Lending Card Case Study

- Submitted by Vijay Khanna

Approach to Solving the Lending Card Case Study

- Research on Lending Domain – Internet Research, Research Papers
- Understand KPI’s and Business Drivers in Lending Domain
- Understand Key Patterns in Data to identify potential Loan Default
- Identify Columns in data which can provide insights. Initial Explore in Excel Pivots
- Explore Data using Python Visualization
- Create Insights from provided data
- Summarize key Insights

Lending Club Case Study : Business Objectives

- Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss)
- Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders.
- Identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- Understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default

Applicant is **likely to repay the loan**

Applicant is **not likely to repay the loan**

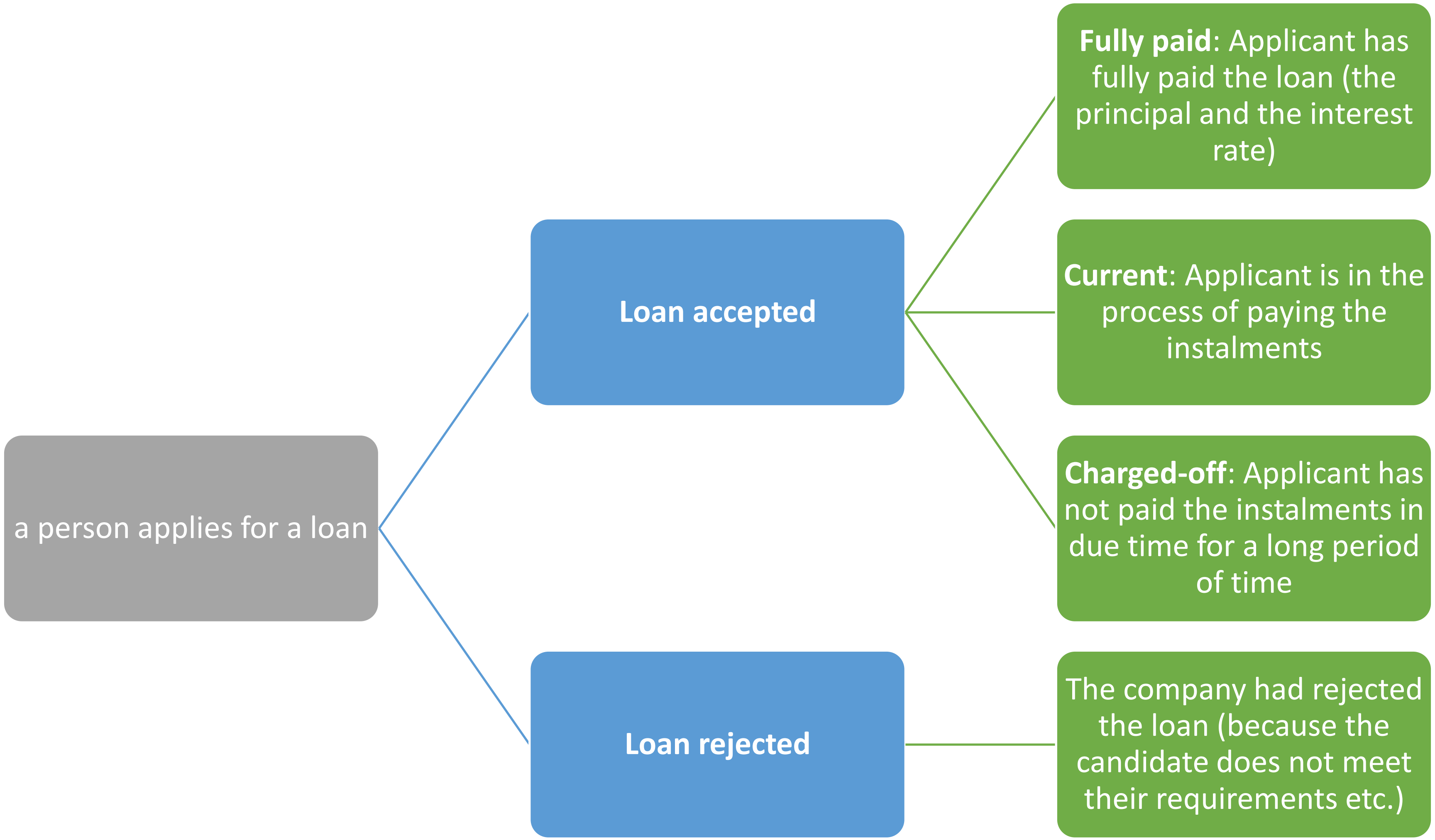
approving the loan results in a **loss of business** to the company

approving the loan may lead to a **financial loss** for the company

Balancing Loan Disbursal

Business Loss vs Profits

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:



Understanding the Domain of **consumer finance company**.
References

Commercial Banks and Consumer Instalment Credit . John M. Chapman NBER

- https://www.nber.org/system/files/chapters/c4732/c4732.pdf

Credit Risk Management of Consumer Finance Based on Big Data

- https://www.hindawi.com/journals/misy/2021/8189255/
- https://downloads.hindawi.com/journals/misy/2021/8189255.pdf

Investigation on Consumer Finance Risk Management – Case Study

- https://www.atlantis-press.com/article/125908355.pdf

Drivers of Performance in Unsecured Personal Loans

- https://hl.com/media/rn2hx2od/drivers-of-performance-in-unsecured-personal-loans.pdf

Domain Understanding from the Reference Reading Materials **"Investigation on Consumer Finance Risk Management – Case Study"**

