

[신용 점수 예측과 맞춤형 마케팅]

딥러닝을 통한 고객 분석

제출자	3조 박지인
제출 일시	2024.09.13

[목차]

01.

시각화 EDA 및 데이터 전처리

02.

모델링 및 사용기법

03.

인사이트 도출 및 제안

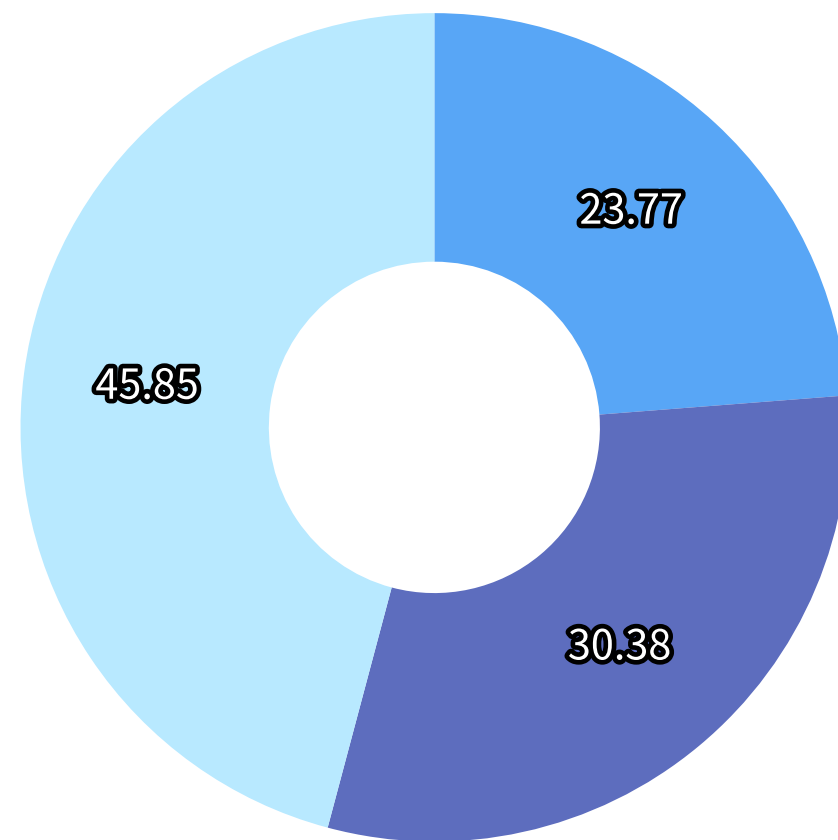
04.

고찰

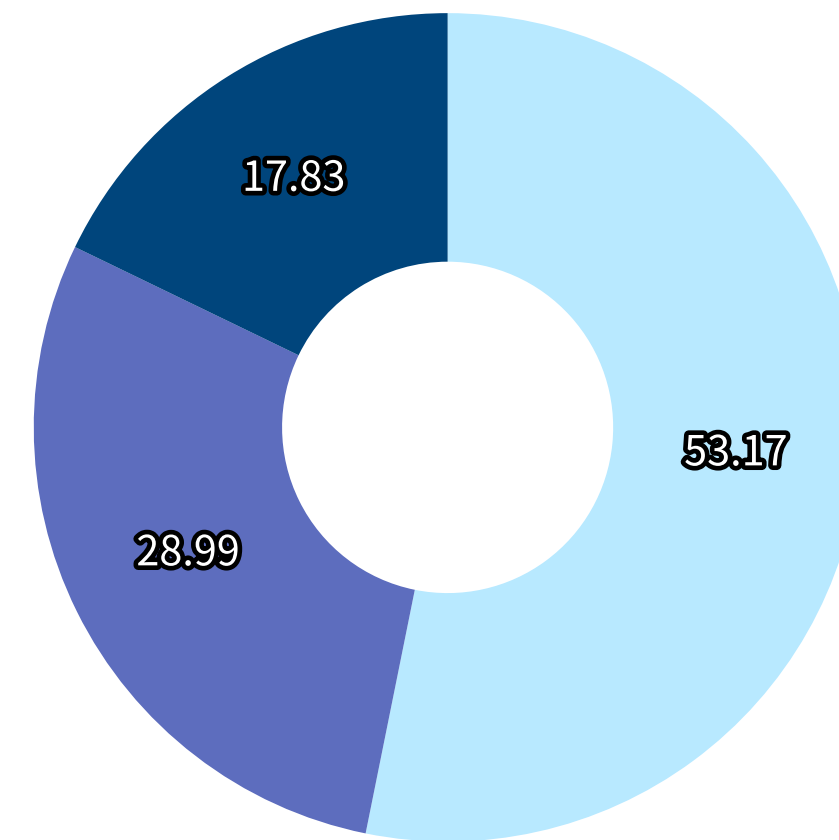
[데이터 전처리]

