DA Internship - Final Project Presentation

Shanney Suhendra
19th June - 16th August, 2023

OBJECTIVES:

1. Compare behavior of Performing Loaners (PL) with Non-Performing Loaners (NPL)
2. Identify significant variables that differentiate NPLs from PLs

DT	AVG_12MTHS_FREK_DB	AVG_6MTHS_FREK_CR	FLAG_RESTRU	~~6to3DB~~	slopeFDB
ID_DEB	AVG_12MTHS_FREK_CR	AVG_3MTHS_CASA	FLAG_DEFERRED	~~12to9FCR~~	interceptFDB
DEBTOR_CATEGORY	AVG_9MTHS_CASA	AVG_3MTHS_DPK	REDFLAG_YELLOW	~~9to6FCR~~	12to3CR
PLAFOND_IDR	AVG_9MTHS_DPK	AVG_3MTHS_AMT_DB	REDFLAG_RED	~~6to3FCR~~	12to3DB
OS_IDR	AVG_9MTHS_AMT_DB	AVG_3MTHS_AMT_CR	REDFLAG_INFORMASI	~~12to9FDB~~	12to3FCR
STRT_PAST_DUE_DT	AVG_9MTHS_AMT_CR	AVG_3MTHS_FREK_DB	FLAG_BLACKLIST	~~9to6FDB~~	12to3FDB
HARI_TUNGGAKAN	AVG_9MTHS_FREK_DB	AVG_3MTHS_FREK_CR	percent_used	~~6to3FDB~~	
CUST_TYPE_CD	AVG_9MTHS_FREK_CR	SALDO_AVG_CASA	total_flags	slopeCR	
COLLECT_KEY	AVG_6MTHS_CASA	SALDO_AVG_DPK	~~12to9CR~~	interceptCR	
AVG_12MTHS_CASA	AVG_6MTHS_DPK	AVG_MUTASI_DB	~~9to6CR~~	slopeDB	
AVG_12MTHS_DPK	AVG_6MTHS_AMT_DB	AVG_MUTASI_CR	~~6to3CR~~	interceptDB	
AVG_12MTHS_AMT_DB	AVG_6MTHS_AMT_CR	FREK_DB	~~12to9DB~~	slopeFCR	
AVG_12MTHS_AMT_CR	AVG_6MTHS_FREK_DB	FREK_CR	~~9to6DB~~	interceptFCR	

DATA PREPARATION

STEPS

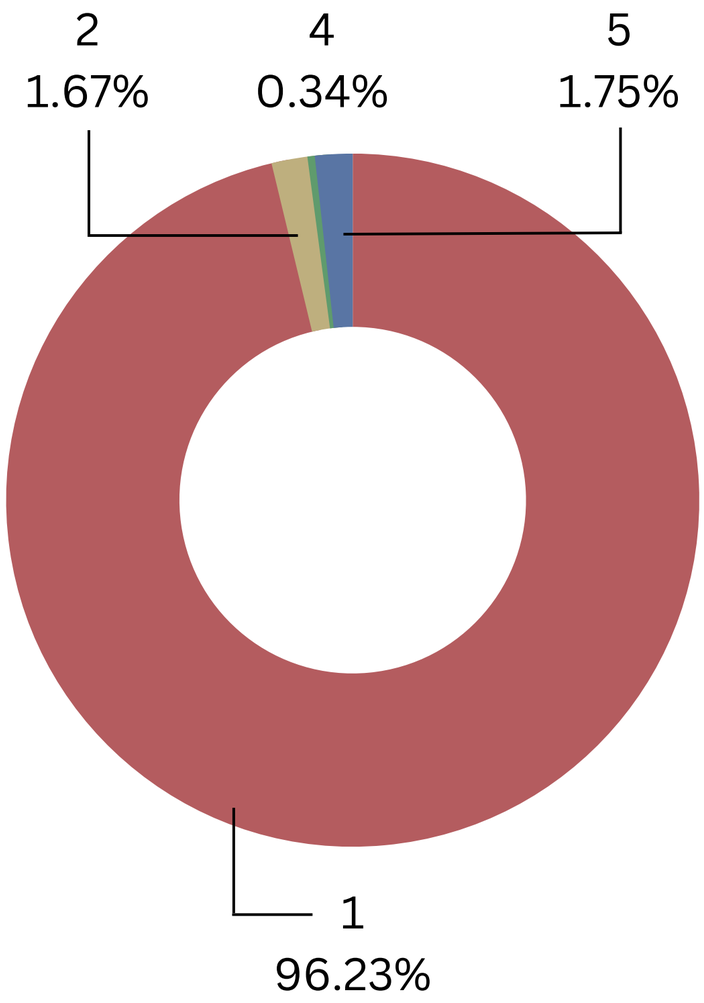
1. remove unwanted variables
2. create new variables
    a. percent_used = OS_IDR/PLAFOND_IDR
    b. total_flags = REDFLAG_YELLOW + 3(REDFLAG_RED) + 5(REDFLAG_INFORMASI) + 7(FLAG_BLACKLIST)
    c. 12to9, 9to6, 6to3 = (9-12)/12, (6-9)/9, (3-6)/6
    d. slope, intercept = derived from linear regression with values from c.
    e. 12to3 = (3-12)/12
3. ~~remove unwanted variables~~ (use correlation matrix to determine (pairs = over 90% correlation)

Loaners Breakdown

PL Performing Loaners

NPL Non-Performing Loaners

Collect Key



1 35,948

2 626

PL 36,574

Customer Type

Organization: 41%
Individual: 59%

Debtor Category

SME: 93.9%
Kommersial: 6.9%

3 127

4 654

NPL 781

Customer Type

Organization: 33.5%
Individual: 66.5%

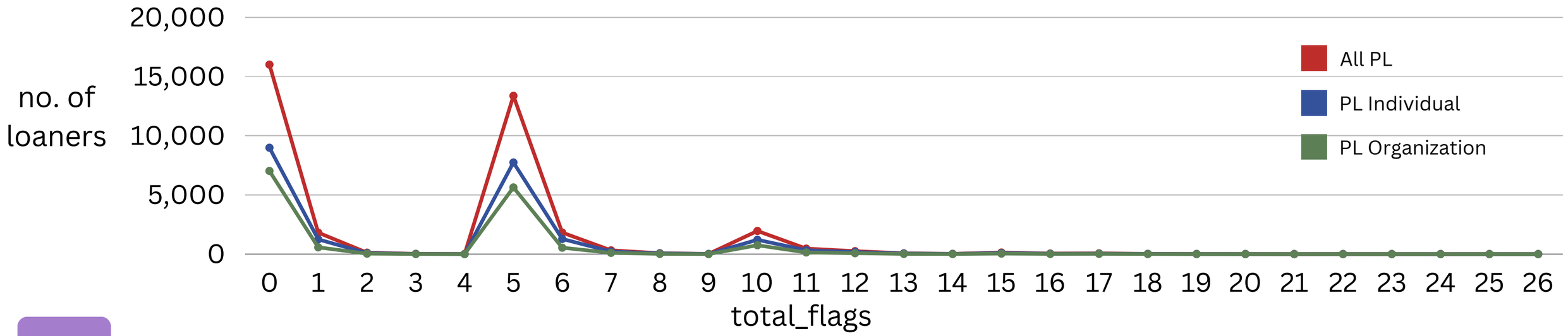
Debtor Category

SME: 92.8%
Kommersial: 7.2%

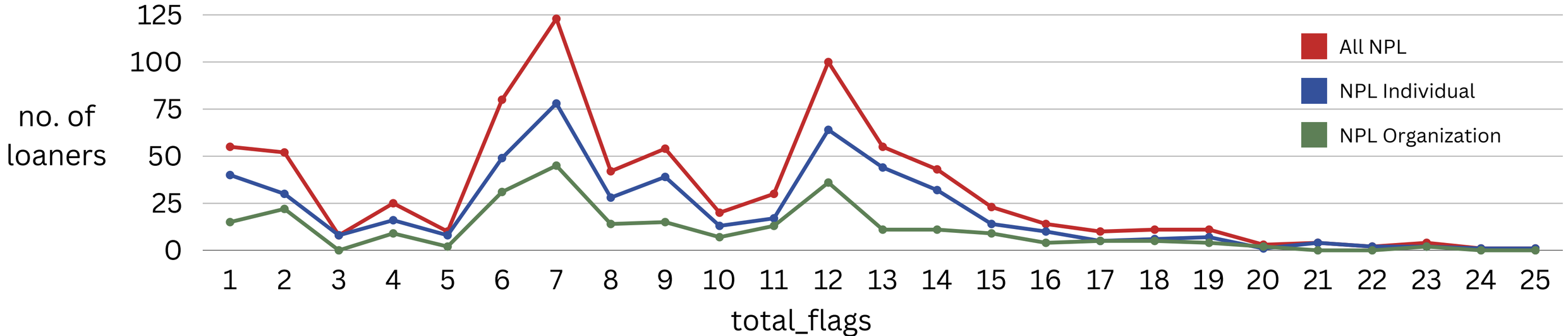
*not significant

Frequency of Flags (Compare Customer Type)