Key Insights and Lessons

- Consumer Finance Growth: The consumer finance market in China, particularly in Zhejiang Province, has experienced rapid growth, significantly influencing regional economic development and improving residents' lives.
- Challenges in Risk Management: The study highlights challenges in consumer finance risk management, such as information asymmetry, high system costs, underdeveloped financial technology, insufficient supervision, lack of professional talent, and limited product variety.
- Risk Management Stages: Effective risk management in consumer finance involves stages like pre-finance, in-finance, and post-finance, focusing on various risks including credit, market, operational, and liquidity risks.
- Employee Skillset and Education: The majority of consumer finance field employees have 1 to 5 years of experience, with a significant proportion holding undergraduate degrees. Key skills include understanding macroeconomic policy, trend analysis, and risk management.
- Target Demographics: A significant portion of consumer finance customers are aged between 25 and 35, indicating high consumption needs. Younger consumers (18-25) are also emerging as an important demographic, characterized by high consumption demand but relatively low and unstable income.

Domain Understanding from the Reference Reading Materials
“Investigation on Consumer Finance Risk – Case Study”

Key Patterns for Loan Default Identification

- **Credit History and Scores**: Analyze borrowers' past credit behavior and scores.
- **Income Stability**: Assess the stability and adequacy of the borrower's income.
- **Debt-to-Income Ratios:** Evaluate borrowers' existing debts against their income.
- **Employment History:** Consider the consistency and length of employment.
- **Age and Demographics:** Understand the risk profile based on age group and other demographic factors.
- **Loan Purpose and Amount**: Scrutinize the purpose of the loan and the amount requested.
- **Current Financial Behavior**: Monitor current spending patterns and financial commitments.

Domain Understanding from the Reference Reading Materials "Commercial Banks and Consumer Instalment Credit– **Case Study**"

- Credit Risk Assessment: Credit officers evaluate the likelihood of loan repayment based on probabilities and intuition, considering an applicant's willingness and ability to repay.
  - Characteristics Influencing Credit Risk:
  - Personal characteristics like age, sex, marital status, dependents, and residence stability.
  - Vocational aspects, including occupation type, employer's business nature, and employment tenure.
  - Financial capacity shown by income, assets, and debts.
- Variability Among Banks: The study acknowledges the variability in experiences among different banks, influenced by regional, customer class, and policy differences.
- Statistical Methods: The Chi-square test was used for statistical significance, ensuring the findings aren't merely due to chance.
- Applicability and Limitations: The findings highlight trends and probabilities rather than certainties, emphasizing they shouldn't be applied mechanically without considering special circumstances. The study also notes that the analysis does not represent potential borrowers who were denied loans or never sought them.
- Influence of Age and Gender: Age is a significant factor, with older borrowers (over 50 years) posing a lower risk than younger ones (21-25 years). Gender also plays a role, with women generally posing a lower risk than men. However, these factors are interrelated and influenced by other aspects like occupation and income.
- Occupational and Industrial Risks: Professional occupations tend to have a lower credit risk compared to wage earners. Industries like utilities, professional services, and public service are associated with lower risk, while building trades and transportation show higher risk indices.
- Tenure of Employment: Longer employment tenure correlates with lower credit risk, highlighting stability as a key factor in assessing creditworthiness.
- Income and Loan Experience: The relationship between income levels and loan experience is not straightforward, with no distinct trend observed across different income brackets. This suggests that banks' lending policies might already factor in the borrower's income adequacy effectively.

Domain Understanding from the Reference Reading Materials **"**Drivers of Performance in Unsecured Personal Loans **– Case Study"**

Key Factors Affecting Loan Performance

- Credit Score (FICO): A higher credit score is associated with lower risk. The study reveals an inverse, nonlinear relationship between credit score and expected loss. Notably, a decrease in FICO score results in a disproportionately higher increase in loss expectation compared to an equivalent increase in score.
- Debt-to-Income Ratio (DTI): Higher DTI indicates greater credit risk. The study finds the relationship between DTI and expected loss to be almost symmetrical; changes in DTI have a roughly similar impact on loss expectations in both directions.
- Homeownership: Owning a home, particularly with a mortgage, correlates with lower expected loss. This finding suggests that homeowners might generally be more responsible credit users.
- Loan Purpose: Different loan purposes exhibit varying levels of risk. For example, loans for medical expenses or moving have higher expected losses compared to those for weddings, which tend to have lower loss expectations. This dispersion in expected loss based on loan purpose underscores the importance of considering the intended use of funds in risk assessment.
- Macroeconomic Factors - Unemployment Rate: Fluctuations in the unemployment rate significantly affect loan performance. An increase in unemployment leads to a higher expected loss, reflecting the borrower's diminished ability to repay the loan. Interestingly, loans to borrowers with lower FICO scores are more sensitive to changes in unemployment rates.

Balancing Loan Disbursal - Business Growth vs Financial Profits

Balancing Loan Disbursal

Business Growth

vs

Financial Profits

Scenario 1: **Potential Loss of Business**.

Risk: Not approving loans to creditworthy individuals.

Strategy: Use advanced credit scoring models that accurately predict creditworthiness. Incorporate a broad range of data, including credit history, employment status, and even behavioral data.

Scenario 2: **Risk of Financial Loss**.

Risk: Approving loans to individuals who may default.

Strategy: Implement stringent risk assessment procedures. Consider factors like debt-to-income ratio, credit score thresholds, and past financial behavior.

