

Lending Club Case Study

Date of Submission: 24th July 2024

Contributors:

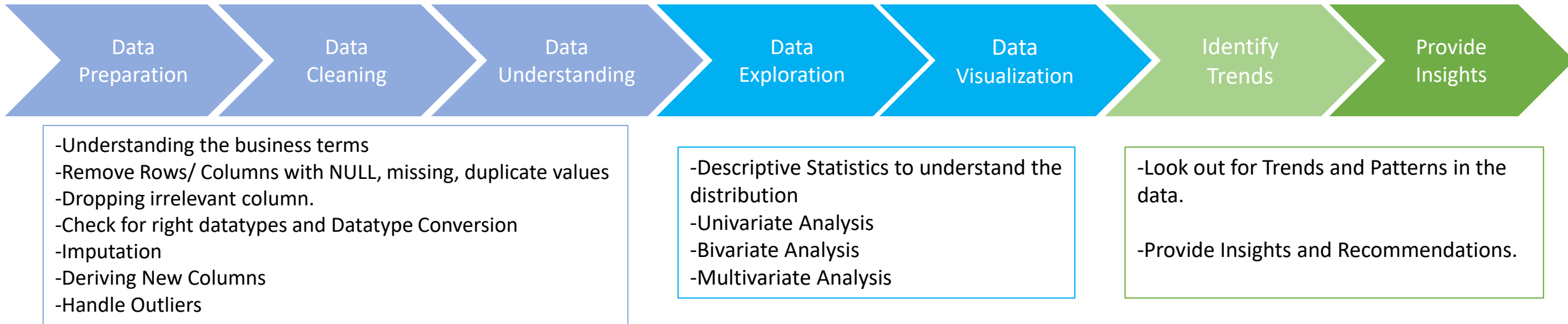
1. Preetam Oswal
2. Ramya Pulipaka

Problem Statement , Objectives & Approach

Problem:

Lending Club needs to identify the driving factors behind the loan default

Approach:



Objectives:

Reduce
Financial
Losses

Enhance
Loan
Approvals

Identify
Driver
Variables

Data Description and Preparation for EDA

Dataset Overview

- The lending club dataset has 111 columns and 39.7k rows

Data Cleaning and Preparation

- 100% Missing Values: **58 columns**
- Irrelevant Columns: **32 columns** (constants, zeros, unique identifiers)
- Post-Approval Columns: Columns that are irrelevant to pre-approval analysis

Data Type Corrections

- Ensured accurate data types for proper analysis

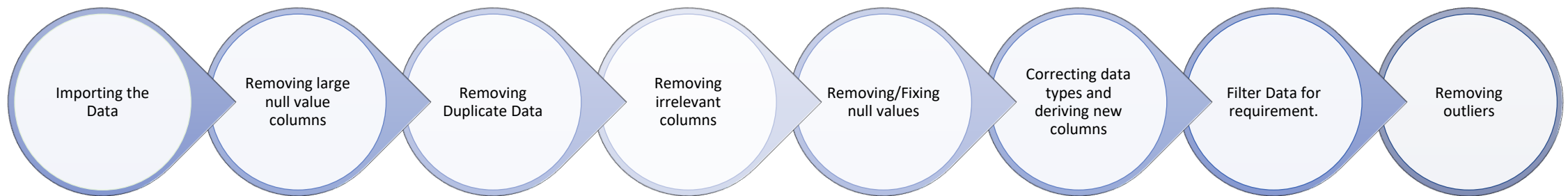
Derived Columns:

- Credit Utilization Ratio: Calculated from existing data
- Other Relevant Metrics: Based on requirements

Data Conversion: Converted continuous variables into categorical ranges for improved analysis and interpretation