파생변수명	변수설명	
	변수 생성 방식	변수의미
Total Debt Ratio (총 대출 부채 비율)	$\text{Outstanding_Debt} / \text{Annual_Income}$	개인의 전체 부채가 연간 소득에 비해 어느 정도인지
Interest Burden (이자 부담)	$\text{Interest_Rate} * \text{Outstanding_Debt}$	개인이 대출이나 신용카드 부채에 대해 매달 부담하는 이자의 양
Credit Usage to Limit Ratio (신용카드 사용량 대비 한도 비율)	$\text{Credit_Utilization_Ratio} / (\text{Changed_Credit_Limit} + 1e-9)$	신용카드 한도 중 얼마나 사용했는지
Delinquency Rate (연체율)	$\text{Num_of_Delayed_Payment} / (\text{Num_of_Loan} + 1e-9)$	대출 중 얼마나 자주 연체했는지
Loans to Credit Cards Ratio (대출 수 대비 신용카드 수 비율)	$\text{Num_of_Loan} / (\text{Num_Credit_Card} + 1e-9)$	대출의 수와 신용카드의 수를 비교하여 신용 계좌의 구성을 평가

[시각화 EDA]

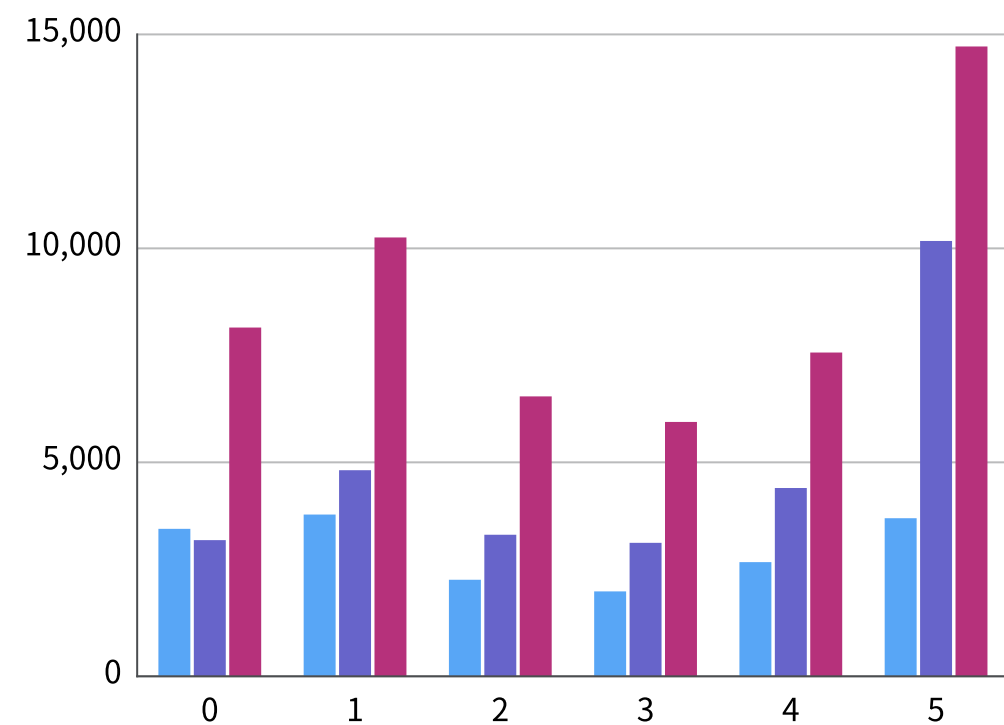


credit score의 분포 비율
standard > poor > good



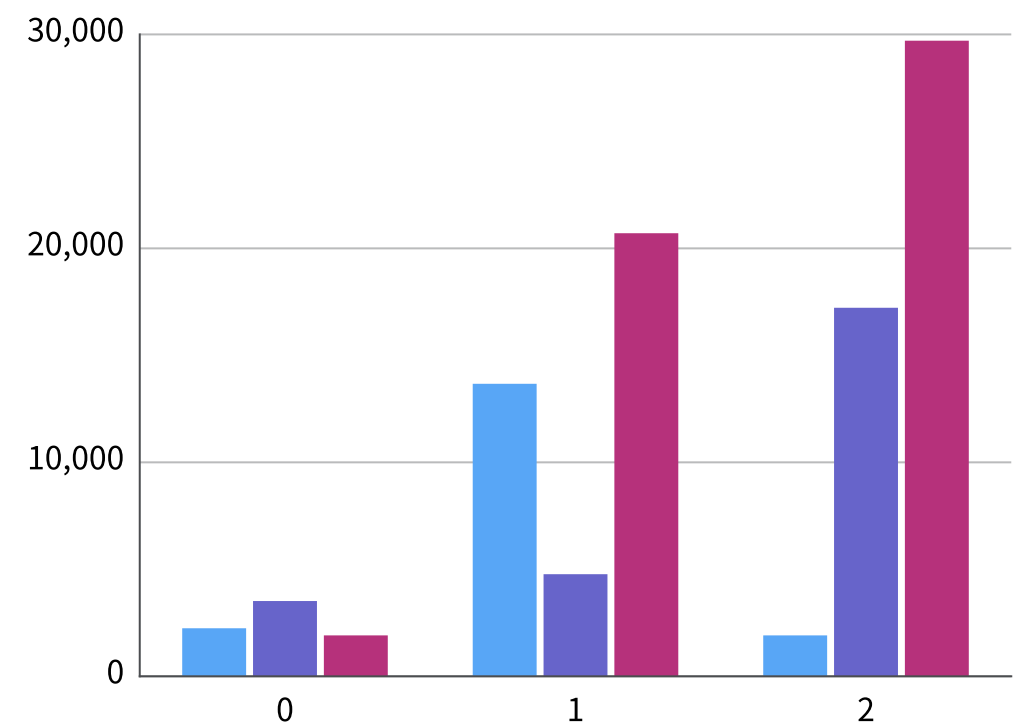
Credit_Mix(상관성이 제일 높음)의 분포 비율
1.6~2.0 > 0.8~1.2 > 0~0.4 > 0.4~0.8, 1.2~1.6 = 0

[시각화 EDA]