PL



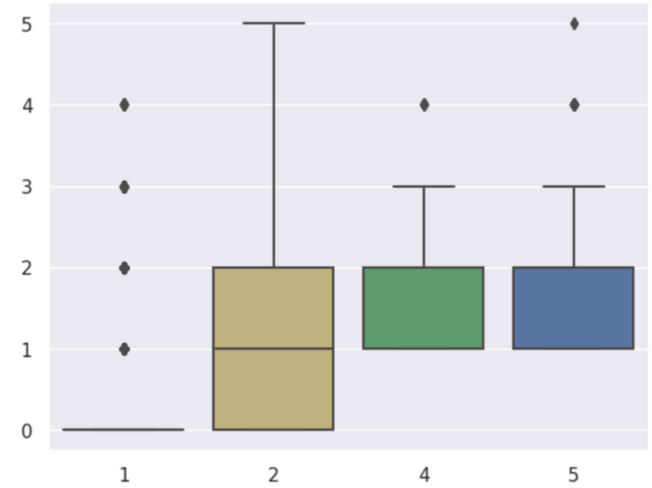
NPL



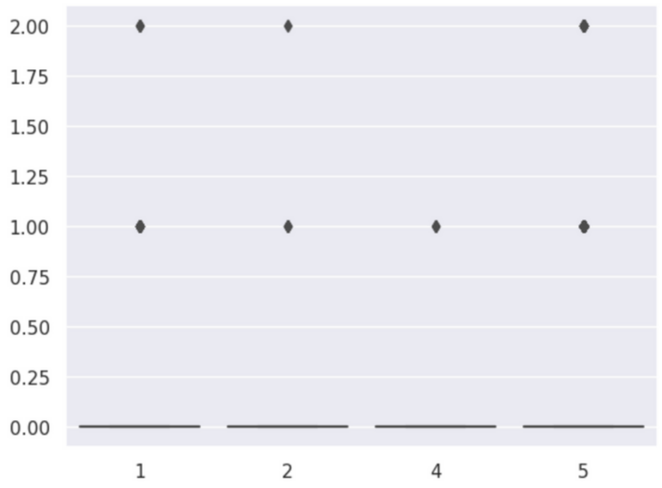
Loaners Category vs No. of Flags

x-axis: no. of flags
y-axis: collect_key

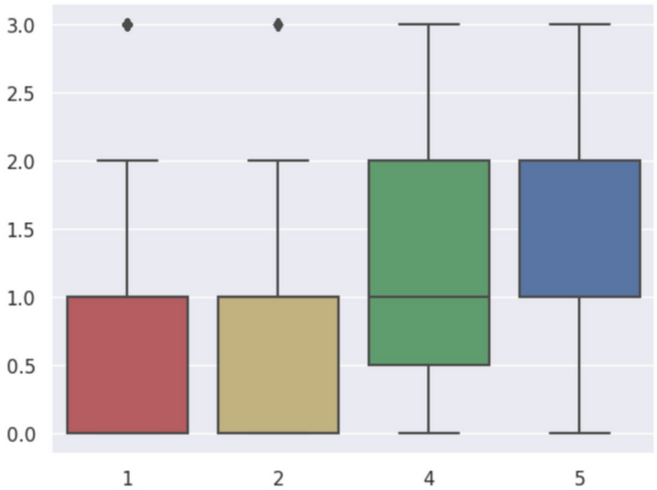
yellow



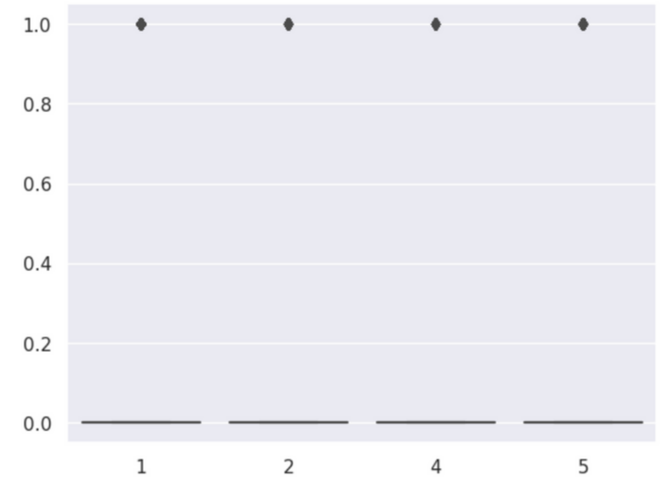
red



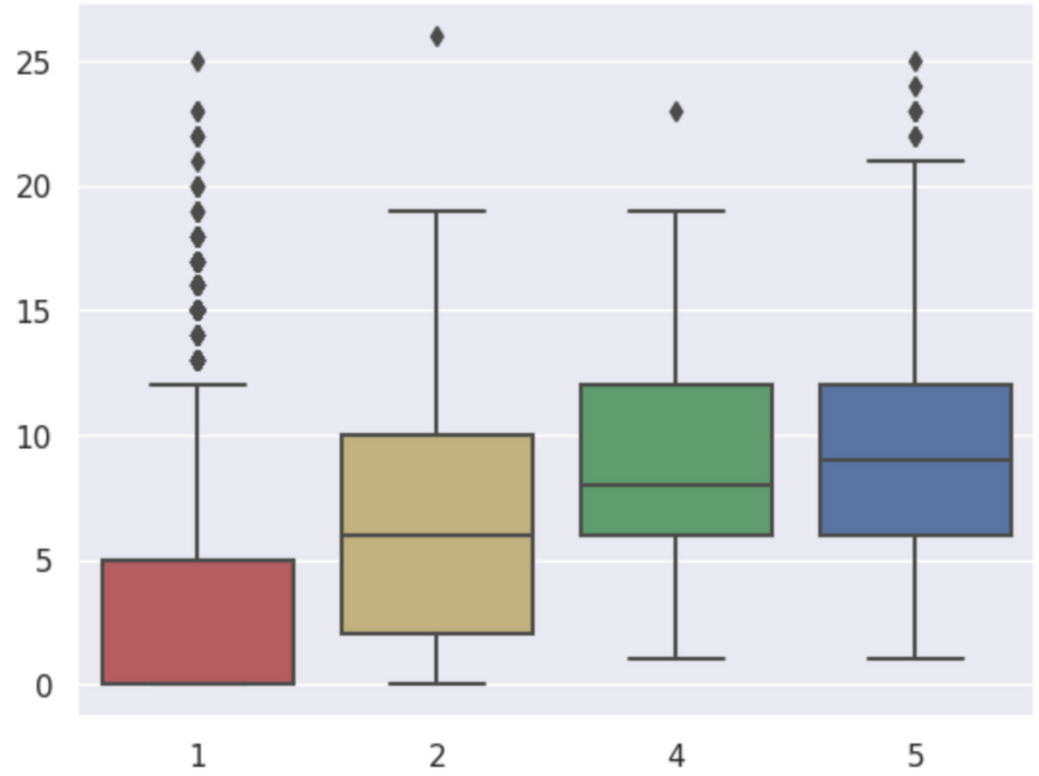
informasi



blacklist



total flags = yellow + 3(red) + 5(informasi) + 7(blacklist)