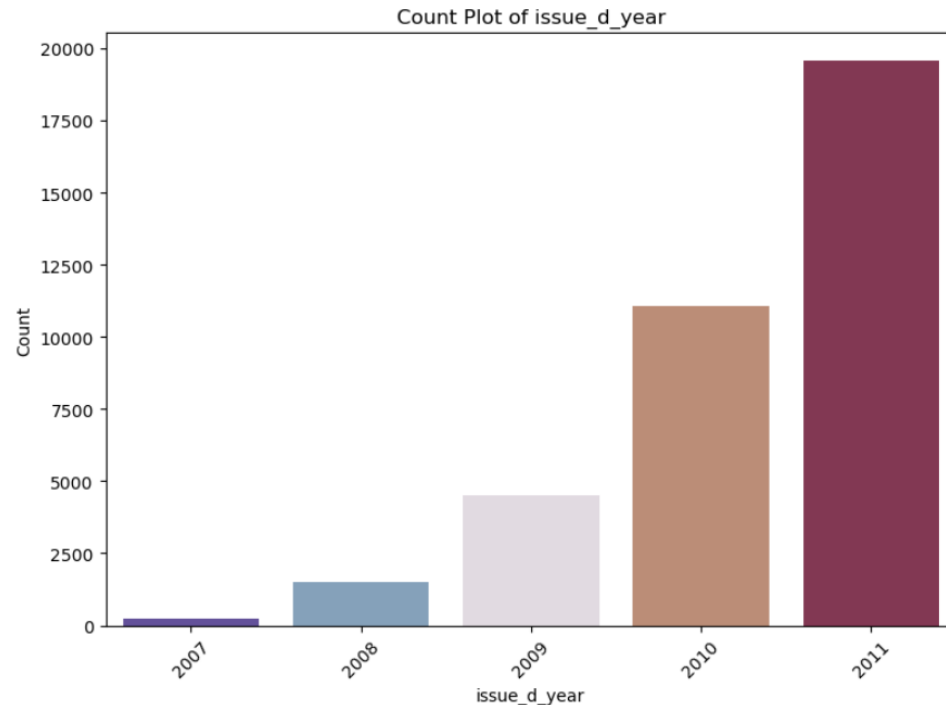
Outlier Handling: Identified and removed outliers to ensure data accuracy and reliability

Columns Retained for Analysis			
loan_amnt	funded_amnt	term	int_rate
grade	installment	emp_length	home_ownership
annual_inc	verification_status	issue_d	loan_status
purpose	addr_state	dti	delinq_2yrs
earliest_cr_line	inq_last_6mths	open_acc	pub_rec,
revol_bal	total_acc	sub_grade	

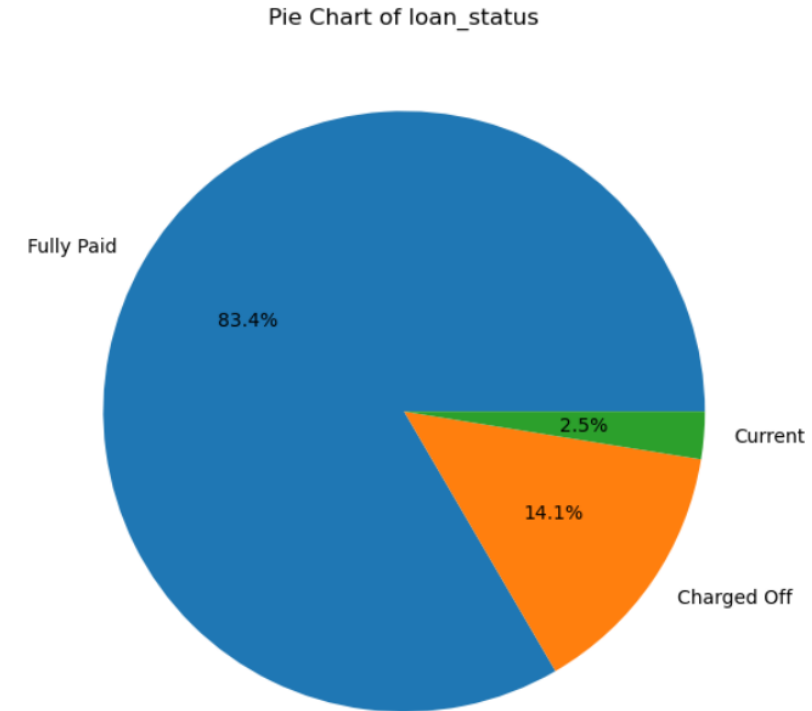


Exploratory Data Analysis (EDA)

Initial Findings:



- **100% increase in Loans** issued from every year starting from 2007 to 2011, indicating the growth of the business