Identifying Patterns for Default Risk

1. **Credit History Analysis**: Look for patterns like late payments, high credit utilization, or defaults on previous loans.
2. **Behavioral Indicators**: Unstable employment history, frequent changes in residence, or erratic financial behavior.
3. **Debt-to-Income Ratio**: High ratios may indicate that the borrower is over-leveraged.
4. **Use of Alternative Data**: Include non-traditional data like utility bill payments, rent payment history, or even social media behavior patterns.

Key Parameters which can provide Insights to Driving Factors behind Loan Default

1. **FICO Scores (fico_range_low, fico_range_high)**: Lower FICO scores typically indicate higher risk of default.
2. **Debt-to-Income Ratio (dti)**: A high DTI suggests that a borrower may have difficulty managing loan repayments.
3. **Delinquency History (delinq_2yrs, mths_since_last_delinq)**: Frequency and recency of past delinquencies can be strong predictors of future defaults.
4. **Utilization Ratios (revol_util, all_util)**: High utilization rates might indicate financial strain or over-reliance on credit.
5. **Credit History Length (earliest_cr_line, mo_sin_old_rev_tl_op)**: A shorter credit history may not provide enough information to gauge a borrower's creditworthiness accurately.
6. **Number of Recent Credit Inquiries (inq_last_6mths, inq_last_12m)**: Multiple recent inquiries can suggest a search for new credit due to financial stress.
7. **Current Loan Amounts (loan_amnt, funded_amnt_inv)**: Larger loan amounts might carry a higher risk of default.
8. **Number of Active Credit Lines (open_acc, num_actv_bc_tl)**: A high number of active credit lines may indicate potential overextension.
9. **Employment Stability (emp_length, emp_title)**: Short or inconsistent employment history might affect a borrower's ability to repay.
10. **Loan Purpose and Characteristics (purpose, term, int_rate)**: Certain loan purposes or characteristics (like high interest rates or longer terms) may be associated with higher default rates.

FICO Score Calculation for given DataSet

1.Payment History (35%)

- Fields: **delinq_2yrs**, **mths_since_last_delinq**, **chargeoff_within_12_mths**, **collections_12_mths_ex_med**, **num_accts_ever_120_pd**, **num_tl_30dpd**, **num_tl_90g_dpd_24m**, **pub_rec**, **pub_rec_bankruptcies**
- These fields reflect the borrower's history of making payments on time.

2.Amounts Owed (30%)

- Fields: **revol_bal**, **revol_util**, **total_bal_ex_mort**, **total_bal_il**, **total_rev_hi_lim**, **il_util**, **bc_util**, **num_rev_tl_bal_gt_0**
- This category considers the total amount of debt and the utilization ratio, which is the balance relative to credit limits.

3.Length of Credit History (15%)

- Fields: **earliest_cr_line**, **mo_sin_old_il_acct**, **mo_sin_old_rev_tl_op**
- The time since accounts were opened and the time since account activity are used to assess credit history length.

4.New Credit (10%)

- Fields: **inq_last_6mths**, **inq_last_12m**, **num_tl_op_past_12m**, **mths_since_recent_inq**
- This includes the number of recently opened accounts and recent credit inquiries.

5.Credit Mix (10%)

- Fields: **open_acc**, **num_actv_bc_tl**, **num_actv_rev_tl**, **num_bc_tl**, **num_il_tl**, **num_op_rev_tl**, **num_rev_accts**, **num_sats**, **num_tl_120dpd_2m**

Initial Exploratory Insights from Data. Using Excel Pivots

- Total Rows in Data : 39717
- Charged off Loans : 5627

Row Labels	Count of loan_status
Charged Off	5627
Current	1140
Fully Paid	32950
(blank)	
Grand Total	**39717**

Average of loan_amnt	Column Labels													
Row Labels	**< 1 year**	**1 year**	**10+ years**	**2 years**	**3 years**	**4 years**	**5 years**	**6 years**	**7 years**	**8 years**	**9 years**	**n/a**	**(blank)**	**Grand Total**
Charged Off	10,308	10,663	14,594	10,738	11,405	11,522	12,044	12,564	12,895	13,051	13,136	[illegible],796		12,104
Current	14,829	15,153	19,209	16,359	15,646	15,886	17,564	15,847	16,552	16,664	16,162	13,692		17,054
Fully Paid	9,450	9,955	12,476	10,031	10,513	10,730	10,958	11,176	11,324	11,526	11,721	8,312		10,866
(blank)														
Grand Total	**9,658**	**10,168**	**13,090**	**10,262**	**10,738**	**10,977**	**11,287**	**11,495**	**11,739**	**11,888**	**12,012**	**8,625**		**11,219**

- People with longer employment length & Average Loan Amount seem to have more Charge off

Count of loan_amnt	Column Labels													
Row Labels	**< 1 year**	**1 year**	**10+ years**	**2 years**	**3 years**	**4 years**	**5 years**	**6 years**	**7 years**	**8 years**	**9 years**	**n/a**	**(blank)**	**Grand Total**
Charged Off	639	456	1,331	567	555	462	458	307	263	203	158	228		5,627
Current	75	71	391	97	83	94	88	61	62	44	32	42		1,140
Fully Paid	3,869	2,713	7,157	3,724	3,457	2,880	2,736	1,861	1,448	1,232	1,068	805		32,950
(blank)														
Grand Total	**4,583**	**3,240**	**8,879**	**4,388**	**4,095**	**3,436**	**3,282**	**2,229**	**1,773**	**1,479**	**1,258**	**1,075**		**39,717**

- People with employment length >10 years have higher count/instances of Charge off

