k$ -means clustering?]*

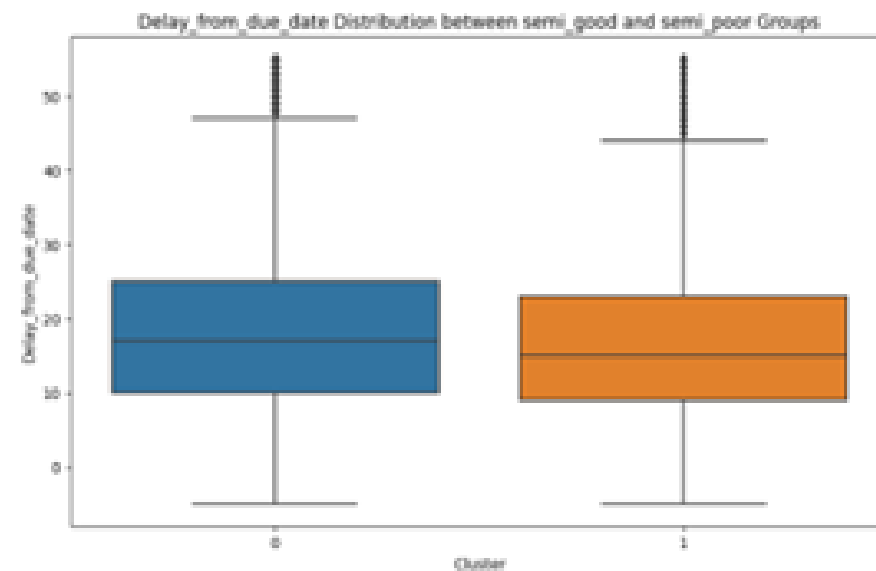
- 입력 변수와 계산이 복잡하지 않아서 대용량 데이터에도 적용 가능
- 사전에 특정 변수에 대한 역할 정의가 필요하지 않아서 적용이 쉬움
- 수렴성 보장

## 4.3 통계 분석

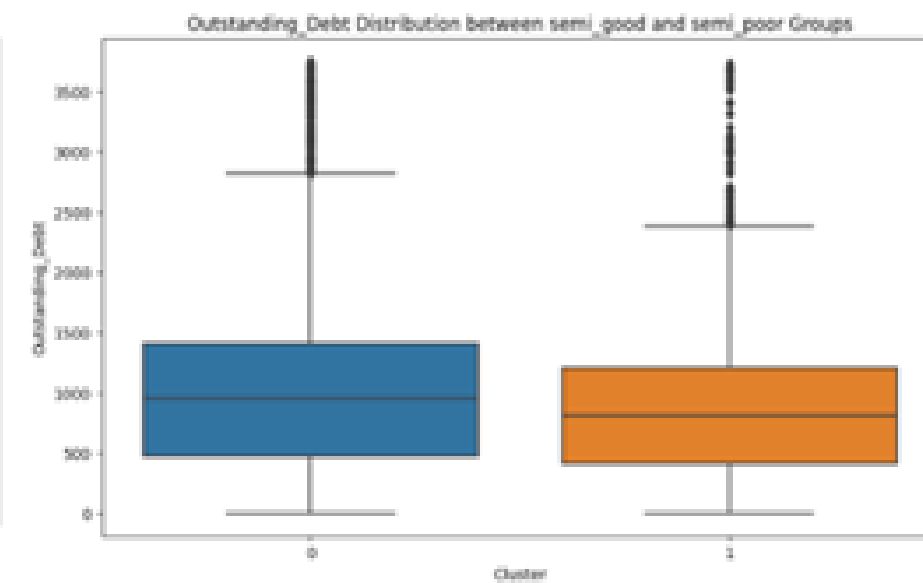
### [t-test]

Delay\_from\_due\_date, Outstanding\_Debt, Credit\_Utilization\_Ratio, Total\_EMI\_per\_month