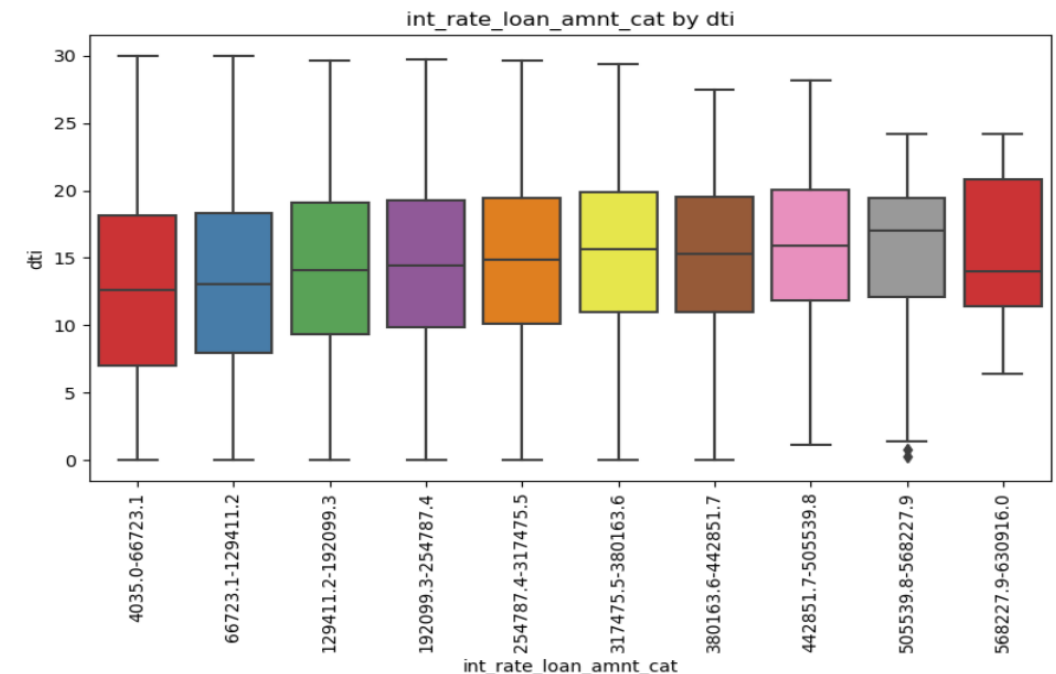
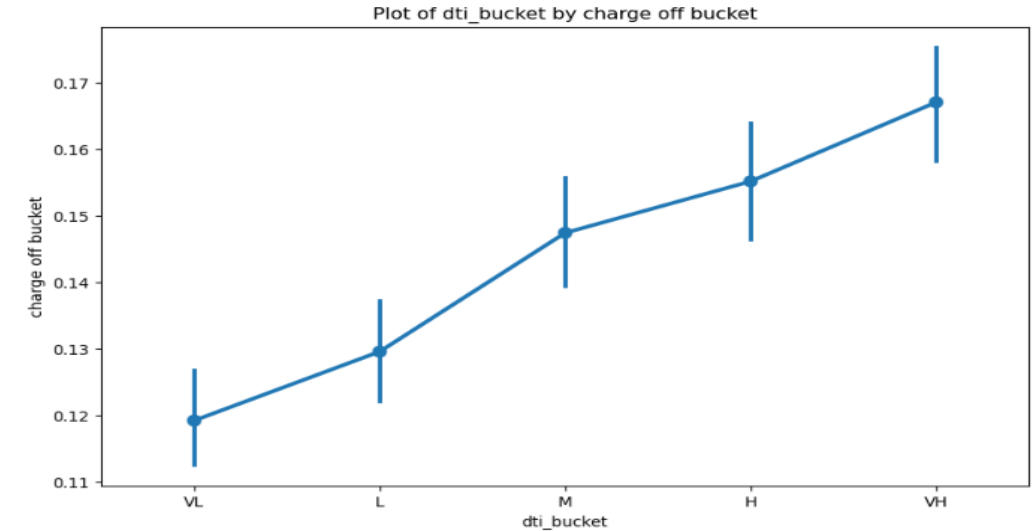


- **83.4%** loans are **Fully paid**
- **14.1%** loans are **Charged Off**

DTI – Key Insights

	min	max
dti_bucket		
VL	0.00	7.00
L	7.01	11.40
M	11.41	15.36
H	15.37	19.75
VH	19.76	29.99

- ❑ Higher Charge Off % for Higher DTI -> **Increasing trend observed**
- ❑ DTI shows a slight positive correlation (0.095) with Loan Amount and Interest Rates.