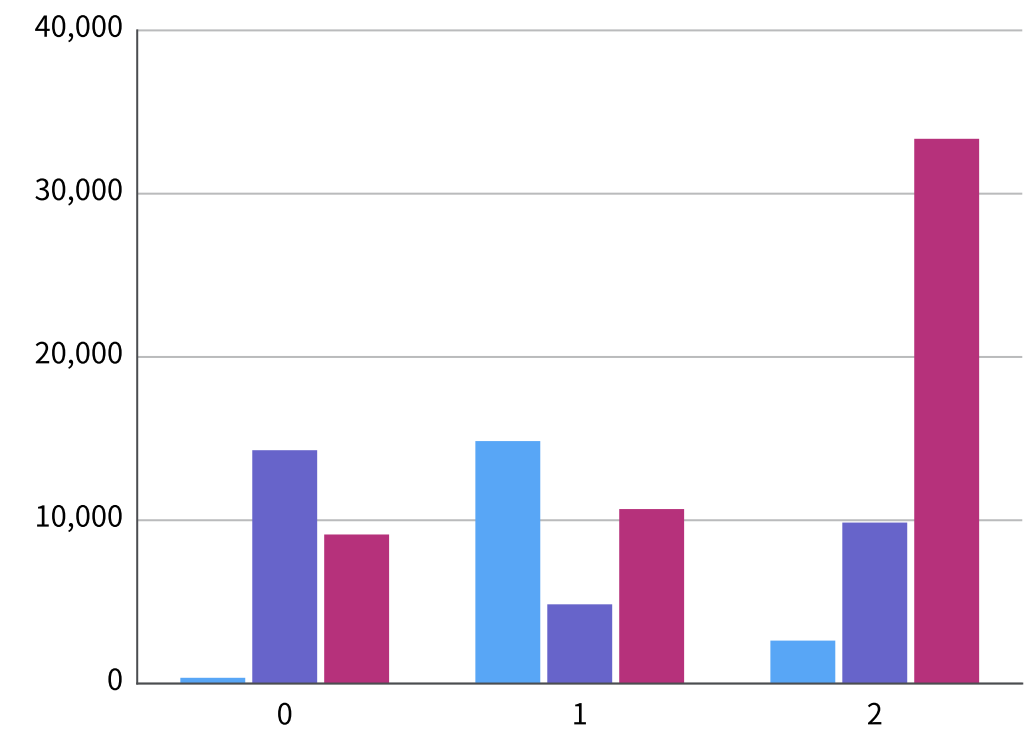
payment behaviour - credit score

최소 지불 금액을 지불하거나
지불해야하는 금액이 없을 때
credit score가 높았습니다



payment of min amount - credit score

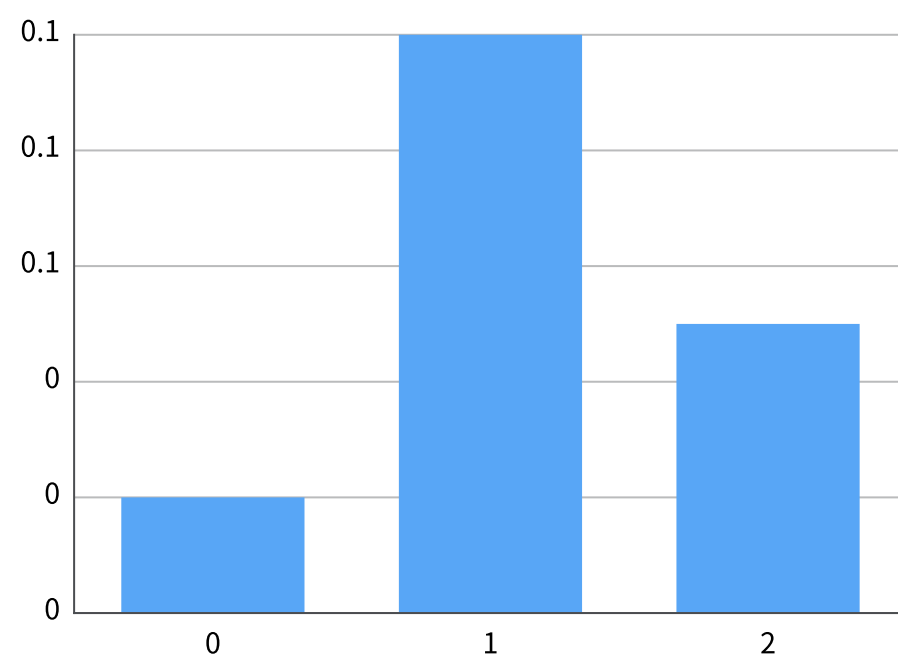