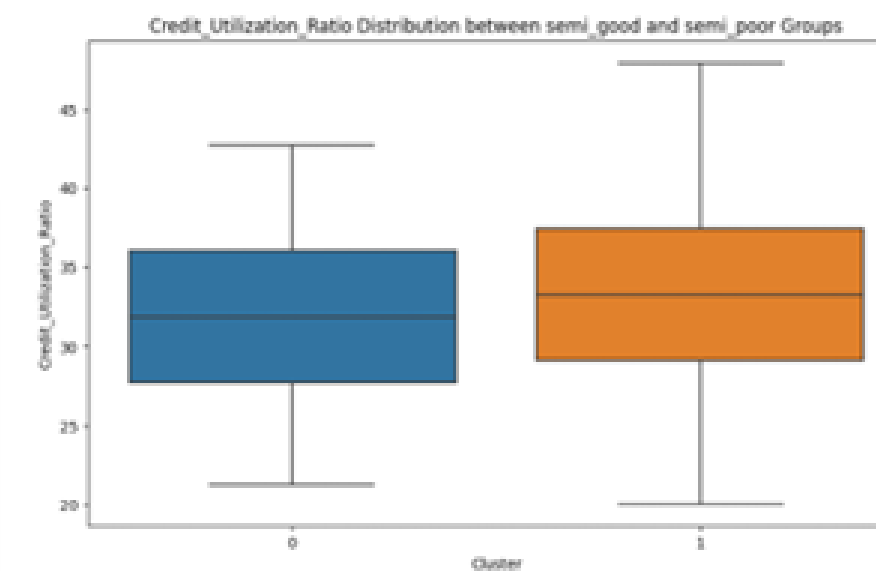
- T 검정 통계 분석 시 귀무가설을 기각하므로 해당 변수들은 모두 유의미하다
- 즉 중신용자를 중-고신용자와 중-저신용자로 분류하는데 유의미한 의미가 있다



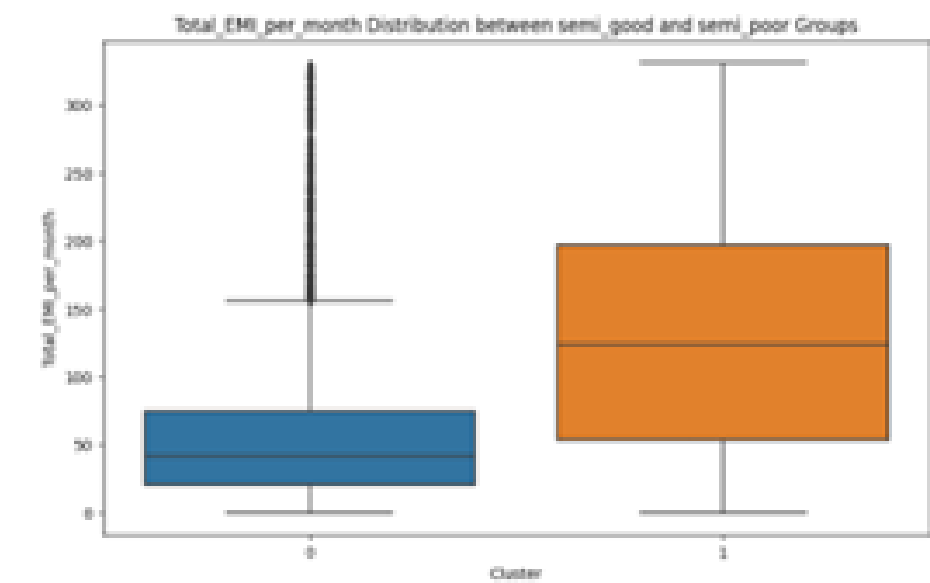
Variable: Delay\_from\_due\_date  
T-statistic: -18.9079  
P-value: 0.00000  
Result: Reject the null hypothesis (significant difference)



Variable: Outstanding\_Debt  
T-statistic: -25.1799  
P-value: 0.00000  
Result: Reject the null hypothesis (significant difference)



Variable: Credit\_Utilization\_Ratio  
T-statistic: 22.8938  
P-value: 0.00000  
Result: Reject the null hypothesis (significant difference)



Variable: Total\_EMI\_per\_month  
T-statistic: 82.6449  
P-value: 0.00000  
Result: Reject the null hypothesis (significant difference)

[Delay\_from\_due\_date]

[Outstanding\_Debt]

[Credit\_Utilization\_Ratio]

[Total\_EMI\_per\_month]

# 4.3 통계 분석

## 수치형 변수

[chi-square statistic]