Initial Exploratory Insights from Data. Using Excel Pivots

Count of member_id	Column Labels								
Row Labels	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**(blank)**	**Grand Total**
Charged Off	602	1,425	1,347	1,118	715	319	101		5,627
Current	40	345	264	222	179	73	17		1,140
Fully Paid	9,443	10,250	6,487	3,967	1,948	657	198		32,950
(blank)									
Grand Total	**10,085**	**12,020**	**8,098**	**5,307**	**2,842**	**1,049**	**316**		**39,717**

Impact of Grade on ChargeOff

Average of annual_inc	Column Labels						
Row Labels	**MORTGAGE**	**NONE**	**OTHER**	**OWN**	**RENT**	**(blank)**	**Grand Total**
Charged Off	75,243		63,433	53,908	53,246		62,427
Current	85,676			57,972	63,289		75,431
Fully Paid	84,253	80,733	73,082	59,760	[illegible],960		69,863
(blank)							
Grand Total	**83,117**	**80,733**	**71,310**	**58,863**	**57,370**		**68,969**

Average Annual Income vs Home Ownership.

Shows lower average Income has direct co-relation on Charge-Off

To Do in Python

- Bin as per Loan Amount, and check bins with most Defaults.
- Home Ownership

Initial Exploratory Insights from Data. Using Excel Pivots

Initial two Numbers of Zip Codes with highest Charge Offs. Check for Media Reports of Gangs colluding with officials for loan frauds in these zip codes / States

Row Labels	Count of Charged Off
44	10
29	10
91	10
01	10
95	10
06	10
61	10
07	10
27	10
08	10
32	10
11	10
43	10
12	10
48	10
15	10
80	10
17	10
22	10
30	10

States with highest ChargeOffs

State	Count of Charged Off
CA	1125
FL	504
NY	495
TX	316
NJ	278
GA	215
IL	197
PA	180
VA	177
MD	162
MA	159
OH	155
WA	127
AZ	123
MO	114
NC	114
NV	108
MI	103

High **debt to income** ratio shows a strong correlation to defaults

Row Labels	Average of dti
Charged Off	14.00
Current	14.75
Fully Paid	13.15

Initial Exploratory Insights from Data. Using Excel Pivots

Count of id	Column Labels				
Row Labels	Charged Off	Current	Fully Paid	(blank)	Grand Total
car	160	50	1339		1549
credit_card	542	103	4485		5130
debt_consolidation	2767	586	15288		18641
educational	56		269		325
home_improvement	347	101	2528		2976
house	59	14	308		381
major_purchase	222	37	1928		2187
medical	106	12	575		693
moving	92	7	484		583
other	633	128	3232		3993
renewable_energy	19	1	83		103
small_business	475	74	1279		1828
vacation	53	6	322		381
wedding	96	21	830		947
(blank)					
Grand Total	**5627**	**1140**	**32950**		**39717**

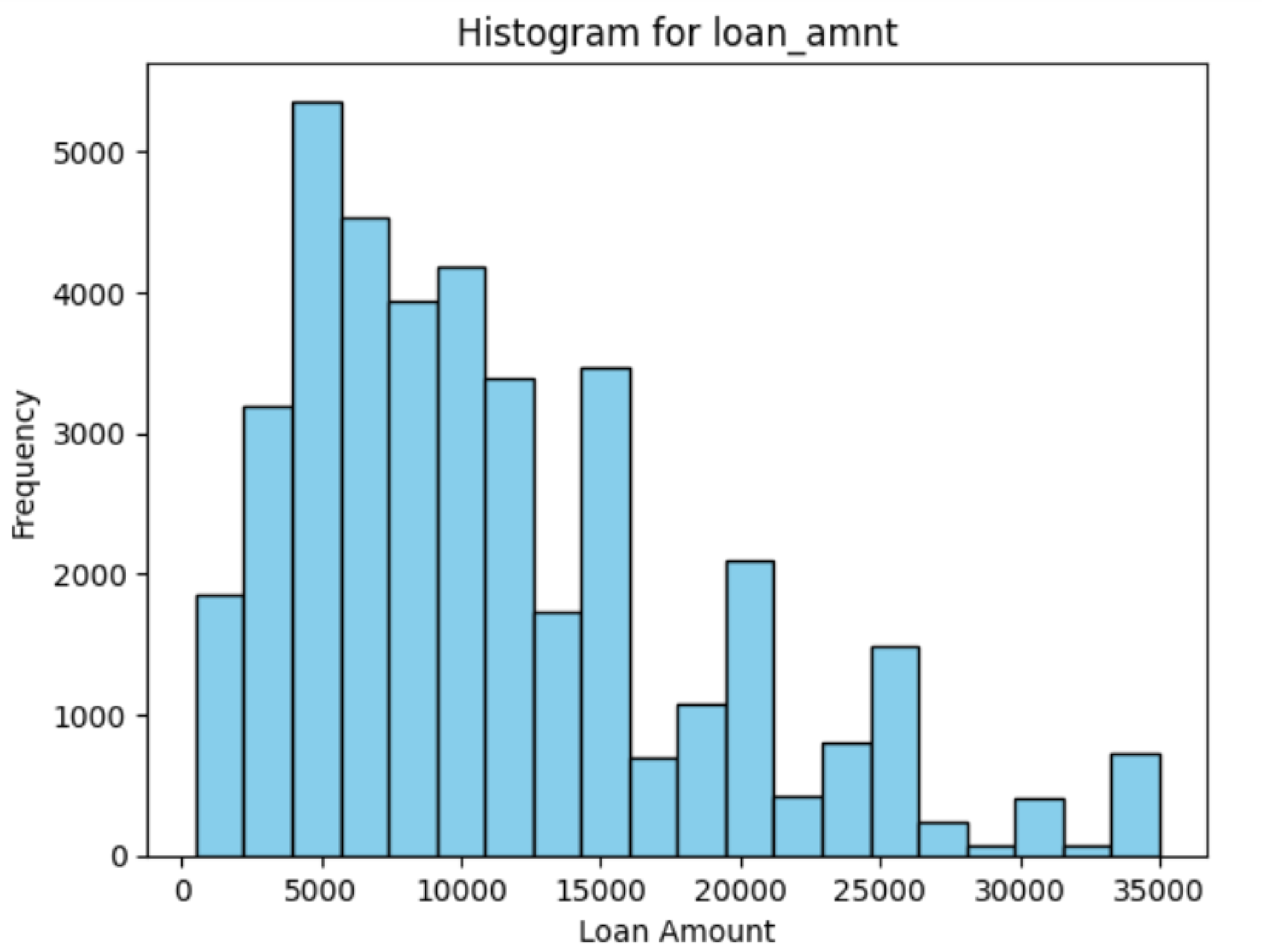
Looks like Loans for Debt_Consolidation, credit_card payments and Small_business have most defaults