2는 지불해야하는 금액이 없는
경우이므로 최소금액을 지불하는
1의 credit score가 높았습니다



credit mix - credit score

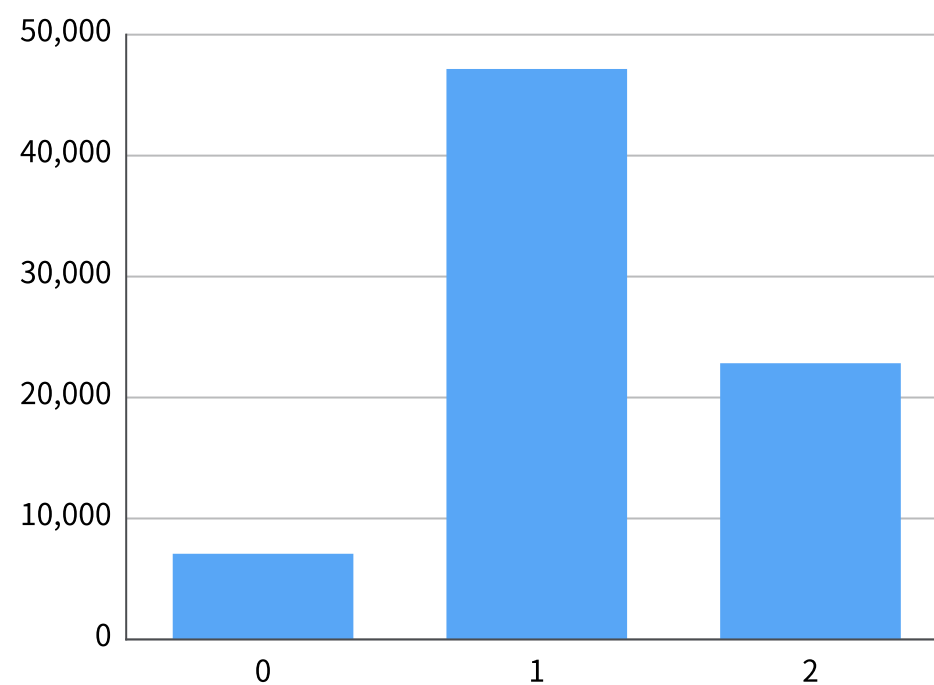
신용관리가 안될 수록 credit
score가 낮고 잘될수록 높은
경향이 있습니다