## 범주형 변수

	Variable	Chi-square statistic	P-value	Result
0	Month	6.470643	4.859897e-01	H1 Reject
1	Occupation	31.808639	4.264588e-03	HO Reject
2	Credit_Mix	815.555269	8.024600e-178	HO Reject
3	Payment_of_Min_Amount	307.178060	8.994444e-69	HO Reject
4	Payment_Behaviour	2678.952086	0.000000e+00	HO Reject

	Variable	F-statistic	P-value	Result
0	Annual_Income	84025.590715	0.000000e+00	HO Reject
1	Monthly_Inhand_Salary	82864.469767	0.000000e+00	HO Reject
2	Num_Bank_Accounts	309.894462	5.431884e-69	HO Reject
3	Num_Credit_Card	196.144344	2.048999e-44	HO Reject
4	Interest_Rate	490.184668	1.096689e-107	HO Reject
5	Num_of_Loan	910.711002	6.669474e-197	HO Reject
6	Delay_from_due_date	360.539348	6.848673e-80	HO Reject
7	Num_of_Delayed_Payment	3.128191	7.695998e-02	H1 Reject
8	Changed_Credit_Limit	253.744469	7.058641e-57	HO Reject
9	Num_Credit_Inquiries	1.454272	2.278539e-01	H1 Reject
10	Outstanding_Debt	632.813077	4.211789e-138	HO Reject
11	Credit_Utilization_Ratio	520.475104	3.687025e-114	HO Reject
12	Credit_History_Age	678.968952	6.558025e-148	HO Reject
13	Total_EMI_per_month	6863.152579	0.000000e+00	HO Reject
14	Amount_invested_monthly	66.262354	4.109951e-16	HO Reject
15	Monthly_Balance	16918.396144	0.000000e+00	HO Reject

💡 유의미하지 않은 변수 : Month , Num\_of\_Delayed\_Payment, Num\_Credit\_Inquiries

⇒ 즉, ‘중-고 신용자’와 ‘중-저 신용자’로 분류하는 데 중요할 것이라고 생각했던 변수인 Num\_of\_Delayed\_Payment 는 영향을 미치지 않는다.

⇒ 채무 상환 의지를 가장 잘 대표한다고 볼 수 있는 변수인 Payment\_of\_Min\_Amount는 ‘중-고 신용자’와 ‘중-저 신용자’로 분류하는데 유의미하다.



## 4.4 Model2 클러스터링 결과

Cluster	0	1	variable_importance
Month	10.497556	10.507726	0.005407
Age	34.119422	35.159521	0.523023
Occupation	6.893328	7.072688	0.091618
Annual_Income	28815.115452	94375.092416	32776.351602
Monthly_Inhand_Salary	2392.611345	7859.290216	2732.864941
Num_Bank_Accounts	5.236507	4.729924	0.254553
Num_Credit_Card	5.258075	4.963221	0.147123
Interest_Rate	12.890034	11.235909	0.830229
Num_of_Loan	3.040055	2.408487	0.316307
Delay_from_due_date	17.970251	15.417410	1.276827
Num_of_Delayed_Payment	33.296005	30.214146	1.548450
Changed_Credit_Limit	10.814080	9.549541	0.633116
Num_Credit_Inquiries	30.166277	34.030250	1.923321
Credit_Mix	1.640353	1.668335	0.013778
Outstanding_Debt	1059.784780	841.416594	108.998623
Credit_Utilization_Ratio	31.953088	33.363105	0.705878
Credit_History_Age	240.864110	268.773667	13.951504
Payment_of_Min_Amount	0.560030	0.442220	0.059061
Total_EMI_per_month	62.679140	133.673670	35.558303
Amount_invested_monthly	566.068171	768.639290	101.108712
Payment_Behaviour	2.967276	2.008705	0.479066
Monthly_Balance	330.623363	589.549786	129.482869

- '0' 컬럼

클러스터 0의 모든 변수에 대한 평균값

- '1' 컬럼

클러스터 1의 모든 변수에 대한 평균값

- 'variable\_importance' 컬럼

클러스터링 결과, 각 변수들의 중요도 수치.

