Credit Scoring Report (HMEQ)
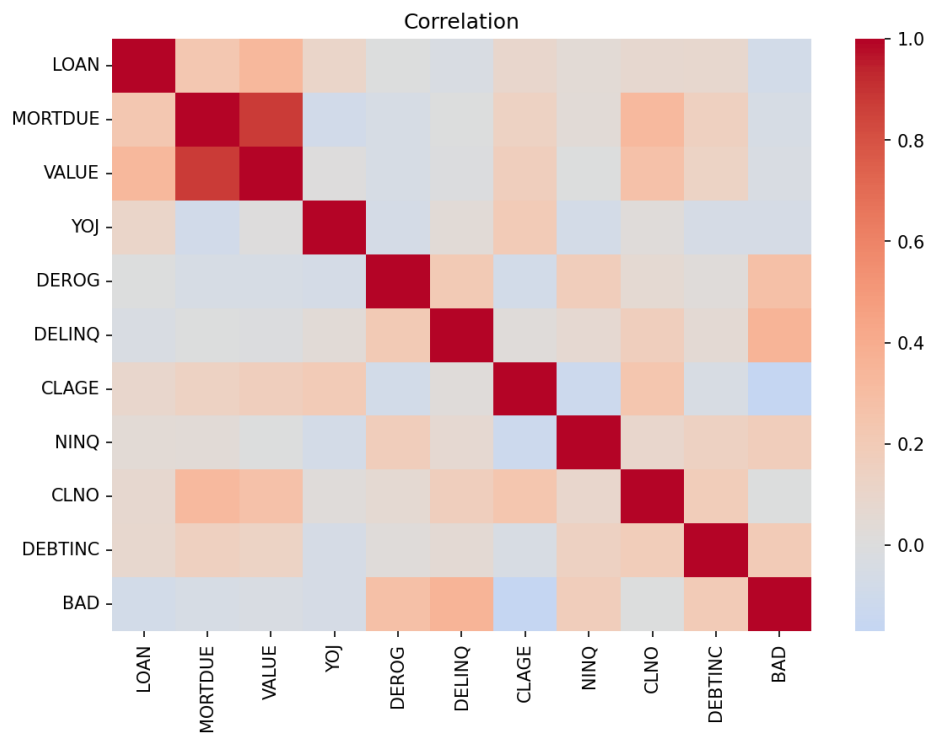
Rows = 5960, Cols = 13, Target = BAD

Exploratory Data Analysis / Phân tích dữ liệu

	BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE	NINQ	
0	1	1100	25860.0	39025.0	HomeImp	Other	10.5	0.0	0.0	94.366667	1.0	
1	1	1300	70053.0	68400.0	HomeImp	Other	7.0	0.0	2.0	121.833333	0.0	1
2	1	1500	13500.0	16700.0	HomeImp	Other	4.0	0.0	0.0	149.466667	1.0	1
3	1	1500	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	0	1700	97800.0	112000.0	HomeImp	Office	3.0	0.0	0.0	93.333333	0.0	1

Missing Values / Giá trị khuyết

DEBTINC	0.212584
DEROG	0.118792
DELINQ	0.097315
MORTDUE	0.086913
YOJ	0.086409
NINQ	0.085570
CLAGE	0.051678
JOB	0.046812
REASON	0.042282
CLNO	0.037248
VALUE	0.018792
LOAN	0.000000
BAD	0.000000

Correlation

```
     Var        VIF
    LOAN   4.214961
 MORTDUE  11.128686
   VALUE  12.152187
     YOJ   2.552094
   DEROG   1.143633
  DELINQ   1.214113
   CLAGE   5.722485
    NINQ   1.587066
    CLNO   6.741402
 DEBTINC   9.068683
```

WOE & IV / Ch n bi n / WOE & IV Feature Selection

```
     Var        IV
 DEBTINC  1.877128
   VALUE  0.470688
  DELINQ  0.452361
   DEROG  0.276585
   CLAGE  0.230638
    NINQ  0.165557
    LOAN  0.160155
     JOB  0.123728
     YOJ  0.083533
    CLNO  0.079103
 MORTDUE  0.048920
  REASON  0.008618
```

Selected variables (8): DEBTINC, VALUE, DELINQ, DEROG, CLAGE, NINQ, LOAN, JOB

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
const	-1.374353	0.046131	-29.792394	4.901588e-195	-1.464768	-1.283937
DEBTINC_WOE	-0.929462	0.031224	-29.767908	1.017120e-194	-0.990659	-0.868265
VALUE_WOE	-0.980790	0.083000	-11.816726	3.199095e-32	-1.143467	-0.818112
DELINQ_WOE	-0.979279	0.067809	-14.441822	2.822796e-47	-1.112181	-0.846376
DEROG_WOE	-0.791717	0.084455	-9.374393	6.957488e-21	-0.957246	-0.626187
CLAGE_WOE	-1.129374	0.094778	-11.916033	9.764832e-33	-1.315135	-0.943613
NINQ_WOE	-0.588538	0.106634	-5.519250	3.404504e-08	-0.797536	-0.379540
LOAN_WOE	-0.346030	0.111754	-3.096361	1.959116e-03	-0.565063	-0.126997
JOB_WOE	-0.940376	0.132242	-7.111000	1.152050e-12	-1.199567	-0.681186

Model Training / Hu n luy n mô hình

Gini and Overfitting diagnostics

Model	AUC_tr	AUC_te	Gini_tr	Gini_te	CV_mean	KS
Logistic	0.820363	0.773767	0.640725	0.547535	0.804198	0.425188
XGBoost	0.978786	0.942572	0.957573	0.885144	0.947628	0.759021
LightGBM	0.989389	0.946656	0.978778	0.893311	0.955596	0.767220

Overfit check

- Logistic: ΔAUC=0.047 → OK
- XGBoost: ΔAUC=0.036 → OK
- LightGBM: ΔAUC=0.043 → OK

Bootstrap AUC Confidence Intervals

Model	AUC_tr	AUC_te	Boot_mean	CI_low	CI_high	CI_width
Logistic	0.820363	0.773767	0.774517	0.741397	0.806872	0.065475
XGBoost	0.978786	0.942572	0.942733	0.927701	0.957899	0.030199
LightGBM	0.989389	0.946656	0.946819	0.931686	0.961540	0.029854