Recommendation:

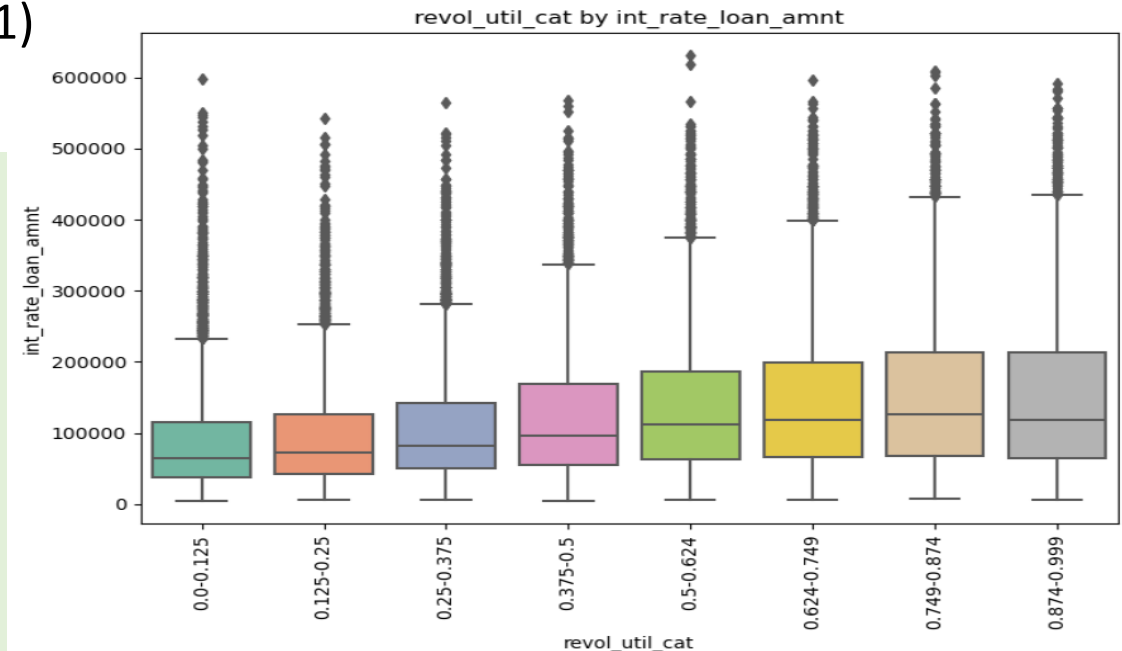
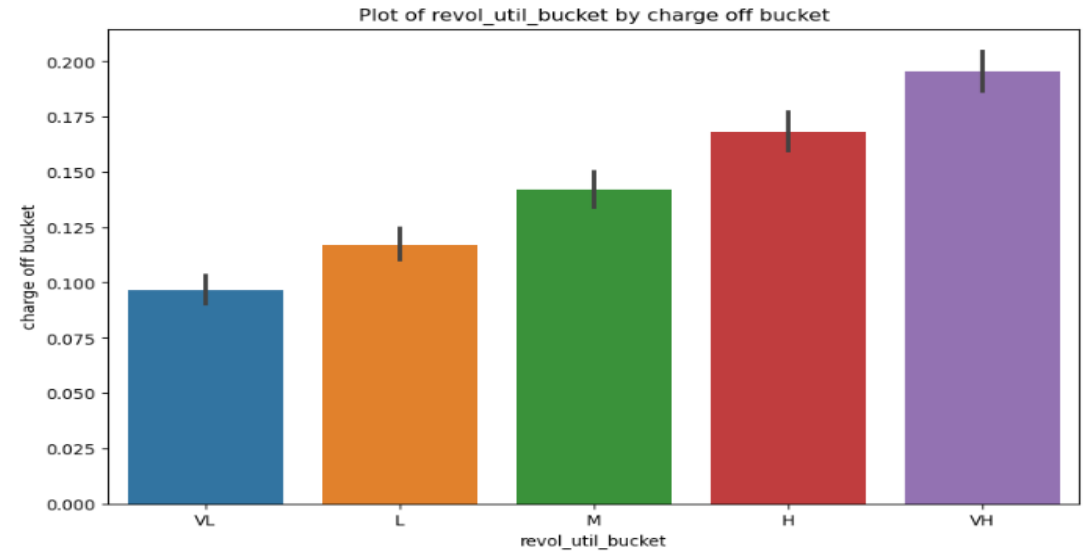
- ✓ **Increased Scrutiny for Higher DTI**
 - Implement stricter approval criteria or higher interest rates for applicants with a DTI above 15% to mitigate default risk.