숫자가 클수록 클러스터간 구분을 잘해준다는 것을 의미



## 4.5 잠재고객의 데이터 시각화

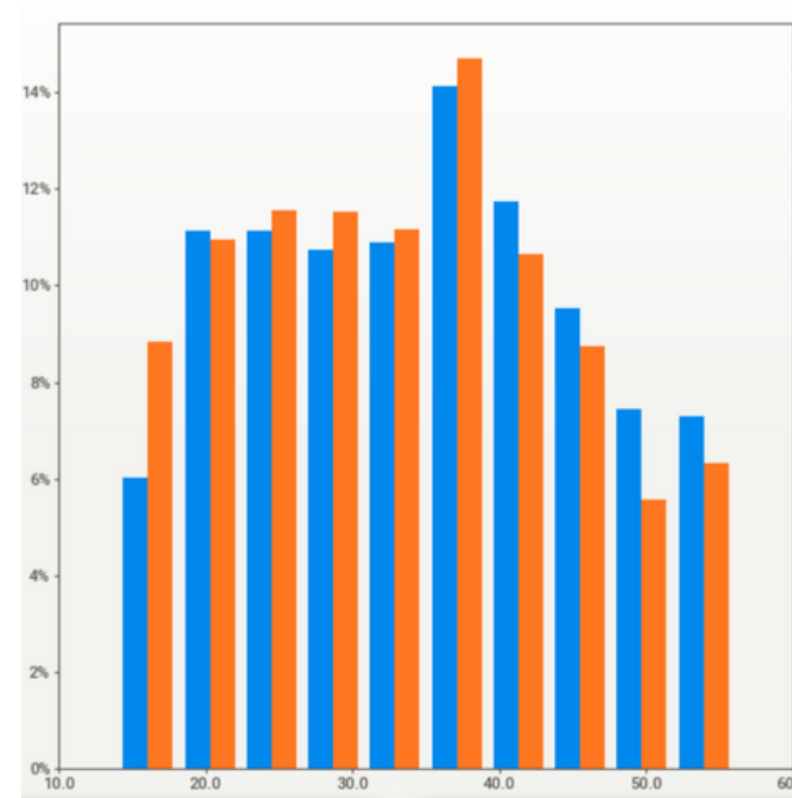
Semi_Good		Semi_Poor	
9559	ROWS	18391	
0	DUPLICATES	0	
1.8 MB	RAM	3.5 MB	
23	FEATURES	23	
6	CATEGORICAL	6	
17	NUMERICAL	17	
0	TEXT	0	

