Lending Club Case Study

ML C54 June 2023 EPGP

Rajaram Karanth

Problem Statement

A consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Aim of this case study is to use EDA over the dataset containing the complete loan data for all loans issued through the time period 2007 to 2011 and assess the driving factors behind the loan default. such that the company can utilise this knowledge for its portfolio and risk assessment.

In this analysis, EDA is used to understand how consumer attributes and loan attributes influence the tendency of default.

Analysis Approach

Data Cleaning:

The dataset was checked for missing or invalid values in the rows and columns and fixed

mths_since_last_delinq column has 25682 NA values. Replaced with zero

mths_since_last_record column has 36931 NA values. Replaced with zero

desc column has 13149 blank values. However, those have corresponding valid purpose column values

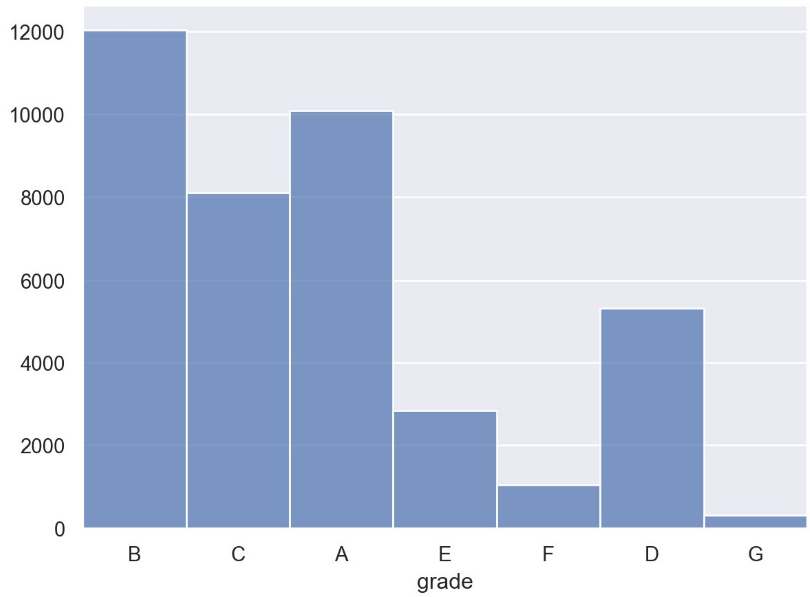
next_pymnt_d column has 38577 blank records

The dataset was checked for unwanted columns and deleted. Below are the columns deleted that only had values as either NA or 0.

mths_since_last_major_derog, annual_inc_joint, dti_joint , verification_status_joint , acc_now_delinq , tot_coll_amt , tot_cur_bal , open_acc_6m , open_il_6m , open_il_12m , open_il_24m , mths_since_rcnt_il, total_bal_il, il_util , open_rv_12m , open_rv_24m, max_bal_bc , all_util , total_rev_hi_lim , inq_fi , total_cu_tl, inq_last_12m , acc_open_past_24mths , avg_cur_bal , bc_open_to_buy , bc_util , chargeoff_within_12_mths, delinq_amnt , mo_sin_old_il_acct , mo_sin_old_rev_tl_op , mo_sin_rcnt_rev_tl_op , mo_sin_rcnt_tl , mort_acc , mths_since_recent_bc , mths_since_recent_bc_dlq , mths_since_recent_inq , mths_since_recent_revol_delinq, num_accts_ever_120_pd , num_actv_bc_tl , num_actv_rev_tl , num_bc_sats , num_bc_tl , num_il_tl , num_op_rev_tl , num_rev_accts , num_rev_tl_bal_gt_0 , num_sats , num_tl_120dpd_2m , num_tl_30dpd , num_tl_90g_dpd_24m , num_tl_op_past_12m , pct_tl_nvr_dlq , percent_bc_gt_75, tot_hi_cred_lim , total_bal_ex_mort , total_bc_limit , total_il_high_credit_limit

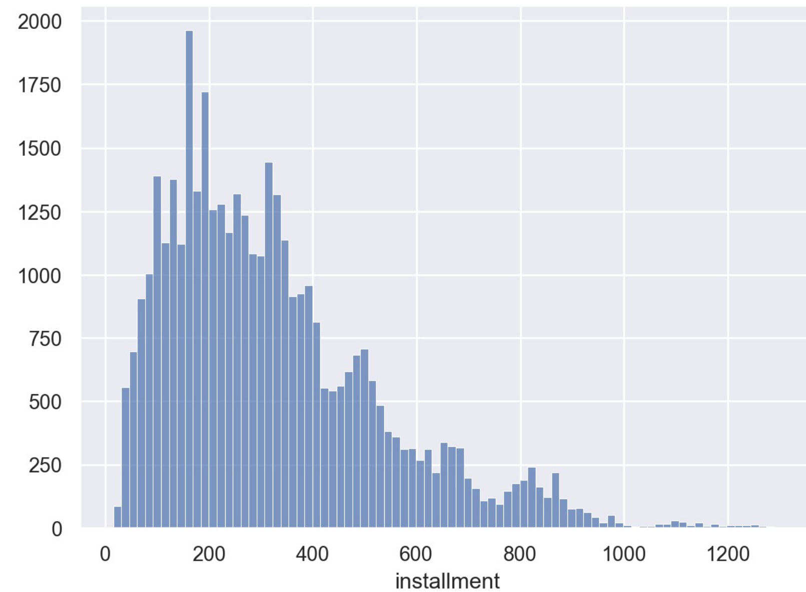
Analysis Approach – key characteristics of the loans issued