Exploratory Insights from Data. Using Python Graphs

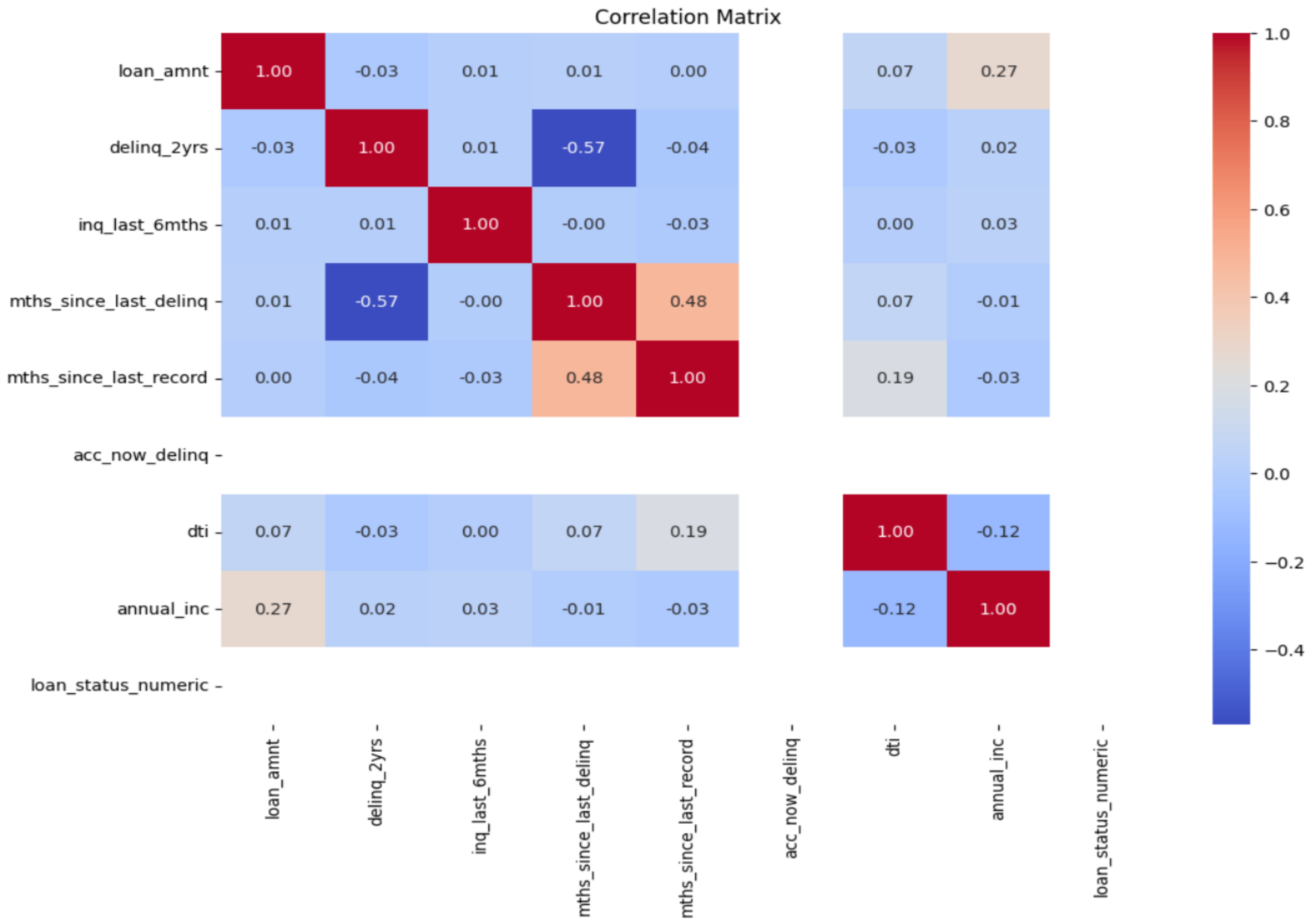
- Based on Excel Analysis, and Understanding the Fields, these are the Fields that can show potential of Co-Relation in the data