ASSOCIATIONS

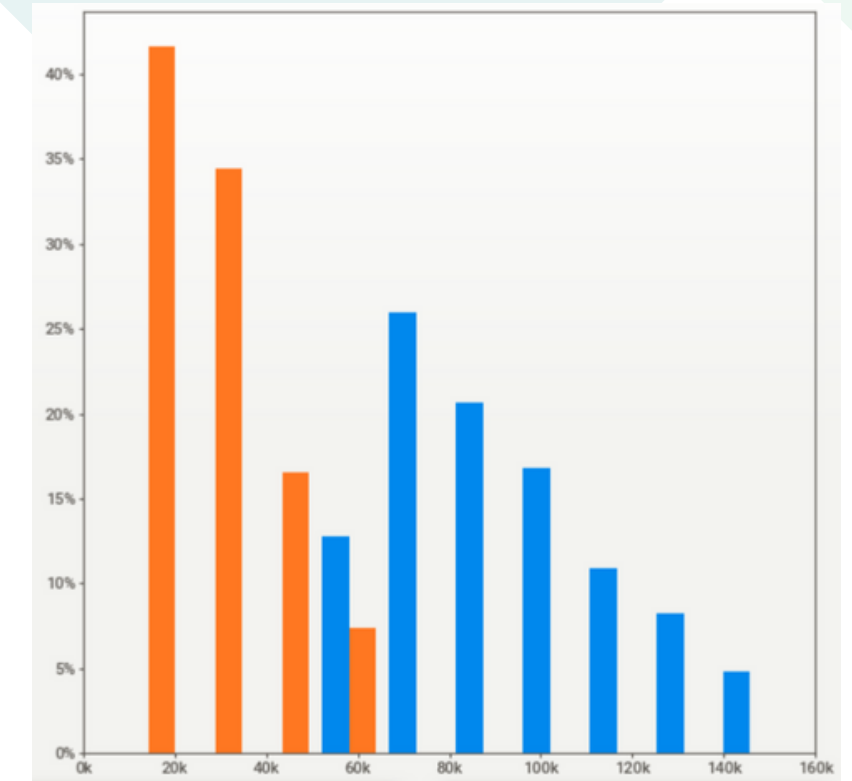
ASSOCIATIONS

Semi\_Good Semi\_Poor

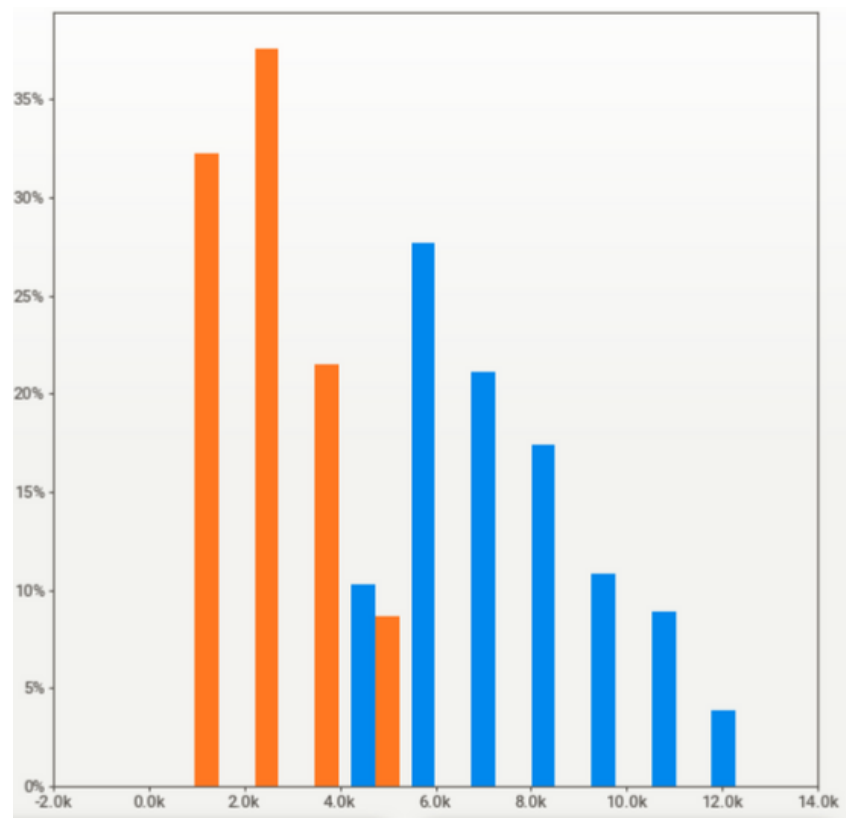
잠재고객: 9559명  
잠재고객 x: 18391명



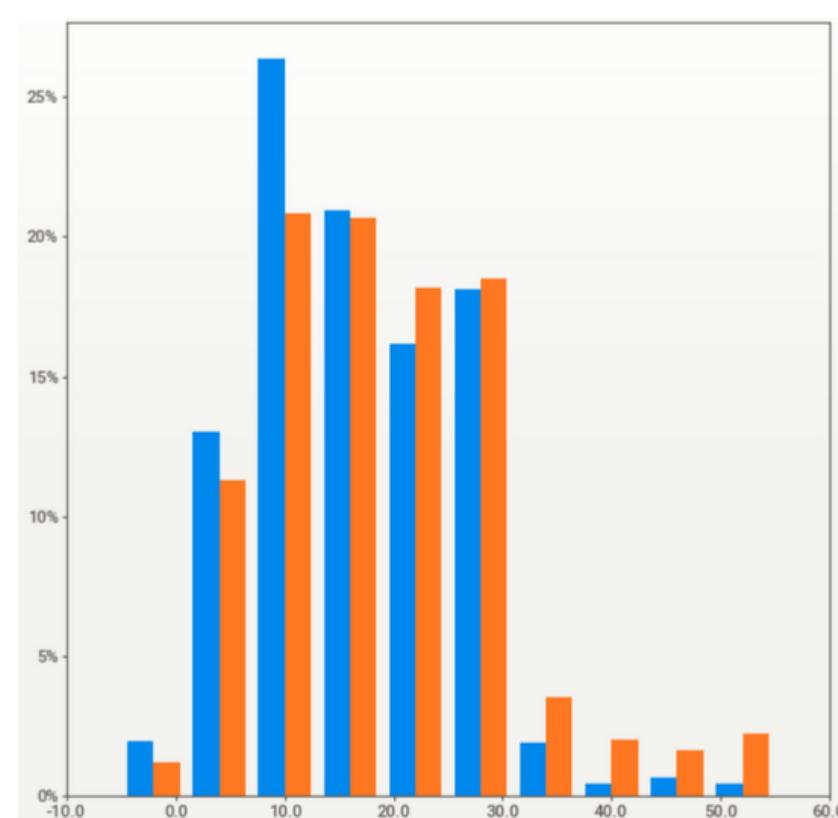
[Age]