시각화 EDA



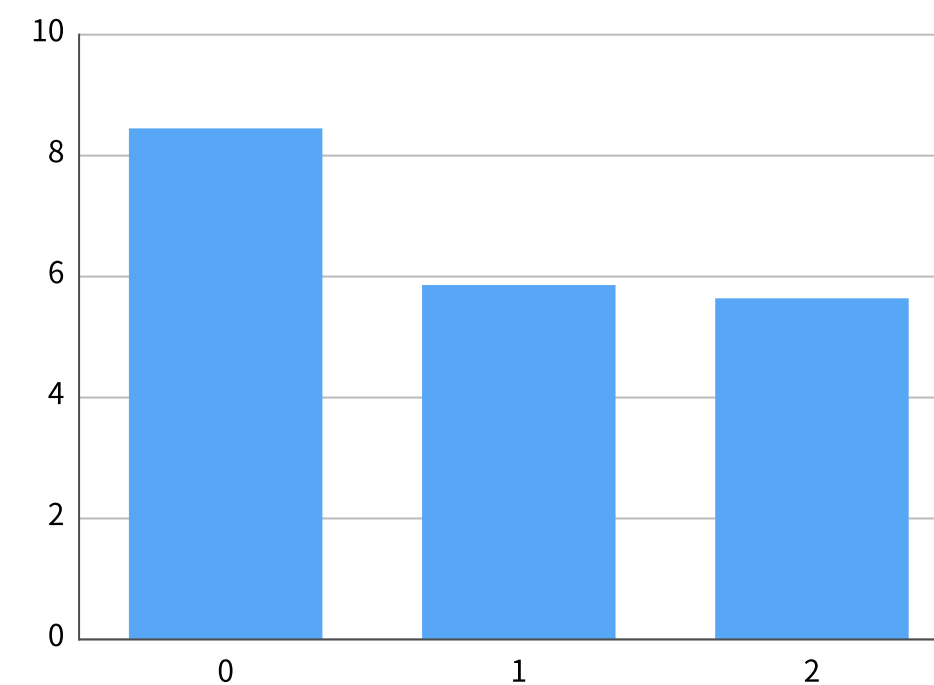
Total debt ratio - credit score

개인 부채율과 신용등급은 반비례 현상을 띄지 않았습니다



interest burden - credit score

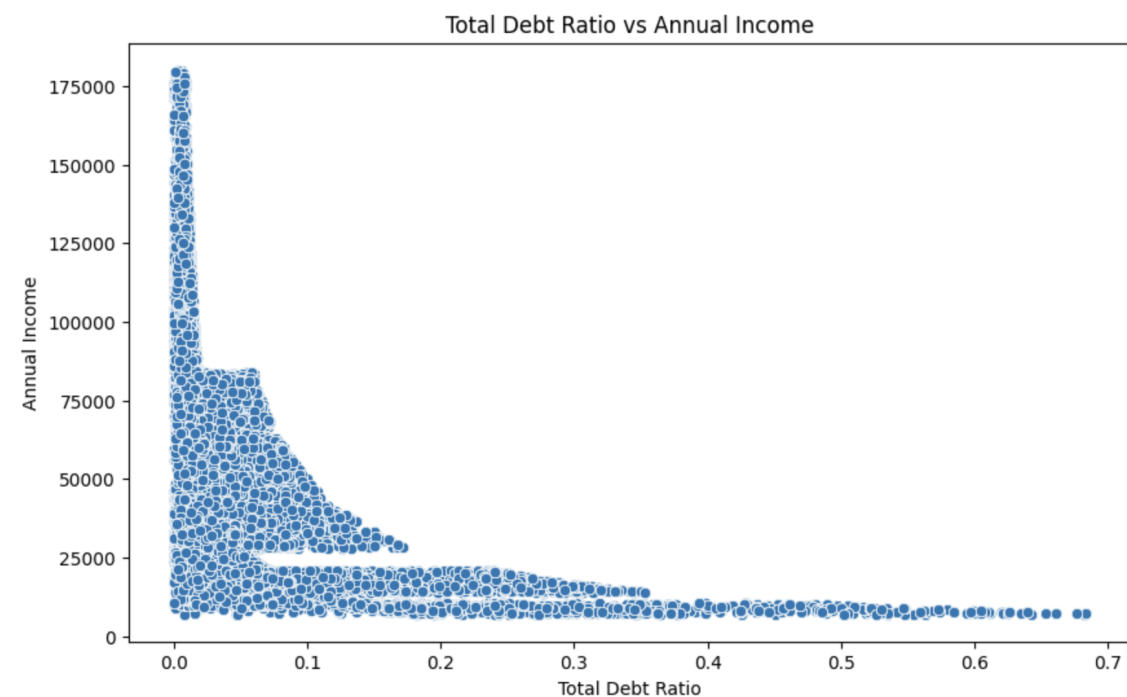
신용등급이 낮으면 매달 감당율이 낮습니다. 신용등급이 높다고 신용 사용율이 높지 않습니다