Fields that can show CoRelation		
acc_now_delinq	last_pymnt_amnt	sub_grade
addr_state	last_pymnt_d	term
annual_inc	loan_amnt	title
application_type	loan_status	total_acc
chargeoff_within_12_mths	mths_since_last_delinq	total_pymnt
delinq_2yrs	mths_since_last_record	verification_status
dti	mths_since_recent_bc_dlq	zip_code
dti_joint	mths_since_recent_inq	zip_code (First 3 Letters)
earliest_cr_line	num_accts_ever_120_pd	inq_last_6mths
emp_length	num_bc_tl	installment
emp_title	num_tl_120dpd_2m	issue_d
funded_amnt_inv	num_tl_op_past_12m	pymnt_plan
grade	open_acc	revol_bal
home_ownership	pub_rec	revol_util
inq_fi	pub_rec_bankruptcies	
inq_last_12m	purpose	

Exploratory Insights from Data. Using Python Analysis



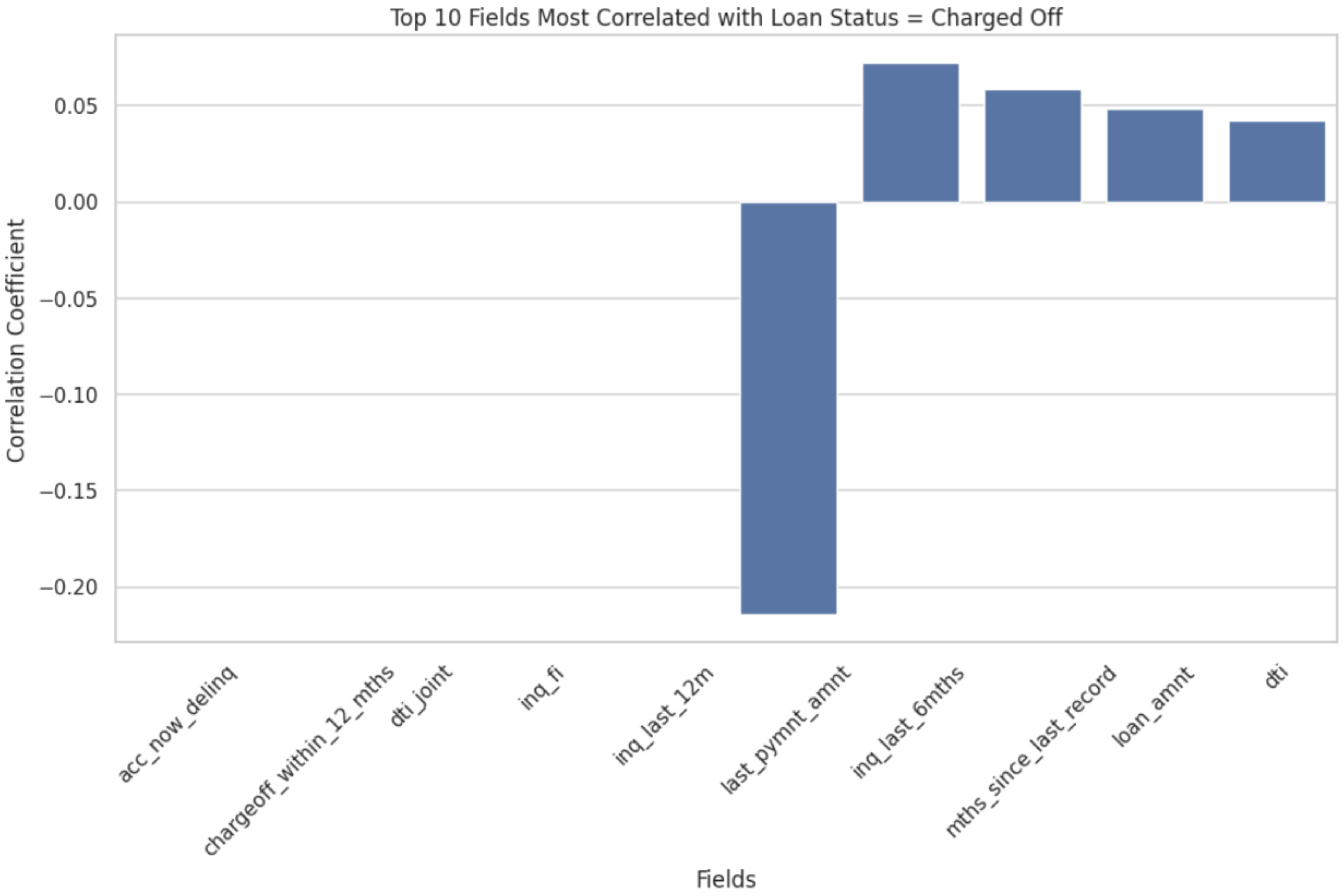
Exploratory Insights from Data. Using Python Analysis



There seems to be a correlation between the fields loan_amount, annual_inc, Dti, Months_since_last_record