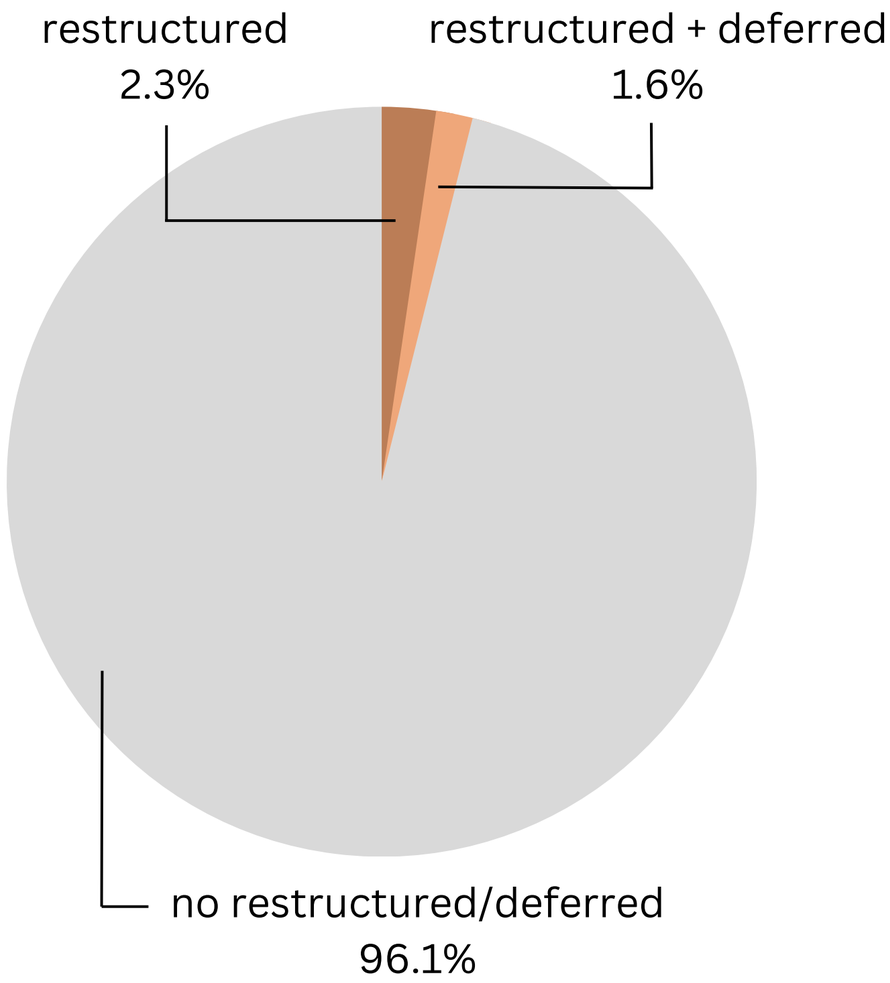


*yellow, informasi, total significant

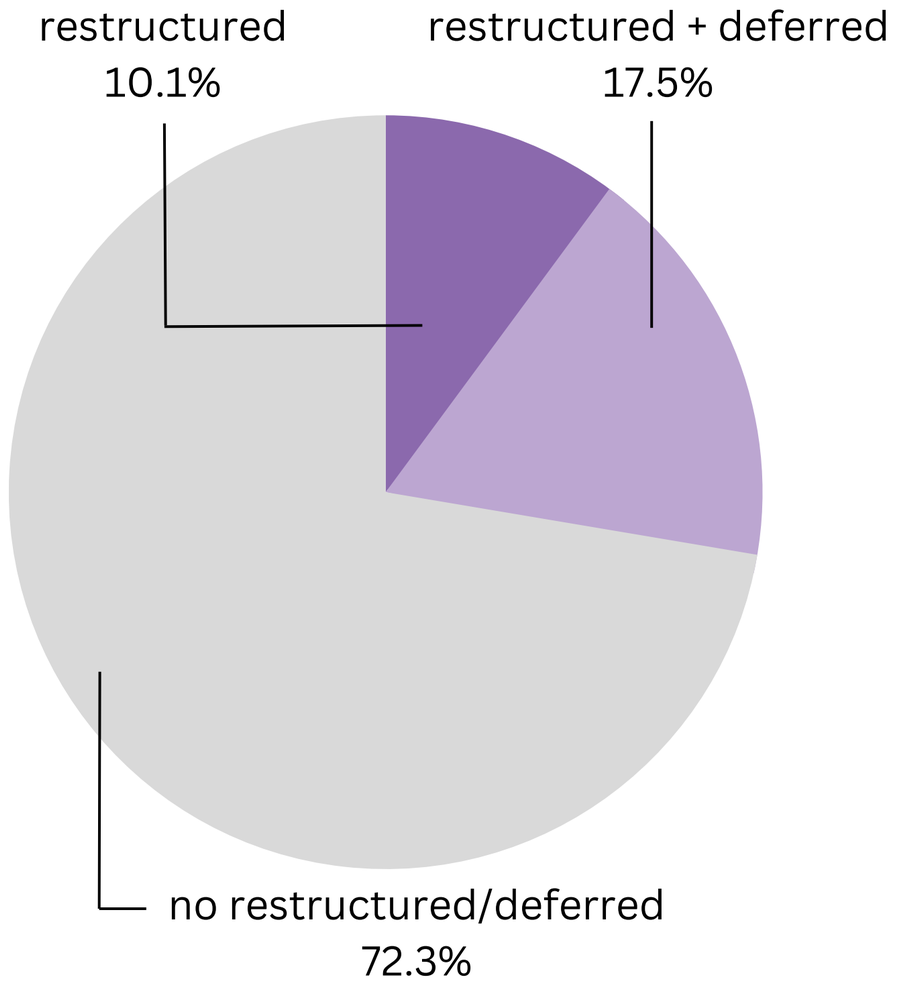
*not significant

Restructured / Deferred Analysis

