

* loan\_amnt :
* recovery : 탕감해준뒤 남은 빚 / 높을수록 부도할 가능성이 높다.
* total\_rec\_int : 현재까지 받은 이자 / 이자가 많다는 것은 많이 빌렸고, 오래빌려서 높은 것(빚에 대한 의존도가 높다)
* **fico\_range\_low : 양의 상관관계라서 해석이 모호함**
* revol\_bal : 리볼빙 금액이 높을수록 부도가능성이 크다.
* ㅇ
* funded\_amnt : 채무자가 처음에 받은 대출금
* funded\_amnt\_inv :
* **fico\_range\_high : 양의 상관관계라서 해석이 모호함**
* installment: 월 납부금 높다는것은 대출원금이 높으니까 돈이 없는사람. -> 부도확률이 높다.
* tot\_acc : 사용한 신용 계좌 갯수 - 돈이 없으니깐 신용을 끌어다 쓴것이고, 부도확률이 돞다.
* out\_prncp\_inv : 투자받은 금액 중 갚고 남은 빚 / 남아있는 빚의 양이 클수록 부도확률이 높다.
* **annual inc : 연소득이 높으면 부도확률이 높다?**
* tot\_coll\_amnt : 신용을 많이 썼을수록 부도확률이 높다
* **last\_fico\_range\_high : 양의 상관관계라서 해석이 모호함**
* tot\_rec\_rate\_fee : 현재까지 받은 연체료 - 연체를 많이할수록 신용이 낮으니… 부도가능성 큼
* out\_prncp : 갚고 남은 빚 / 남아있는 빚의 양이 클수록 부도확률이 높다.
* revol\_until : 높을수록 리볼빙을 많이 한것으로 - 부도확률이 높다.
* delling\_2yrs : 연체된 횟수는 높을수록 부도확률이 높다.
* debt\_settlement\_flag1 : 신용회복위원회의 도움을 받아서 상환을 하고 있는지. / 스스로 상환하는 사람은 0, 도움을 받는사람은 1 / 도움을 받는경우 부도확률이 높다.
* int\_rate : 이율이 높다는 것은 그사람이 신용이 낮다는 뜻이니깐 -> 부도확률이 높다.
* 부도확률이 높다.

// 더미들은 pass?

* mths\_since\_recent\_bc6 : 계좌를 개설한지 얼마나 지났는지.
* mths\_since\_recent\_bc4 :
* **emp\_length12:** 재직기간이 길면… 이제 곧 짤릴 수 있다던지/돈을 모을 수 없는 환경이라던지… ->**|| 제3의 변수로 인해 confounding factor 가 존재해서 둘이 양의 상관관계를 나타낸 것. (-> 부도낸 사람들이 대부분 income이 낮지않고 credit rate도 높아서 emp\_length12 가 높게 나왔을 수 있다)**
* home\_ownership5:
* purpose10:
* mths\_since\_revol\_deling10 :
* mths\_since\_rcnt\_il3:
* mths\_since\_rcnt\_il5:
* addr\_state43:
* issue\_d84
* add\_state3:
* **elapsed\_t : 변수 설명이 없음**
* chargeoff\_within\_12\_mths : 지난 12개월간 탕감해준 횟수 / 탕감해준 횟수가 많을수록 부도확률이 높다
* <음의 상관관계 >
* last\_pymnt\_amnt : 마지막에 갚은 금액 / 갚을수록 부도확률이 낮아진다.
* tot\_cur\_bal : 전체 계좌 잔고 / 계좌 잔고가 높을수록 부도확률이 낮다.
* last\_fico\_range\_low : 신용등급이 높아질수록 부도확률이 낮아진다.
* total\_pymnt\_inv :
* total\_pymnt: 현재까지 수령한 총 금액
* last\_pymnt\_amnt: 최근에 받은 총액 / 남은 대출금+이자가 줄어서 부도확률이 낮아진다.
* total\_rec\_prncp: 현재까지 받은 원금 / 남은 대출금이 줄어서 부도확률이 낮아진다.

더미들을 어떻게 처리할것인지에대한 방향 제시.