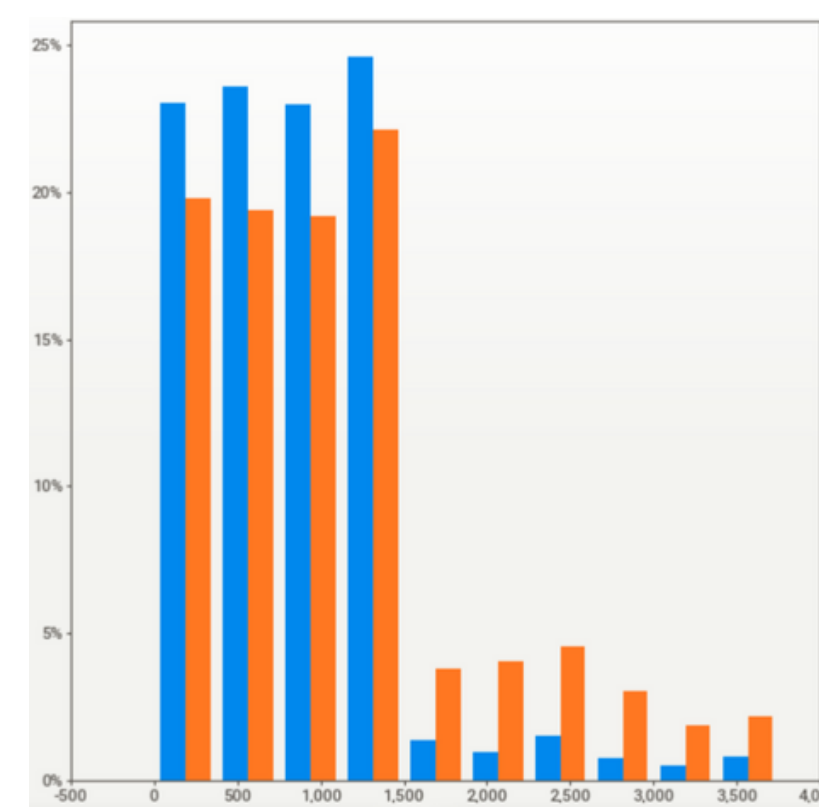
[Annual\_Income]



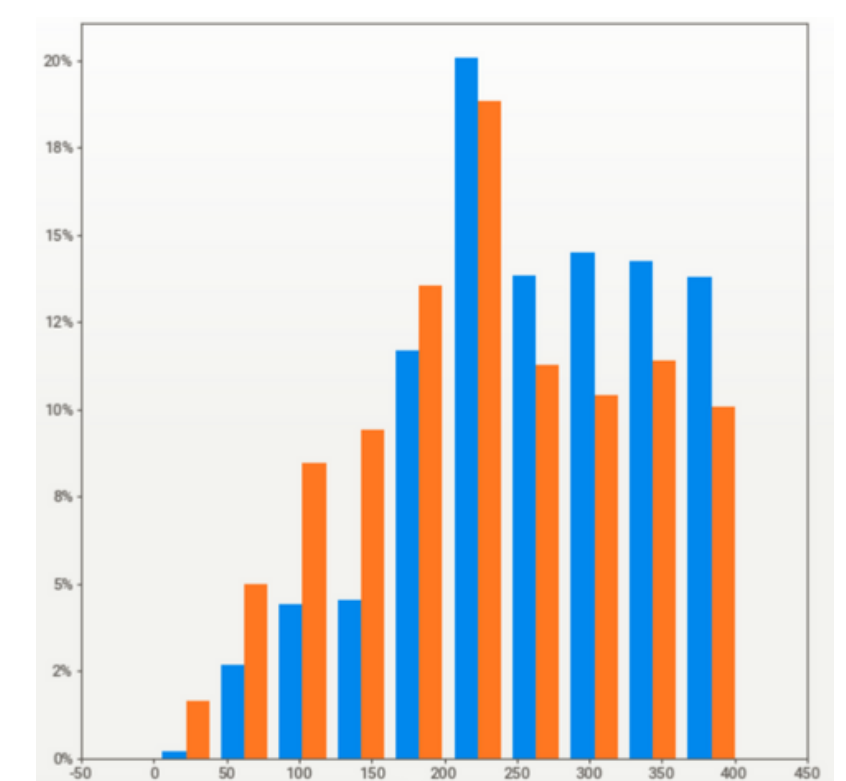
[Monthly\_Inhand\_Salary]