PL



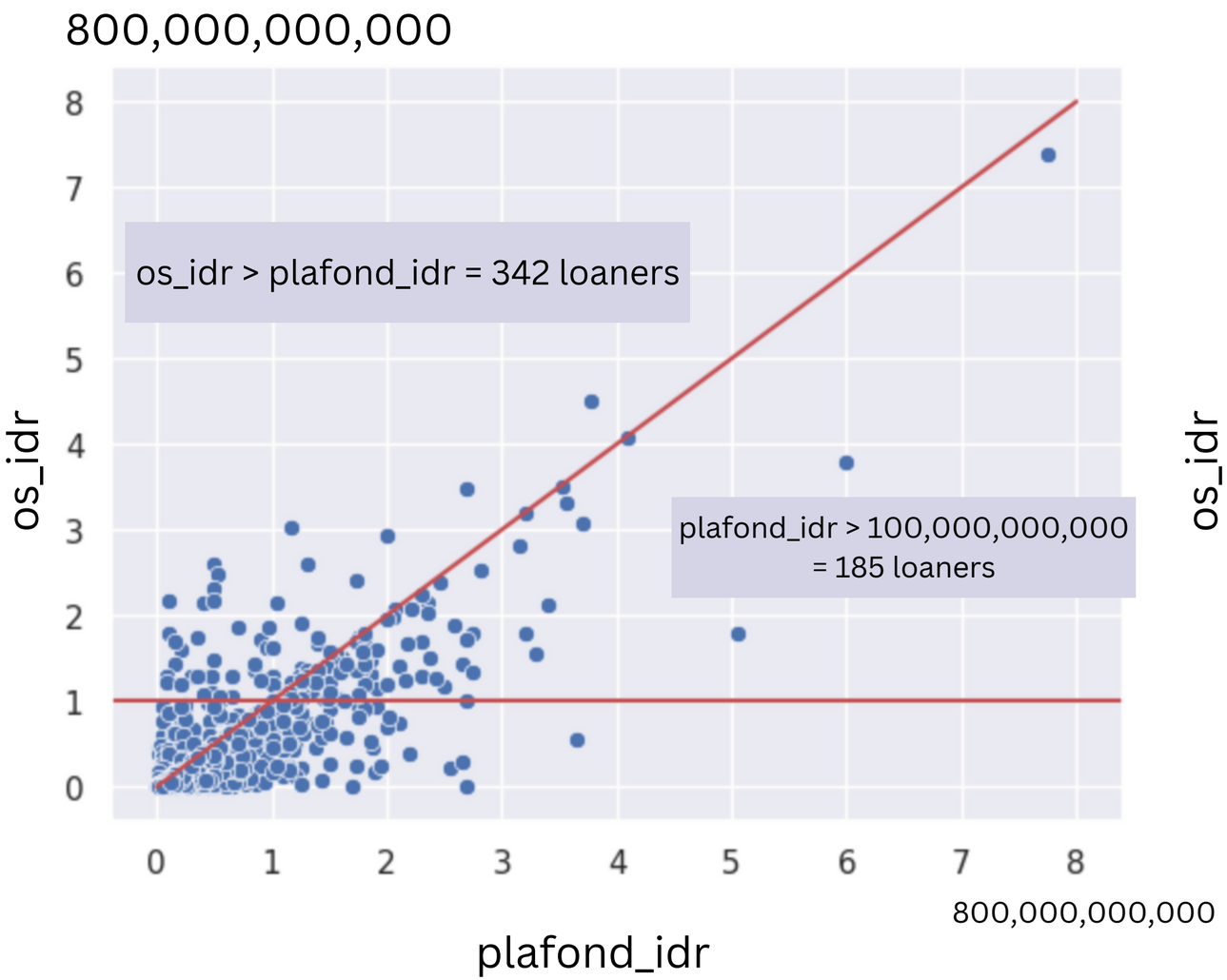
NPL



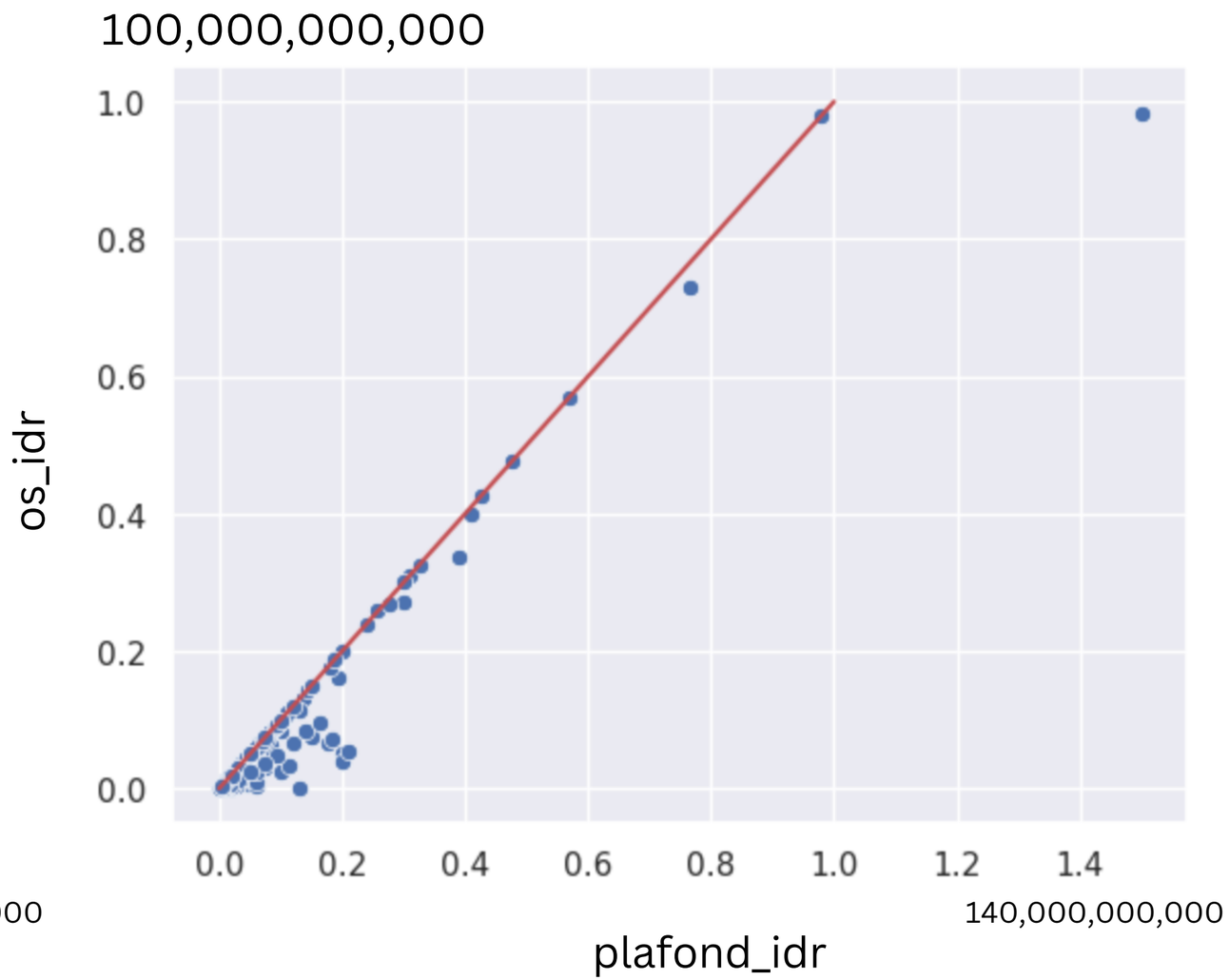
*not significant

Plafond vs OS

PL (36,574 loaners)



NPL (781 loaners)