This becomes a driver for the business in several ways.

Revolving Utilization Rate – Key Insights

	min	max
revol_util_bucket		
VL	0.0	20.3
L	20.4	40.3
M	40.4	58.5
H	58.6	77.4
VH	77.5	99.9

- ❑ Higher Charge Off % for Higher Revolving Utilization -> **Increasing trend observed**
- ❑ Revolving Utilization shows positive correlation with Interest Rates (~0.5) and Loan Amount (~0.1)



Recommendation:

- ✓ **Implement Revol Utilization Thresholds**
 - Strict credit utilization thresholds during the loan approval process, say above 0.50 (50%) utilization rates should undergo additional scrutiny.

This becomes a driver for the business in several ways.

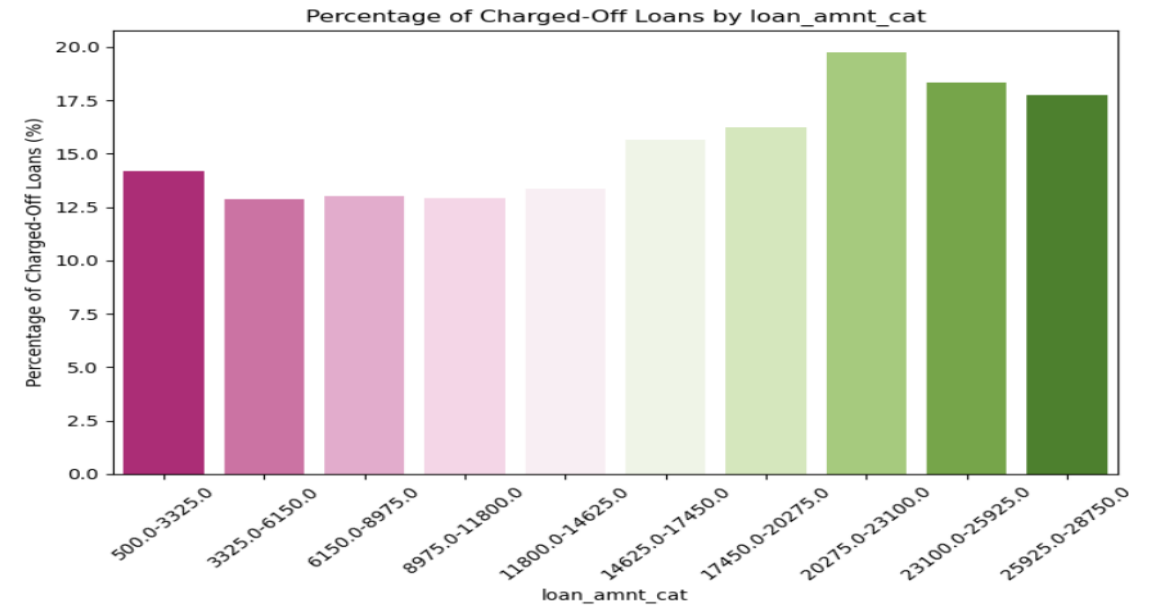
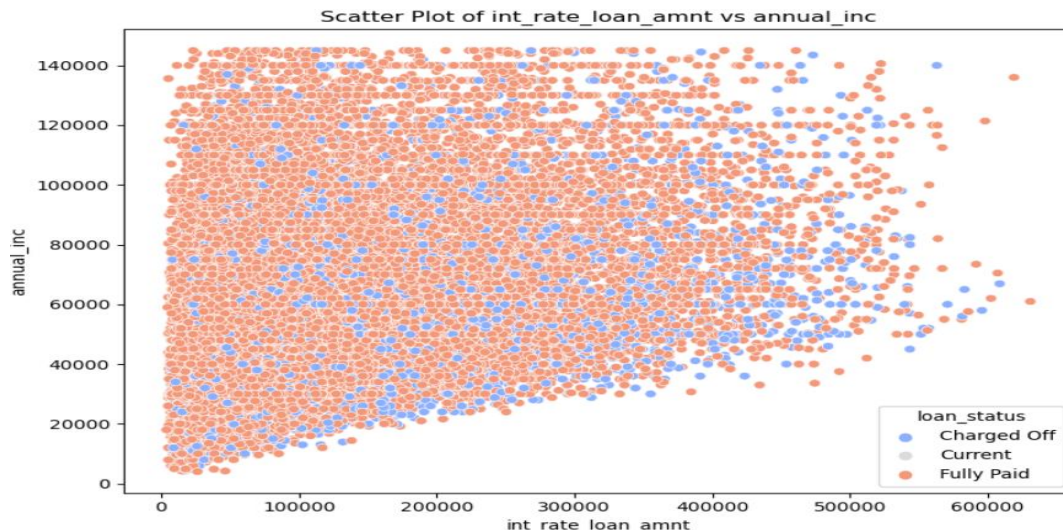
Loan Amount - Key Insights