credit usage limit ratio - credit score

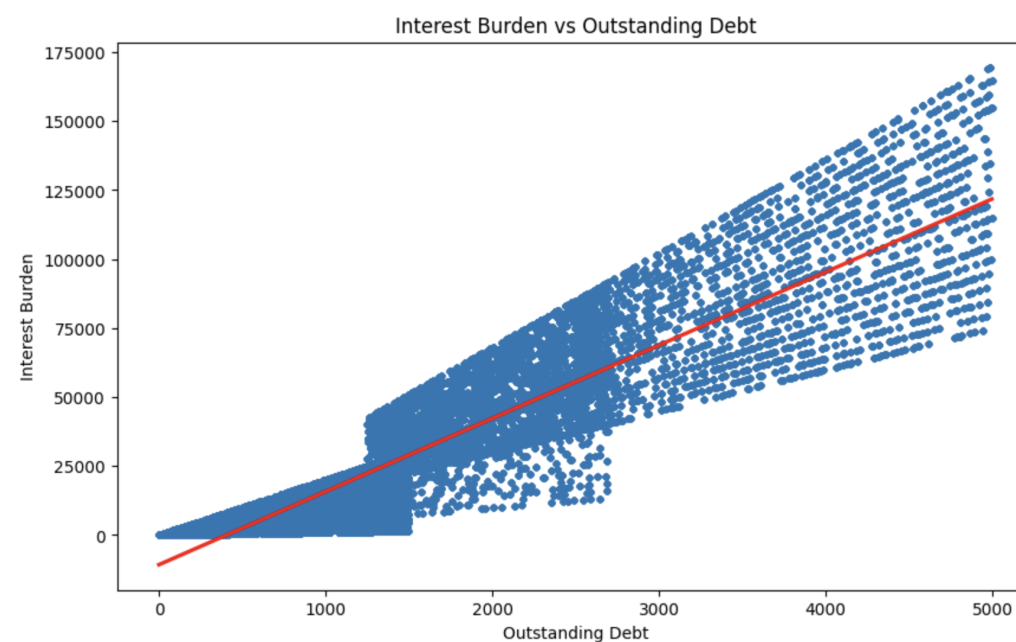
한도 신용을 모두 사용한 경우에 신용등급이 낮은 경우가 많았습니다

시각화 EDA



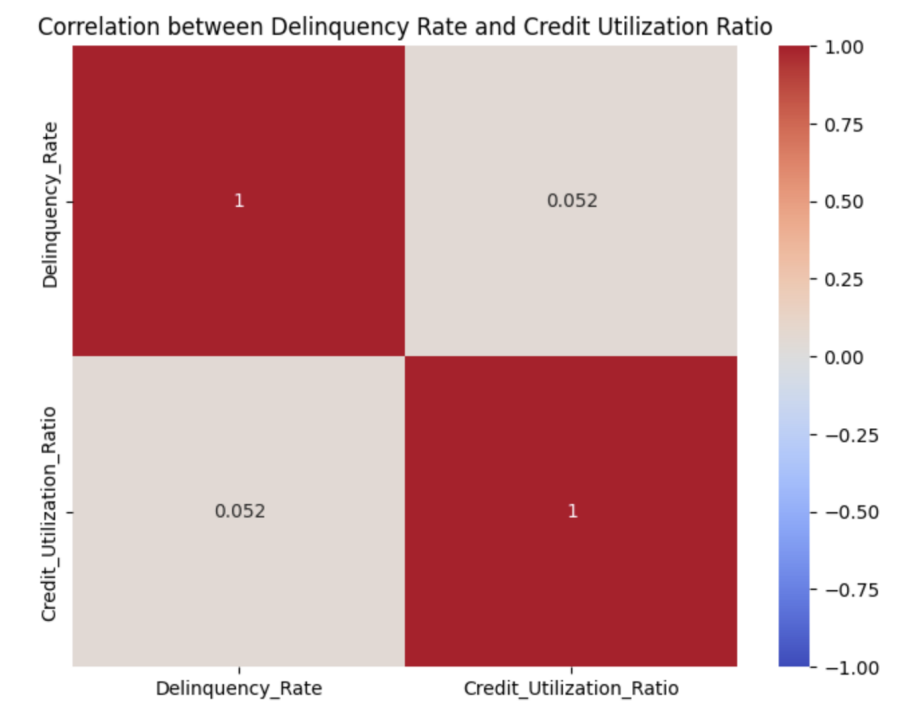
연간 수입 - 부채율

연간 수입과 부채율음 음의 곡선적 관계를 가집니다. 즉 수입이 크면 부채율이 낮고 수입이 일정이하이면 부채율이 높은 경향을 가집니다



이자부담 - 미지급액

이자부담이 커질수록 미지급액도 커지는 선형 비례 경향을 가집니다. 하지만 완전히 비례하지는 않지만 예외적인 상황이 없습니다.



신용카드 사용률 - 연체율

상관계수 값이 0.052로 매우 낮습니다. 이는 두 변수 간에 상관관계가 거의 없다는 것을 의미합니다.

[모델링 및 사용기법]

모델 종류	MLP
옵션	3개의 완전연결층(64, 32, 3) ReLU 활성화 함수 20 에포크 동안 학습
점수	Epoch [20/20] Loss: 0.5947 Valid Loss: 0.6270 Validation Accuracy: 71.48%

모델 종류	선형회귀
옵션	손실함수 : MSE 옵티마이저 : Adam
점수	Epoch [20/20] Loss: 0.4872 Valid Loss: 0.4877 R ² score: 0.1610

모델 종류	MLP
옵션	파생변수 추가 20에폭
점수	Epoch 50/50, Loss: 0.5572 Validation Accuracy: 72.45%

파생변수를 추가한 MLP모델의 성능이 제일 좋아
이 모델을 기반으로 여러가지 옵션을 추가함

[모델링 및 사용기법]

모델 종류	MLP
옵션	50 에포크
점수	Epoch 50/50 Loss: 0.5572 Validation Accuracy: 72.45%

모델 종류	MLP
옵션	30 에포크
점수	Epoch 30/30 Loss: 0.5767 Validation Accuracy: 71.89%

모델 종류	MLP
옵션	100 에포크
점수	Epoch 100/100 Loss: 0.5335 Validation Accuracy: 72.79%

에폭을 바꿔가며 여러가지 비교해 보았을 때,
100에포크의 성능이 제일 좋다

[모델링 및 사용기법]