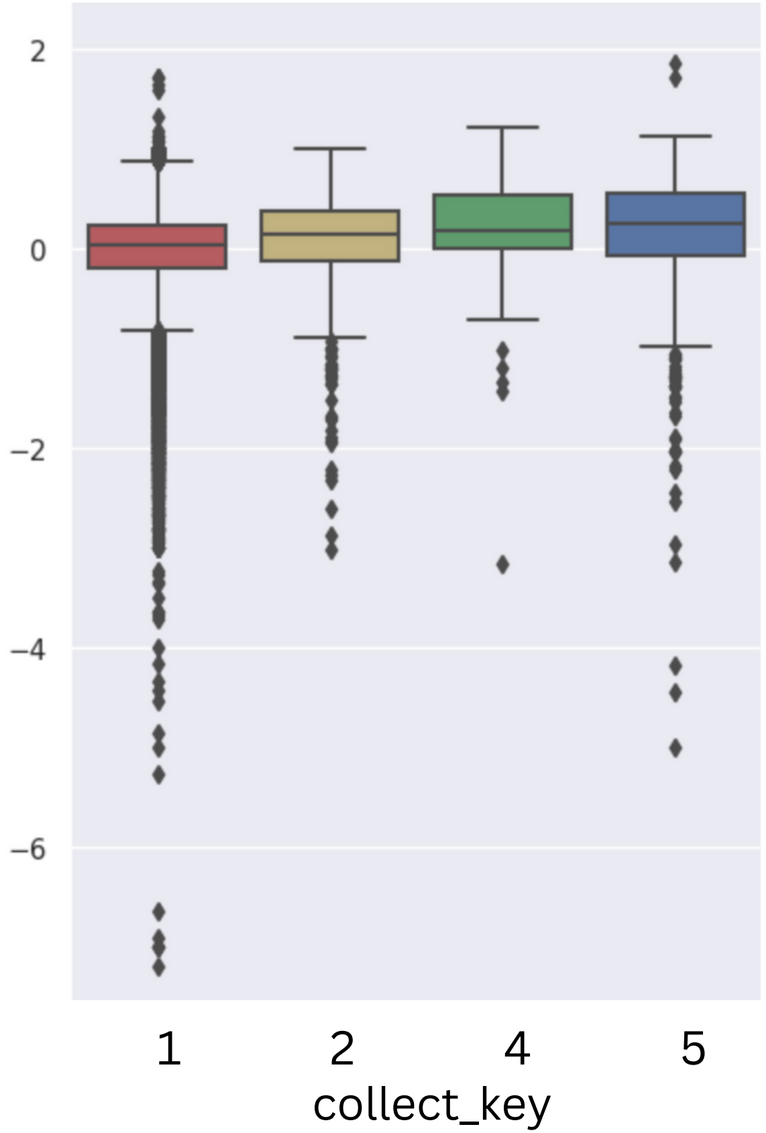


*significant

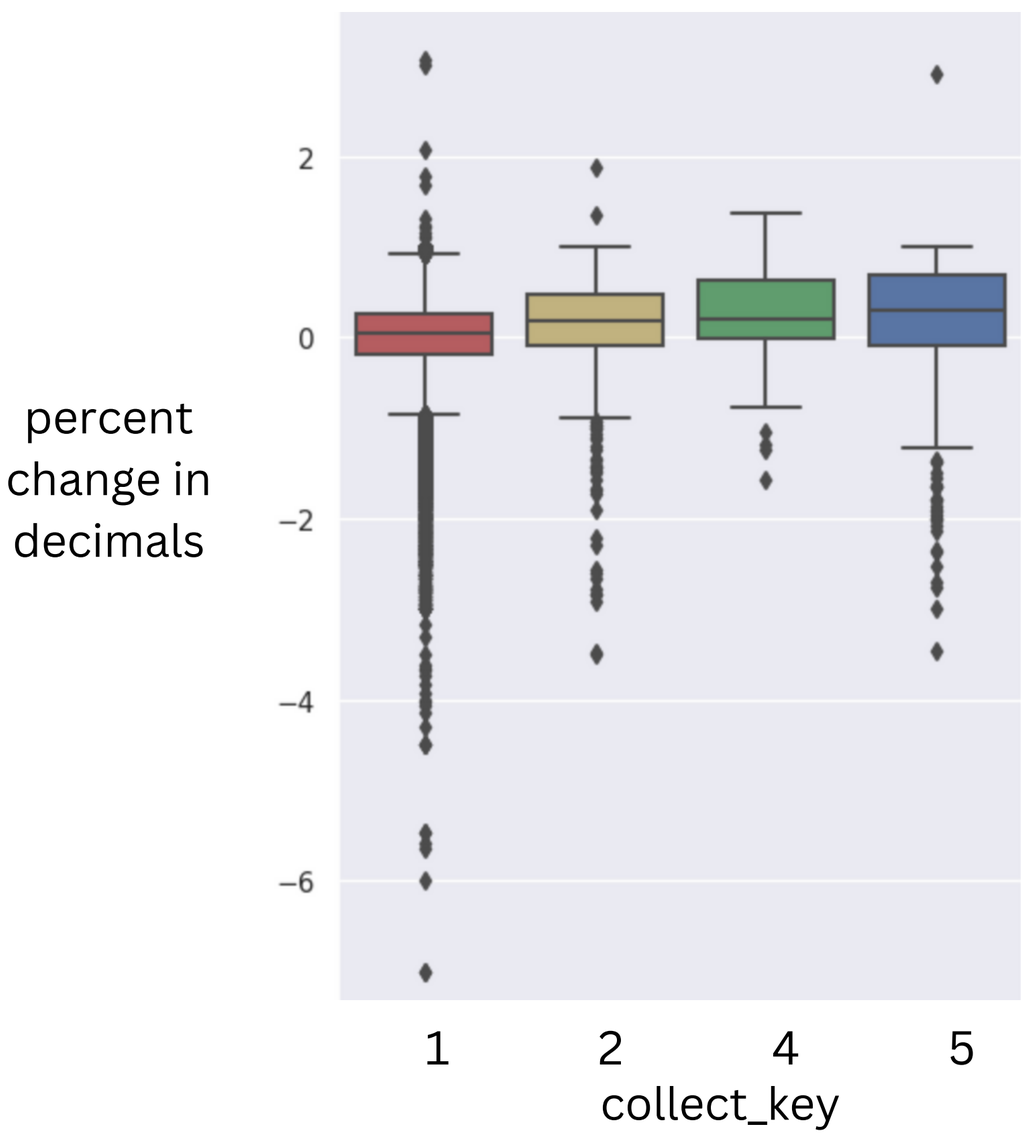
Change in Transaction Behavior
(transaction amount)