Univariate Analysis #1 - number of loans by grade



Highest number of loans disbursed are of grade B, followed by grade A

Univariate Analysis #2 - loan installments preferred



Highest bins installments opted are between 0 - 200 months and 200 - 400 months.

Analysis Approach – key characteristics of the loans issued

Univariate Analysis #3 - Purpose for which most of the loans are taken

debt_consolidation	18641
credit_card	5130
other	3993
home_improvement	2976
major_purchase	2187
small_business	1828
car	1549
wedding	947
medical	693
moving	583
vacation	381
house	381
educational	325
renewable_energy	103

Maximum number of loans taken are for the purpose of debt consolidation. More chances of default in this category as this option just provides breathing time for otherwise a possible delinquent loan in the previous institution

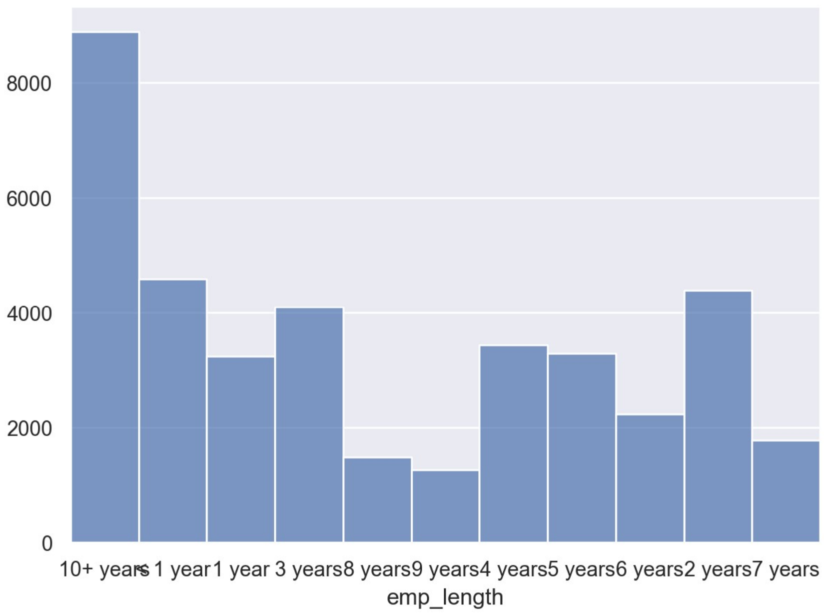
Univariate Analysis #4 - States where maximum number of loans disbursed

CA	7099	NC	788	WV	177
NY	3812	CT	751	HI	174
FL	2866	MI	720	NH	171
TX	2727	MO	686	DE	114
NJ	1850	MN	615	MT	85
IL	1525	NV	497	WY	83
PA	1517	SC	472	AK	80
VA	1407	WI	460	SD	64
GA	1398	AL	452	VT	54
MA	1340	OR	451	MS	19
OH	1223	LA	436	TN	17
MD	1049	KY	325	IN	9
AZ	879	OK	299	ID	6
WA	840	KS	271	IA	5
CO	792	UT	258	NE	5
		AR	245	ME	3
		DC	214		
		RI	198		
		NM	189		

Most number of loans are issued in big cities

Analysis Approach – key characteristics of the loans issued

Univariate Analysis #5 - Number of loans by length of employment



Maximum number of loans are availed by customers who are employed for 10+ years. The next highest bin is of the customers who are employed between 1 and 2 years

Analysis Approach – key characteristics of the loans issued

Bivariate Analysis #1 - Correlation between the delinquency and the number of open credit lines

	delinq_2yrs	open_acc
delinq_2yrs	1.000000	0.011656
open_acc	0.011656	1.000000

Number of open credit lines has a positive correlation with delinquency although not very strong. However more number of open credit lines increases the risk profile of the customer.