모델 종류	MLP
옵션	100에폭 레이어 추가
점수	Loss: 0.3636 Validation Accuracy: 75.51%

모델 종류	MLP
옵션	50에폭 레이어 추가
점수	Epoch 50/50 Train Loss: 0.4115 Valid Loss: 0.5967 Valid Accuracy: 76.13% Final Validation Accuracy: 76.13%

성능이 좋았던 두 에폭에 레이어를 추가해보았더니
두가지 모두 각각 2.72%, 3.34% 성능이 개선되었다.

[모델링 및 사용기법]

모델 종류	MLP
옵션	히든사이즈 변경(256, 128, 64)
점수	Validation Accuracy: 71.14% Validation Loss: 0.6264



모델 종류	MLP
옵션	학습률 0.001로 조정 추가
점수	Train Loss: 0.5753 Valid Loss: 0.6264 Valid Accuracy: 71.70% Final Validation Accuracy: 71.70%

모델 종류	MLP
옵션	학습률 0.0001로 조정 추가
점수	Epoch 50/50 Train Loss: 0.5797 Valid Loss: 0.6238 Valid Accuracy: 71.52% Final Validation Accuracy: 71.52%

히드사이즈를 변경하고 학습률을 더 작게 조정한
모델의 성능이 계속해서 개선되고 있다
조기 종료 없이 높은 성능의 모델이 버려지고 있

[모델링 및 사용기법]

모델 종류	MLP
옵션	조기종료 추가 Early stopping at epoch 22
점수	Final Validation Accuracy: 73.45% bad f1 score : 0.65 neutral f1 score : 0.72 good f1 score : 0.77



모델 종류	MLP
옵션	드롭 아웃 추가
점수	Early stopping at epoch 16 Final Validation Accuracy: 71.14% bad f1 score : 0.63 neutral f1 score : 0.69 good f1 score : 75

드롭아웃을 진행하지않은 제안이 덜 걸린 모델이
드롭아웃을 진행한 모델보다 성능이 뛰어나다. 과적합 방지 옵션이 없는
모델과 비교가 필요할 것 같다.