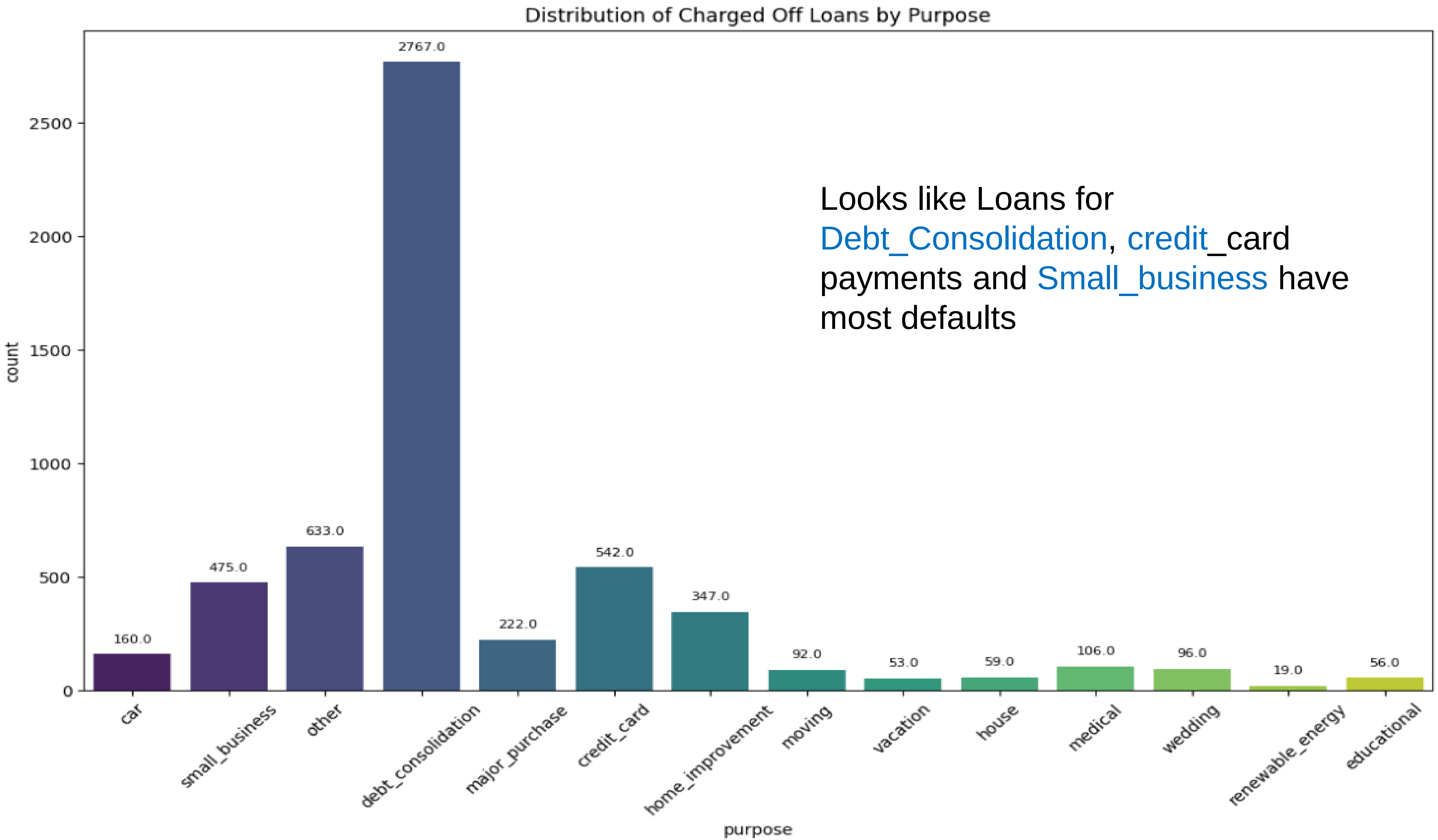
These may be able to provide early insights into upcoming default.

Exploratory Insights from Data. Using Python Analysis



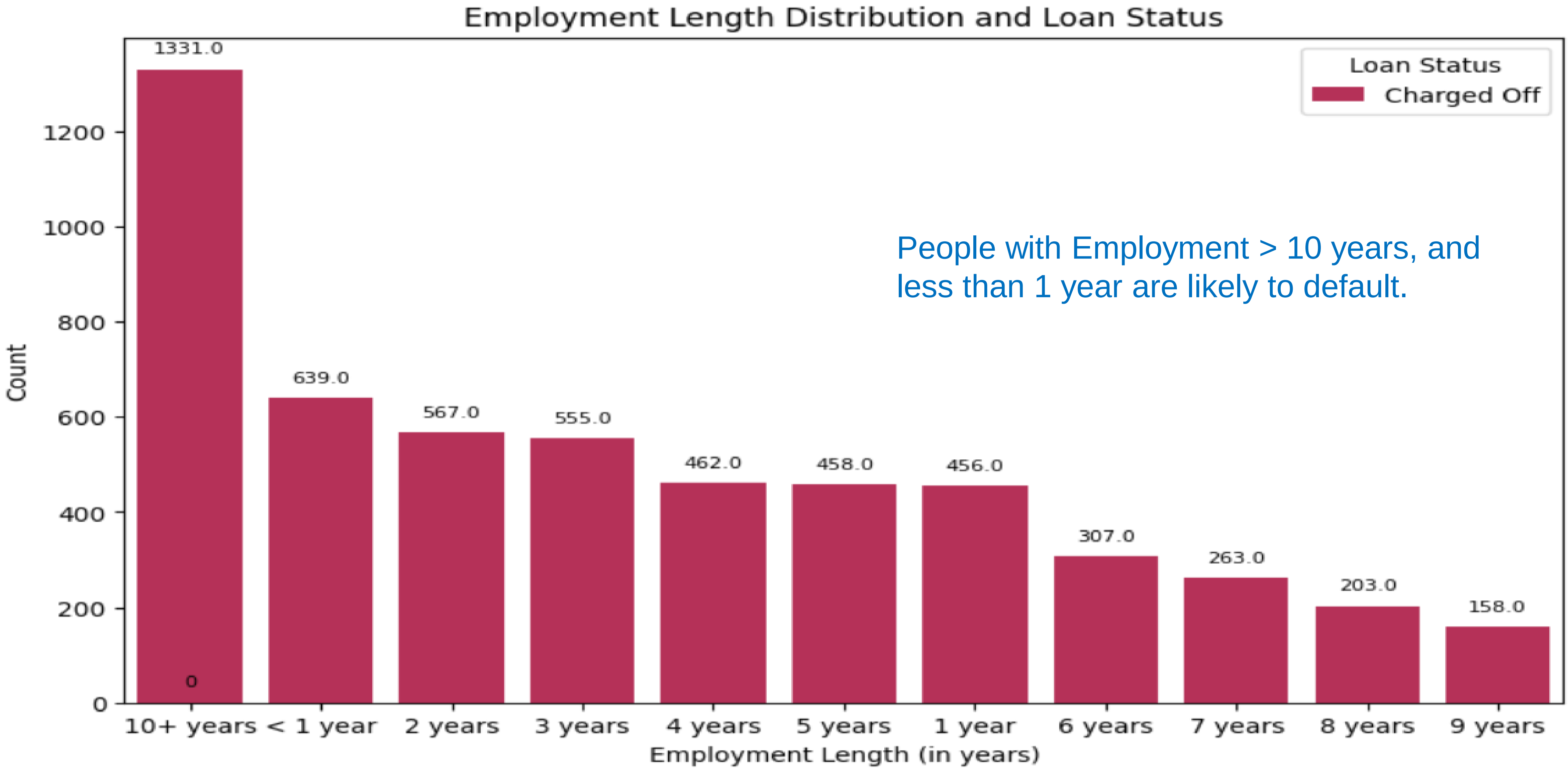
These are the fields which seem to be most corelated to Charged Off status.
Providing early signals to the company.