[Delay\_from\_due\_date]



[Outstanding\_Debt]



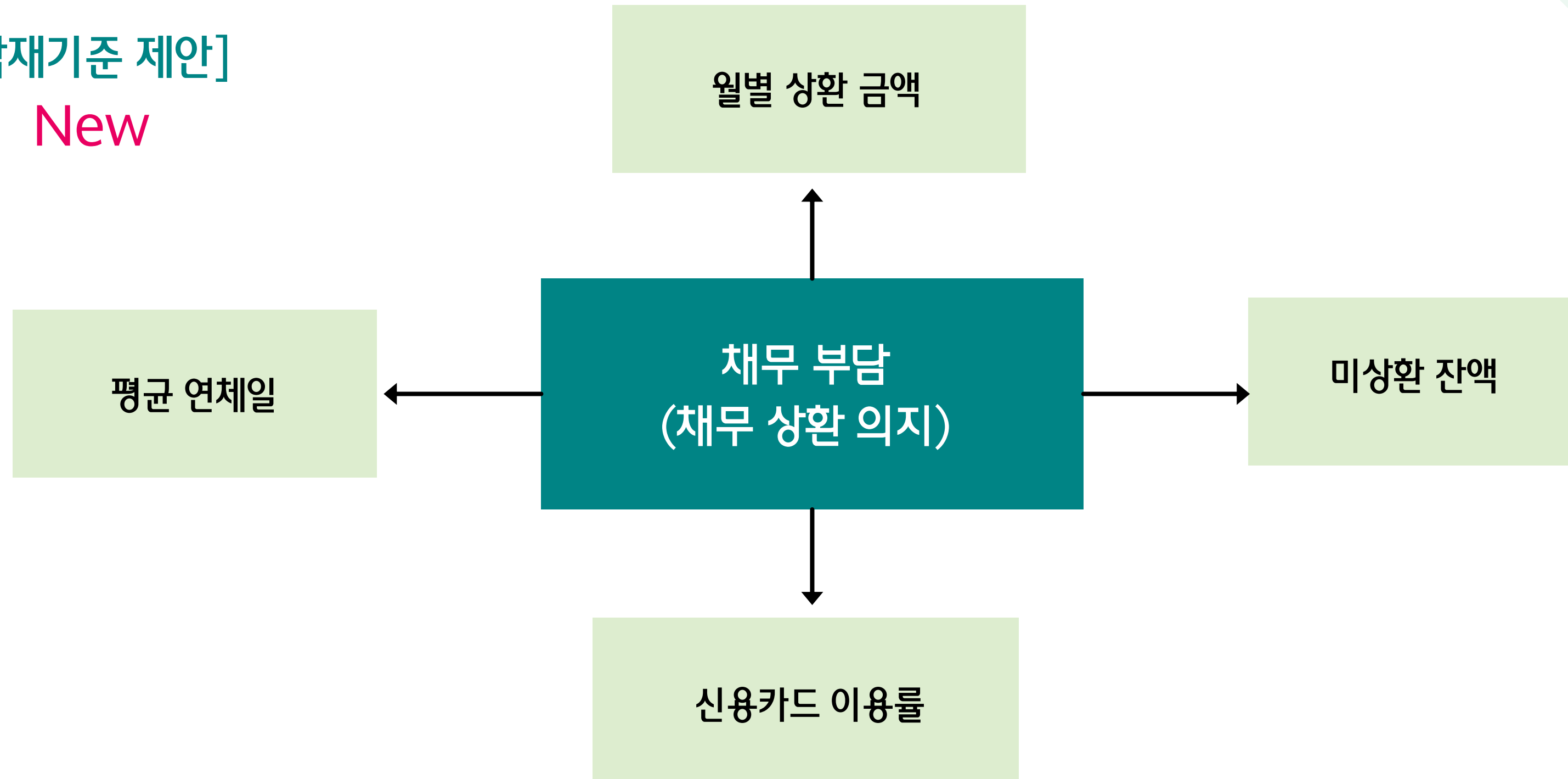
[Credit\_History\_Age]