Y-axis = $\frac{\text{3MTHS - 12MTHS}}{\text{12MTHS}}$

AMT CR



AMT DB

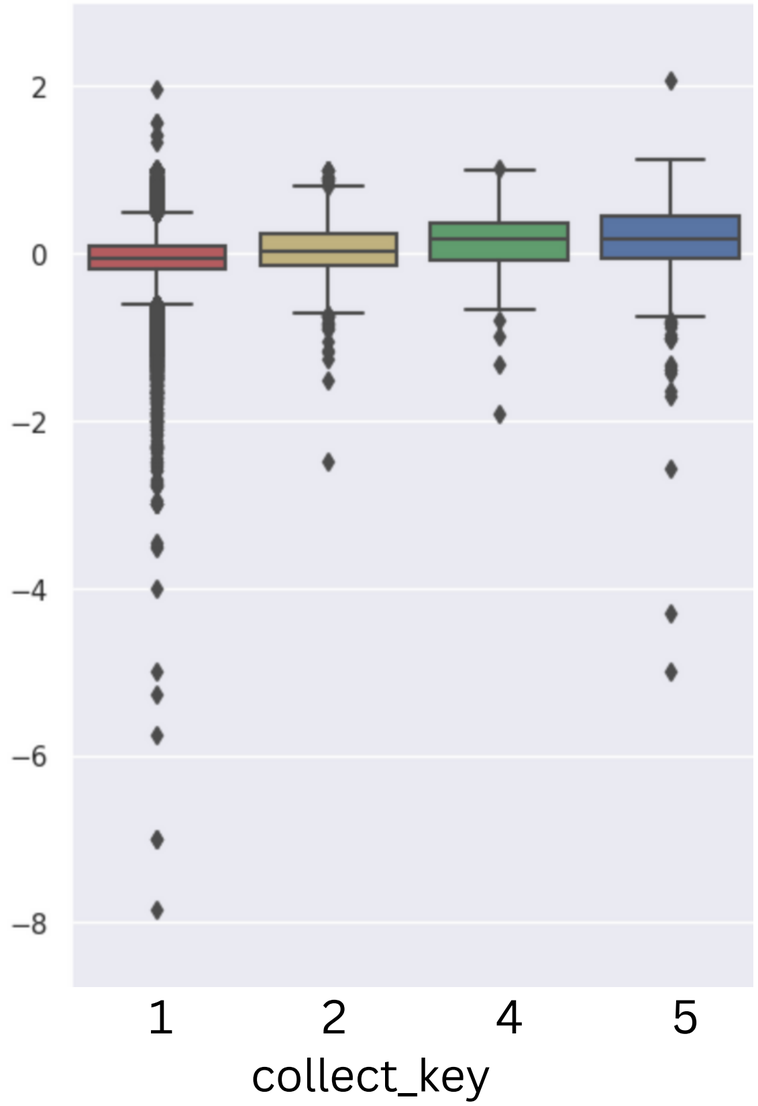


*not significant

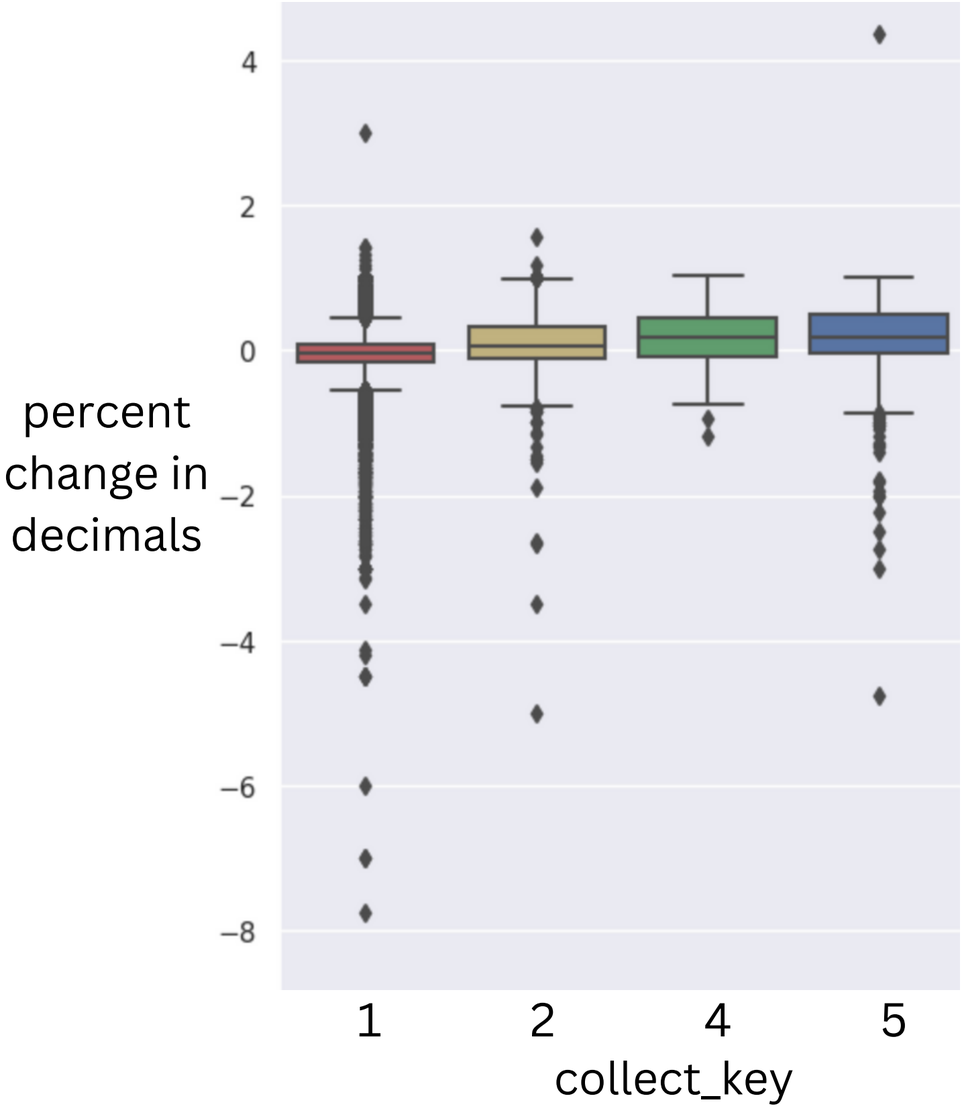
Change in Transaction Behavior
(transaction frequency)

Y-axis = $\frac{\text{3MTHS - 12MTHS}}{\text{12MTHS}}$

FREK CR



FREK DB



12to3CR					
key		1	2	4	5
statistic	count	35948	626	127	654
	mean	-0.02079	0.062522	0.176803	0.13867
	std	0.482124	0.550824	0.555789	0.690041
	min	-7.194555	-3.009541	-3.164119	-5
	25%	-0.189766	-0.131165	-0.000072	-0.070857
	50%	0.031207	0.143814	0.186785	0.244202
	75%	0.23595	0.382989	0.528213	0.552754
	max	1.720222	1	1.20793	1.862868

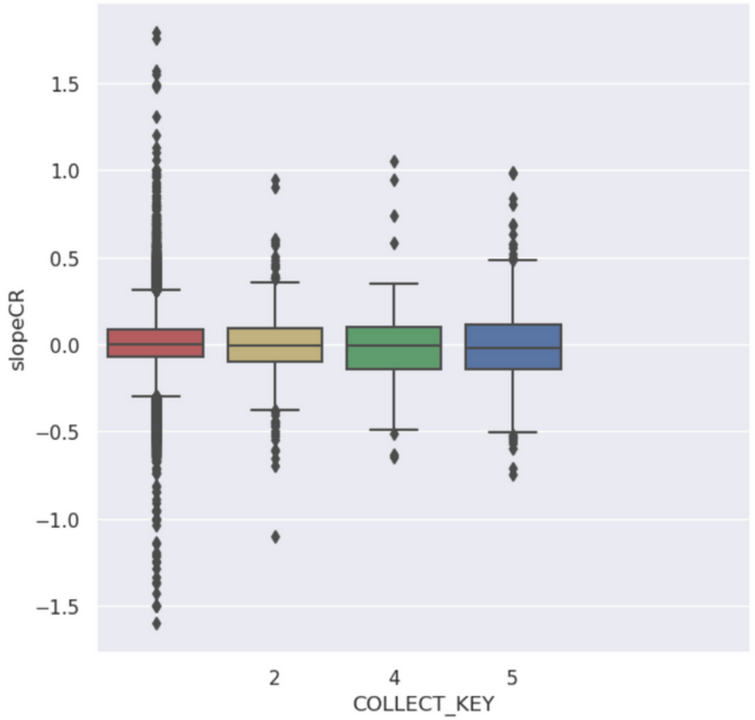
12to3DB					
key		1	2	4	5
statistic	count	35948	626	127	654
	mean	-0.019688	0.109433	0.265653	0.208409
	std	0.548806	0.711982	0.523027	0.749481
	min	-16.2	-7.524561	-1.574157	-3.469467
	25%	-0.187506	-0.088648	-0.003747	-0.09247
	50%	0.041446	0.184205	0.211193	0.300917
	75%	0.254321	0.466906	0.628686	0.687958
	max	3.063641	1.887009	1.367719	7.683775

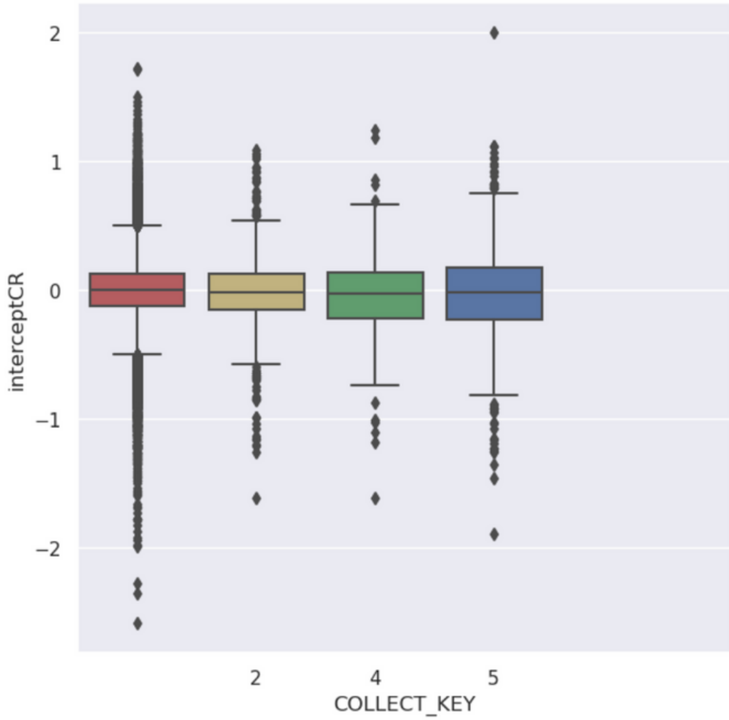
12to3FCR					
key		1	2	4	5
statistic	count	35948	626	127	654
	mean	-0.069162	0.036091	0.129803	0.151356
	std	0.32927	0.370483	0.431738	0.512084
	min	-7.854167	-2.479029	-1.909091	-5
	25%	-0.192575	-0.145089	-0.07009	-0.049405
	50%	-0.054093	0.030865	0.167286	0.168448
	75%	0.078534	0.238092	0.356471	0.4375
	max	1.969388	1	1.017857	2.08