* Decision Tree Classifier에서 predict한 이유는 depth size를 5, 10, 20으로 다양하게 해봤었고 10이 그나마 적절한 accuracy를 도출해냈다.
* roc[best] = 1- roc[fpr]+roc[tpr] 로 선택한 이유 설명.
* loan\_amnt :
* recovery : 탕감해준뒤 남은 빚 / 높을수록 부도할 가능성이 높다.
* total\_rec\_int : 현재까지 받은 이자 / 이자가 많다는 것은 많이 빌렸고, 오래빌려서 높은 것(빚에 대한 의존도가 높다)
* **fico\_range\_low : 양의 상관관계라서 해석이 모호함**
* revol\_bal : 리볼빙 금액이 높을수록 부도가능성이 크다.
* ㅇ
* funded\_amnt : 채무자가 처음에 받은 대출금
* funded\_amnt\_inv :
* **fico\_range\_high : 양의 상관관계라서 해석이 모호함**
* installment: 월 납부금 높다는것은 대출원금이 높으니까 돈이 없는사람. -> 부도확률이 높다.
* tot\_acc : 사용한 신용 계좌 갯수 - 돈이 없으니깐 신용을 끌어다 쓴것이고, 부도확률이 돞다.
* out\_prncp\_inv : 투자받은 금액 중 갚고 남은 빚 / 남아있는 빚의 양이 클수록 부도확률이 높다.
* **annual inc : 연소득이 높으면 부도확률이 높다?**
* tot\_coll\_amnt : 신용을 많이 썼을수록 부도확률이 높다
* **last\_fico\_range\_high : 양의 상관관계라서 해석이 모호함**
* tot\_rec\_rate\_fee : 현재까지 받은 연체료 - 연체를 많이할수록 신용이 낮으니… 부도가능성 큼
* out\_prncp : 갚고 남은 빚 / 남아있는 빚의 양이 클수록 부도확률이 높다.
* revol\_until : 높을수록 리볼빙을 많이 한것으로 - 부도확률이 높다.
* delling\_2yrs : 연체된 횟수는 높을수록 부도확률이 높다.
* debt\_settlement\_flag1 : 신용회복위원회의 도움을 받아서 상환을 하고 있는지. / 스스로 상환하는 사람은 0, 도움을 받는사람은 1 / 도움을 받는경우 부도확률이 높다.
* int\_rate : 이율이 높다는 것은 그사람이 신용이 낮다는 뜻이니깐 -> 부도확률이 높다.
* 부도확률이 높다.

// 더미들은 pass?

* mths\_since\_recent\_bc6 : 계좌를 개설한지 얼마나 지났는지.
* mths\_since\_recent\_bc4 :
* **emp\_length12:** 재직기간이 길면… 이제 곧 짤릴 수 있다던지/돈을 모을 수 없는 환경이라던지… ->**|| 제3의 변수로 인해 confounding factor 가 존재해서 둘이 양의 상관관계를 나타낸 것. (-> 부도낸 사람들이 대부분 income이 낮지않고 credit rate도 높아서 emp\_length12 가 높게 나왔을 수 있다)**
* home\_ownership5:
* purpose10:
* mths\_since\_revol\_deling10 :
* mths\_since\_rcnt\_il3:
* mths\_since\_rcnt\_il5:
* addr\_state43:
* issue\_d84
* add\_state3:
* **elapsed\_t : 변수 설명이 없음**
* chargeoff\_within\_12\_mths : 지난 12개월간 탕감해준 횟수 / 탕감해준 횟수가 많을수록 부도확률이 높다
* <음의 상관관계 >
* last\_pymnt\_amnt : 마지막에 갚은 금액 / 갚을수록 부도확률이 낮아진다.
* tot\_cur\_bal : 전체 계좌 잔고 / 계좌 잔고가 높을수록 부도확률이 낮다.
* last\_fico\_range\_low : 신용등급이 높아질수록 부도확률이 낮아진다.
* total\_pymnt\_inv :
* total\_pymnt: 현재까지 수령한 총 금액
* last\_pymnt\_amnt: 최근에 받은 총액 / 남은 대출금+이자가 줄어서 부도확률이 낮아진다.
* total\_rec\_prncp: 현재까지 받은 원금 / 남은 대출금이 줄어서 부도확률이 낮아진다.

더미들을 어떻게 처리할것인지에대한 방향 제시.