[모델링 및 사용기법]

MLP+ 파생변수 + 에폭 조절 + 레이어 추가

모델 종류	MLP
옵션	50에폭 레이어 추가
점수	Epoch 50/50 Train Loss: 0.4115 Valid Loss: 0.5967 Valid Accuracy: 76.13% Final Validation Accuracy: 76.13%

>

MLP + 파생변수 + 에폭 조절 + 레이어 추가 + 조기종료

모델 종류	MLP
옵션	조기종료 추가 Early stopping at epoch 22
점수	Final Validation Accuracy: 73.45% bad f1 score : 0.65 neutral f1 score : 0.72 good f1 score : 0.77

과적합 방지 옵션이 없는 모델의 성능이 더 뛰어나지만
큰 차이를 보이지 않고 둘다 70이상의 성능을 보임

[인사이트 도출 및 제안]

▶ 재정관리가 잘되는 그룹에게 프리미엄 서비스 제공

대출을 적절하게 관리하고, 연체율이 낮고 신용카드 사용 비율이 높지 않고, 신용 한도를 잘 관리하는 경향이 있는 소득 대비 부채 비율이 낮은 그룹에게 금융 세미나나 투자자 네트워크 행사 초대, 프리미엄 클래스 제공 (예: 와인 시음회, 골프 투어 등)한다.

▶ 재정관리가 안되는 그룹에게 맞춤형 서비스 제공

소득 대비 부채 비율이 높고, 연체가 잦은 경향이 있습니다. 주로 최소 금액 납부를 자주 선택하고, 상환 능력이 부족할 수 있는 그룹이므로 부채 관리나 저금리 대출로의 전환을 유도하여 부채를 줄이는 전략을 제안하고 신용이 매우 낮지 않은 신용카드 이용률이 30% 이하로 유지되도록 제안, 이를 준수할 경우 리워드 제공하는 서비스를 제공한다.

[인사이트 도출 및 제안]

▶ 신용카드 사용 대비 한도 비율이 높은 고객 관리 서비스

신용카드 사용 비율이 높을수록 신용 점수가 떨어지는 경향이 있다. 그러므로 신용카드 한도에 가까운 사용을 계속하는 고객들에게는 신용 한도 상향을 제안하거나, 신용카드 부채 상환을 돕는 상품을 제공하여 신용 점수를 개선할 수 있는 기회를 제공한다

▶ 게임화 전략을 통한 고객 참여율 개선 시스템

신용 점수 개선 활동을 지속적으로 독려할 수 있는 게임화 전략을 도입한다. 고객이 신용 점수 개선 행동을 취할 때마다 포인트를 지급하거나, 레벨 업 시스템을 도입해 고객이 더 높은 신용 점수를 목표로 지속적인 노력을 기울이도록 만든다. 이를 통해 고객의 신용 점수 개선에 대한 동기를 부여할 수 있다.

[고찰]

느낀점 및 아쉬운 점

평소 팀으로 분배해 진행하던 프로젝트를 혼자서 하려고 하니 시간 분배가 어려웠다.
아이디어나 방향성 또한 혼자 생각해야하는 부분이 막막했으며 진행 중에도 잘 하고 있는지에 대해 의문이 계속 들었다.

하지만 서로서로 의견을 나누기도 하고 꼭 팀이 아니여도 도움을 주고 받다보니 잘 진행할 수 있었던 것 같다.

아쉬운 점은 처음 시도해보는 딥러닝 프로젝트이다보니
성능을 많이 개선시키지 못해 아쉽다.

하지만 이론으로만 배우던 내용을 적용하며 스스로 탐구하다보니 이해가 안되고 어려웠던 부분도 알게되었고 프로젝트이 진행 흐름에 대해 깨달을 수 있었다.

다음 프로젝트에는 더 촘촘하게 계획하고 방향성을 고민해보고싶다