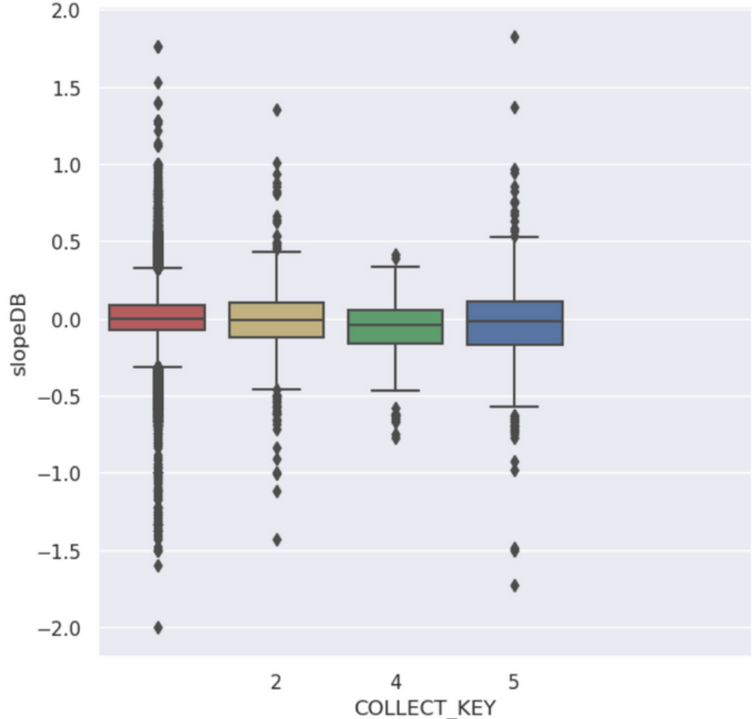
12to3FDB					
key		1	2	4	5
statistic	count	35948	626	127	654
	mean	-0.057772	0.070088	0.203766	0.185612
	std	0.358161	0.526085	0.425269	0.595897
	min	-16.2	-5	-1.185792	-4.76
	25%	-0.171007	-0.123193	-0.084326	-0.043226
	50%	-0.04	0.049745	0.176471	0.185108
	75%	0.079086	0.315508	0.44535	0.505422
	max	3	1.56134	1.027513	4.375

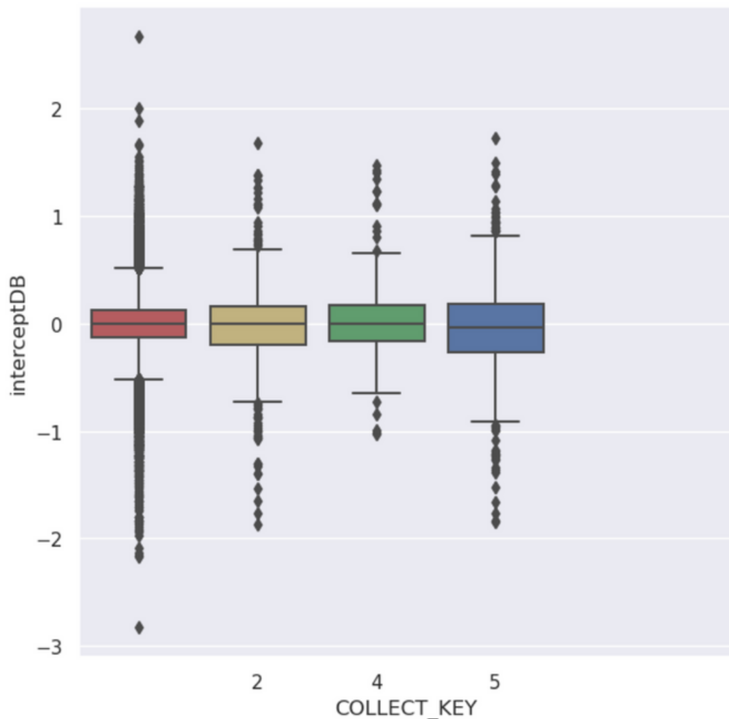
*not significant

Slope + Intercept of amount and frequency (CR + DB)