Bivariate Analysis #2 - Correlation between the delinquency and the dti

	delinq_2yrs	dti
delinq_2yrs	1.000000	-0.034452
dti	-0.034452	1.000000

In the given dataset, dti is found to be lower in the delinquent cases.
Lower dti is a strong indicator of a possible default as it indicates lower debt repayment capacity of the borrower.

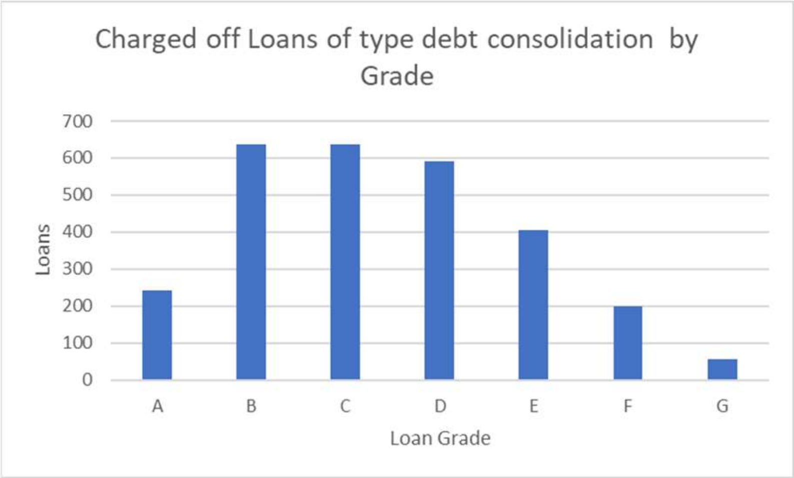
Analysis Approach – key characteristics of the loans issued

Bivariate Analysis #3 - Correlation between the delinquency and the principal outstanding

	mths_since_last_delinq	out_prncp
mths_since_last_delinq	1.000000	-0.012297
out_prncp	-0.012297	1.000000

Months since delinquency is inversely correlated with the repayment. recent delinquencies indicates chances of default

Bivariate Analysis #4 - Charged off Loans of type debt consolidation by Grade

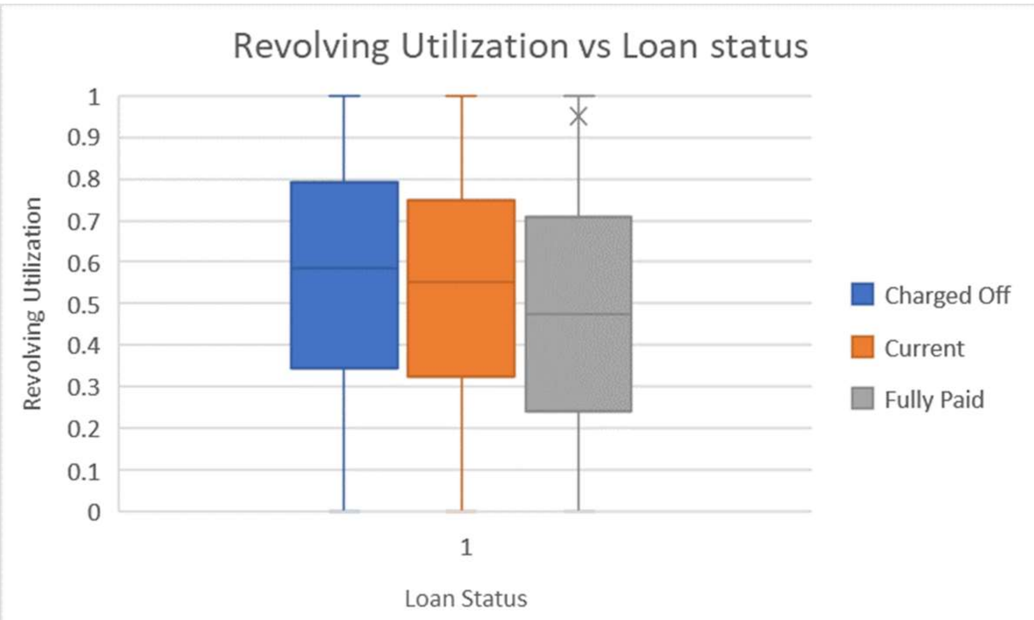


Plotted in xls

Maximum number of charged off loans for the purpose of debt consolidation are of grade B. More caution to be exercised on the loan applications with these attributes

Analysis Approach – key characteristics of the loans issued

Bivariate Analysis #5 - Revolving Utilization vs Loan status



Plotted in xls

- Revolving utilization that indicates risk profile based on the balances on the revolving accounts as against the available credit.
- The Median value of the Revolving utilization in case of the Charged off loans was close to 60%.
- The Median value of the Revolving utilization in case of the Fully paid loans was under 50%.
- The currently running loans have the Median value for Revolving utilization is very near to the Median of Charged off loans.
- This is a strong indicator of possible default.

Thank You