## 5. 자체 평가 의견

## 5.1 중신용자 대출 승인을 위한 잠재 기준 제안

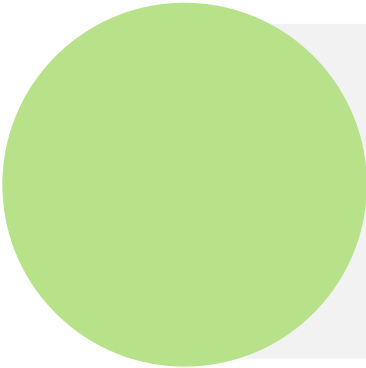
[잠재기준 제안]

New

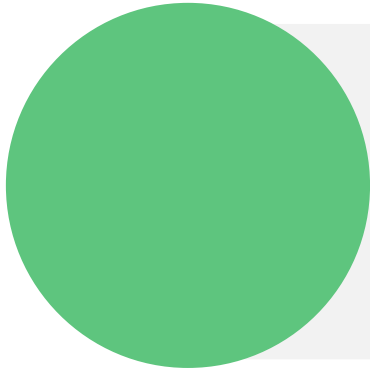


**채무 부담(채무 상환 의지)**라는 새로운 대출 승인 잠재 기준 제안

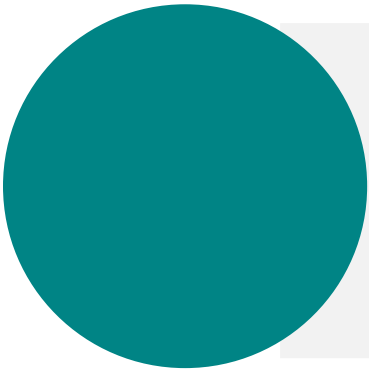
## 5.2 기대효과



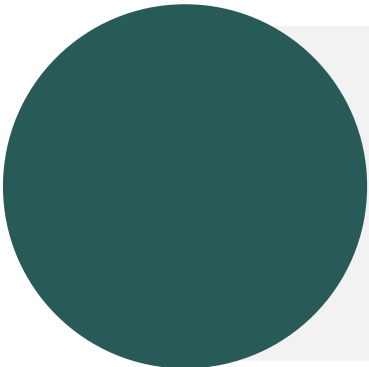
정교한 신용등급분류 모델링으로  
개인 대출 고객 디폴트 리스크 최소화



그레이존 고객의 대출 심사 기준을  
재정립함으로써 대출 규모 확대



기존 금융 기관들의 새로운 시장 개척으로  
기업의 필요 자금 확보 및 수익률 제고



향후 비재무 정보와 결합하여 저신용자를 위한  
더 정교한 대출 승인 기준 제안 가능

# 5.3 역할 분담 및 자체 평가

## 역할 분담

이름	역할	담당업무
구영지	팀장	▶ 데이터 문제정의 및 전처리, 데이터 품질 확인(결측치, 이상치 처리), 하이퍼파라미터 튜닝
김재현	팀원	▶ 데이터 문제정의 및 전처리, 통계분석, 머신러닝 분류 모델
김정민	팀원	▶ 데이터 문제정의 및 전처리, 통계분석, 데이터 품질 확인(결측치, 이상치 처리), ppt 디자인
이한비	팀원	▶ 데이터 문제정의 및 전처리, 머신러닝 분류 모델, 데이터 품질 확인(결측치, 이상치 처리)
임채영	팀원	▶ 데이터 문제정의 및 전처리, 통계분석, 하이퍼파라미터 튜닝, 발표

## 자체 평가

분석에 사용할 수 있는 데이터가 한정적이어서  
보다 깊은 분석에 대한 어려움

2주라는 단기 프로젝트에 대한 아쉬움

모델 튜닝과 데이터 인사이트 도출 과정에서  
트렌드 기술 학습의 필요성

금융권을 준비하는 교육생들과 함께  
프로젝트를 진행하면서 쌓인 도메인 지식



THANK YOU