Exploratory Insights from Data. Using Python Analysis



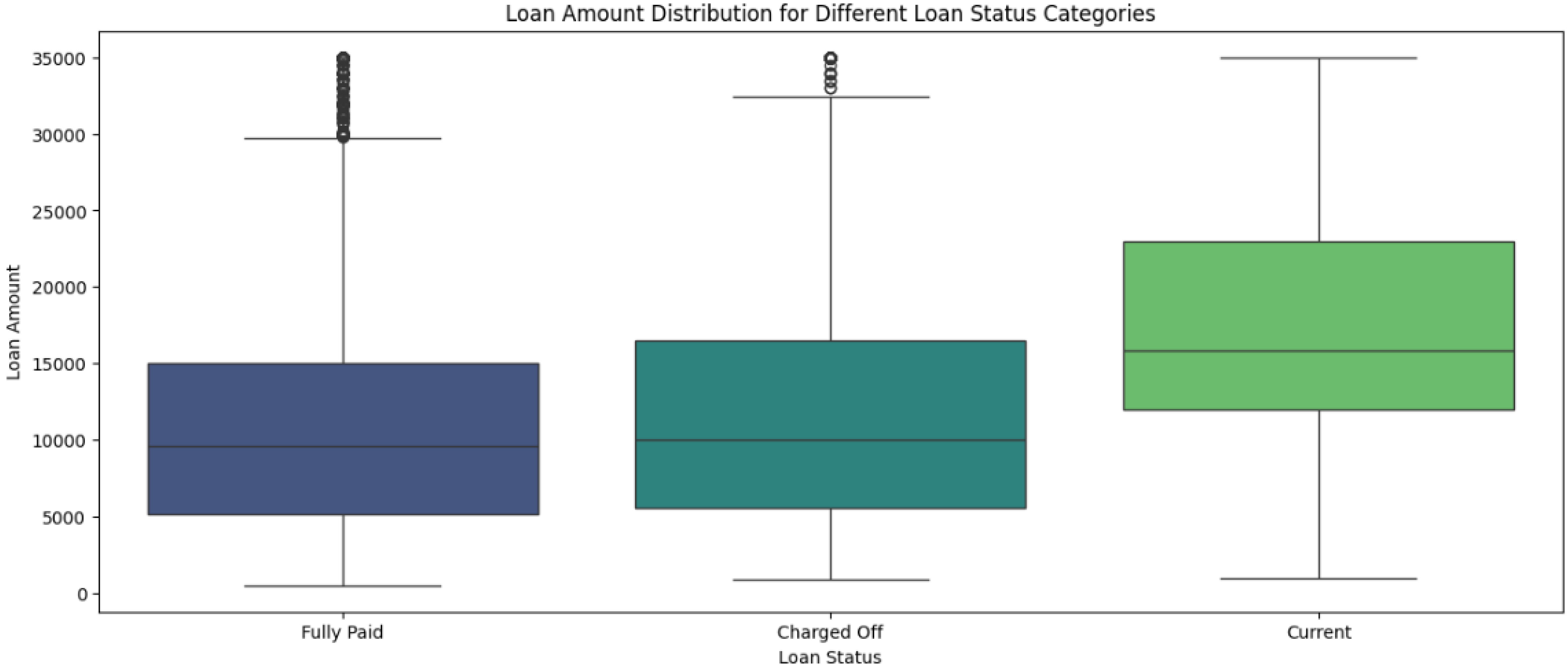
Looks like Loans for Debt_Consolidation, credit_card payments and Small_business have most defaults

Exploratory Insights from Data. Using Python Analysis



People with Employment > 10 years, and less than 1 year are likely to default.

Exploratory Insights from Data. Using Python Analysis



The Box Plot shows that the Average Loan Size for Charged Off segment is Higher than Fully Paid. This also leads to the insight that the Current average is much higher, which could point to upcoming loan defaults, and the company should proactively work on working with those customers and monitor the triger signals better.