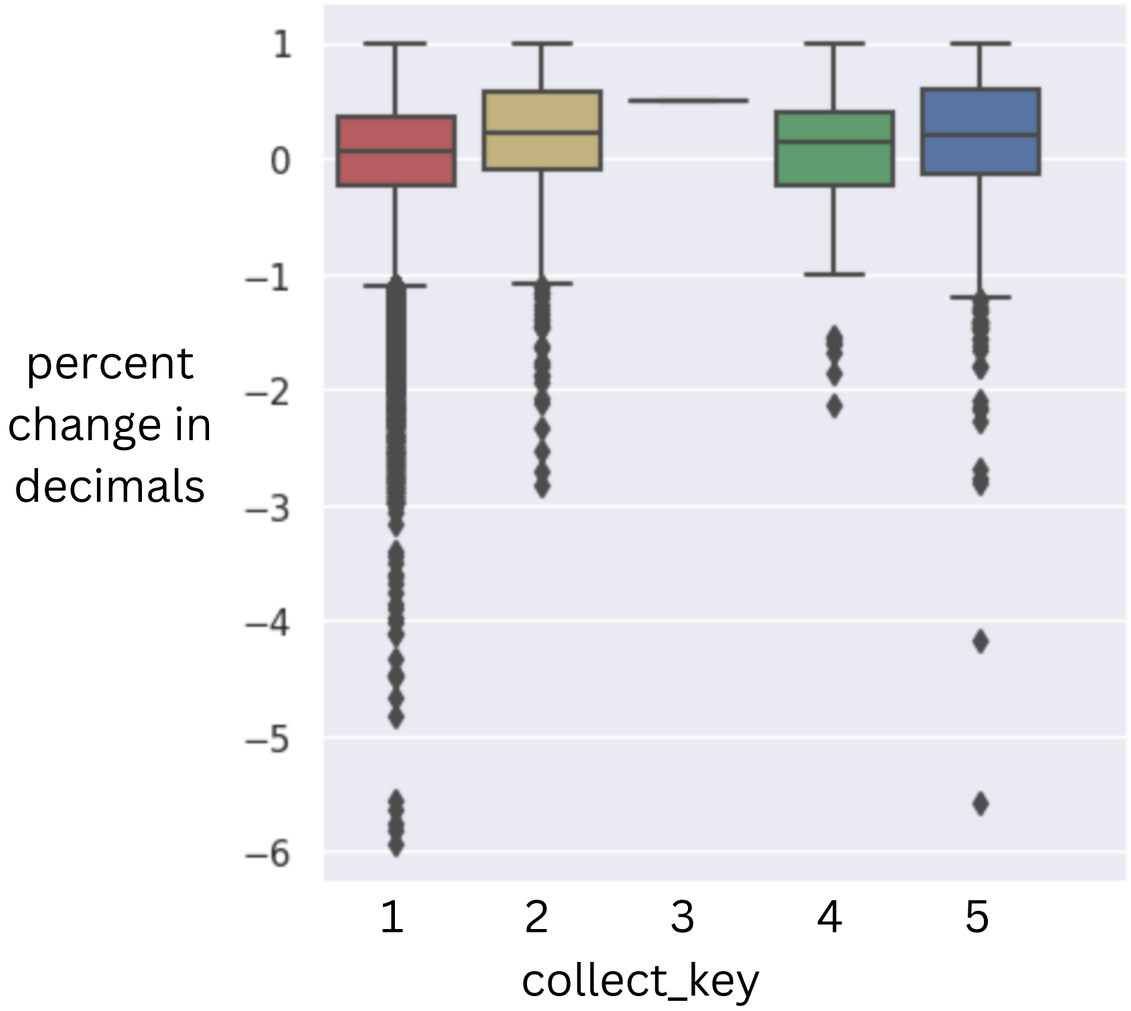


extra

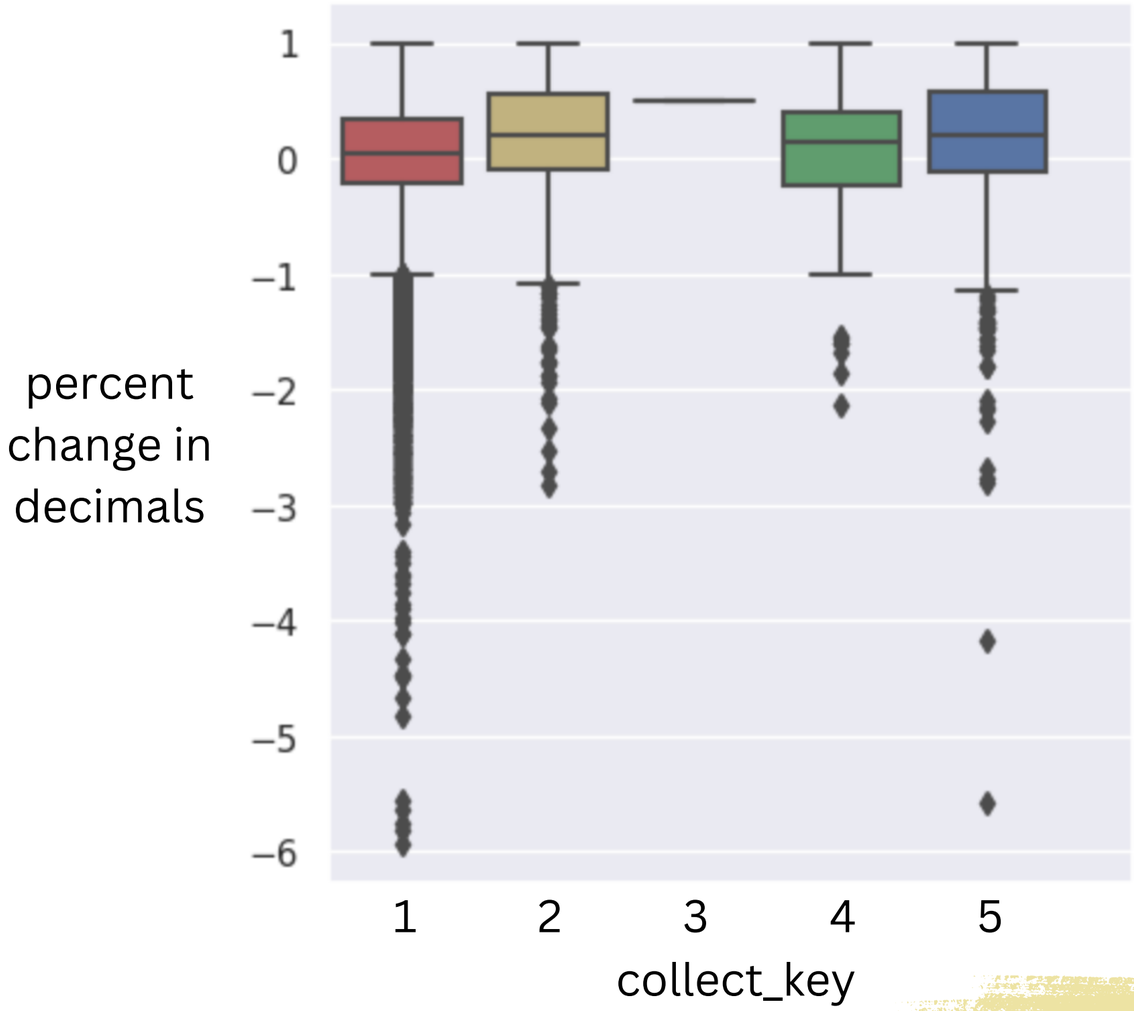
Additional Analysis

$$\text{Y-axis} = \frac{\text{3MTHS - 12MTHS}}{\text{12MTHS}}$$

CASA



DPK

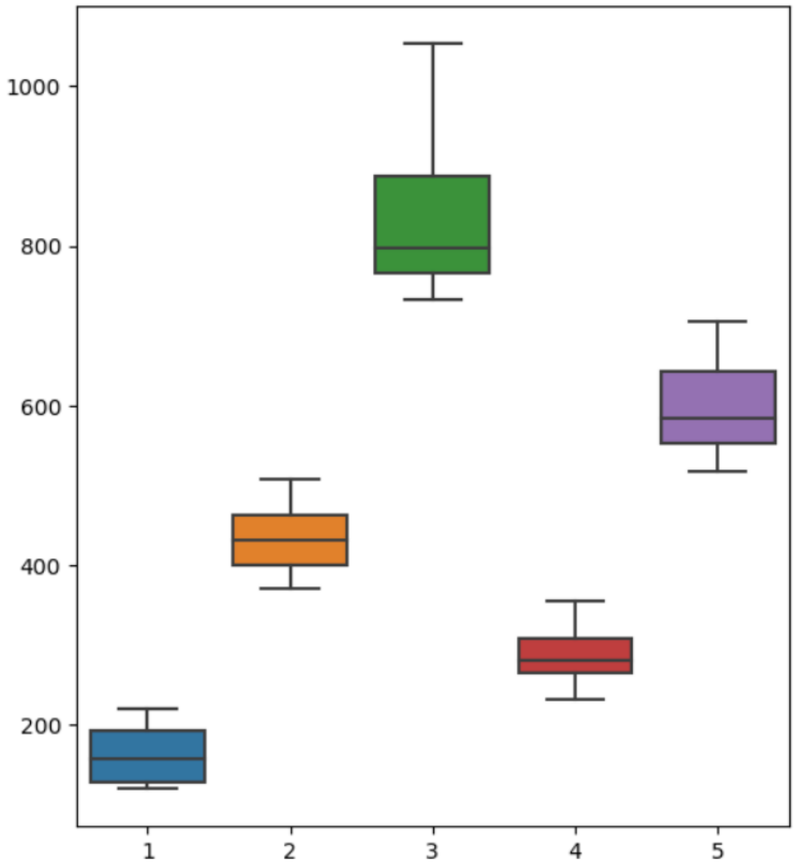


*not significant

extra

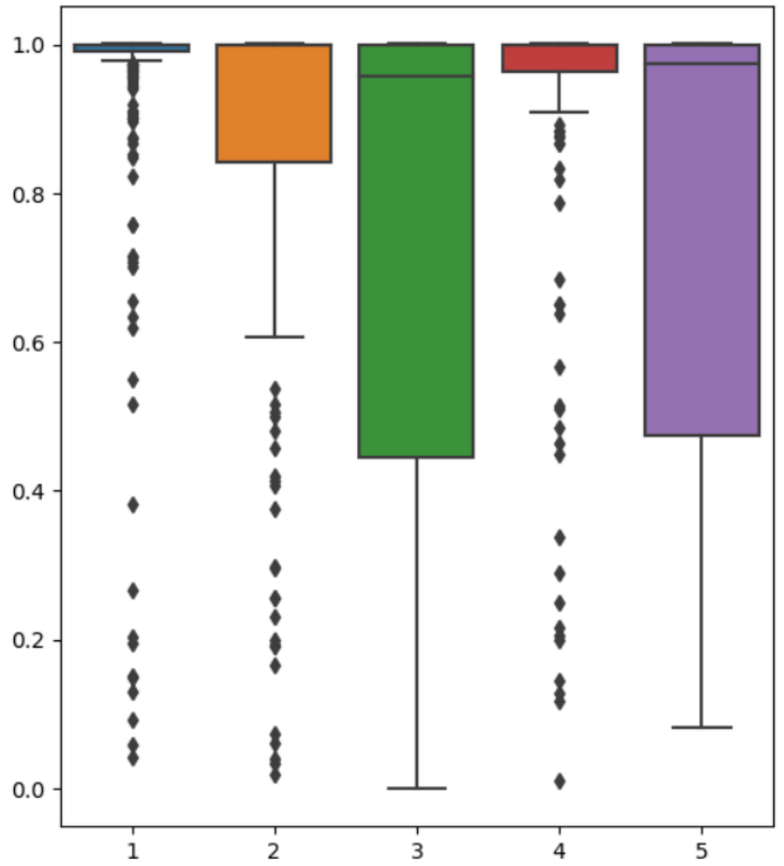
Additional K-means Analysis NPL

Hari Tunggakan



categories made by k-means

Percent Used



categories made by k-means

*loaners who used more borrowed money tend to have less arrears duration

*significant

Variable Summary

ID_DEB
DEBTOR_CATEGORY
PLAFOND_IDR
OS_IDR
HARI_TUNGGAKAN
CUST_TYPE_CD
COLLECT_KEY
FLAG_RESTRU
FLAG_DEFERRED
REDFLAG_YELLOW
REDFLAG_RED
REDFLAG_INFORMASI
FLAG_BLACKLIST
percent_used
total_flags
slopeCR
interceptCR
slopeDB
interceptDB
slopeFCR
interceptFCR
lopeFDB
interceptFDB
12to3CR
12to3DB
12to3FCR
12to3FDB

Logistic Regression

linear_model.LogisticRegression(solver='saga', max_iter=300)

Confusion Matrix (all loaners)

Actual Key \ Predicted Key	1	2	4	5
1	7189	0	0	0
2	129	0	0	0
4	0	0	2	16
5	0	0	2	133

PL: 98.23% accuracy
NPL: 86.92% accuracy