❑ High Default Risk for Large Loans:

- ✓ Loans above 12000 exhibit a significantly higher risk of default, with elevated charge-off percentages compared to smaller loans.

❑ Clear Trend in Default Rates:

- ✓ Annual Income, Interest Rate, Loan Amount are very strongly correlated and the chances of getting defaulted increase these parameters increase.



Recommendation:

✓ *Scrutinize Large Loan Applications:*

- Implement additional screening for higher amount loans

✓ *Adjust Interest Rates:*

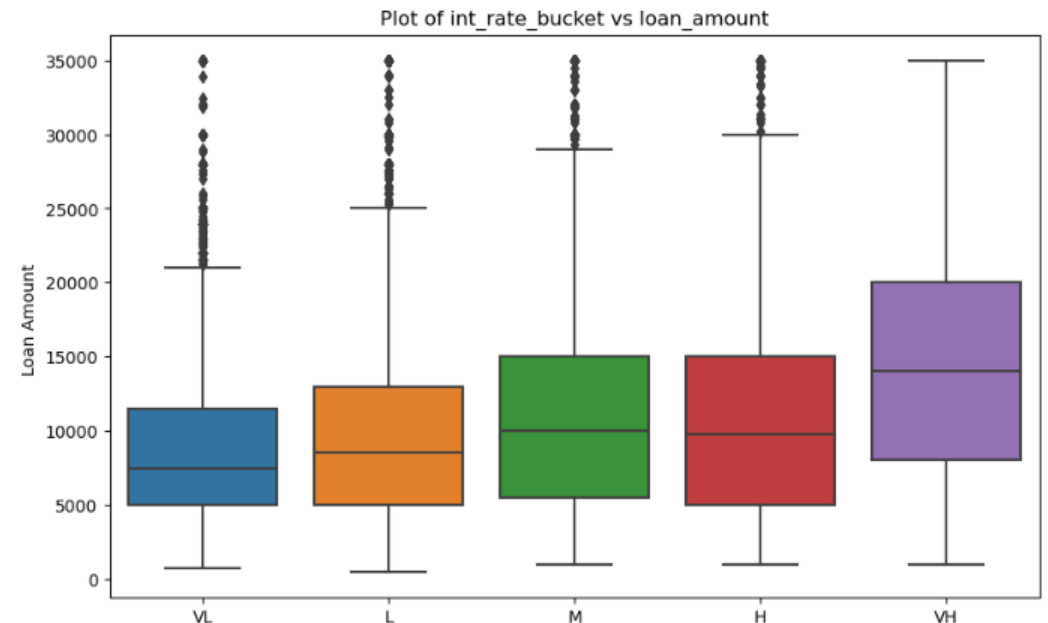
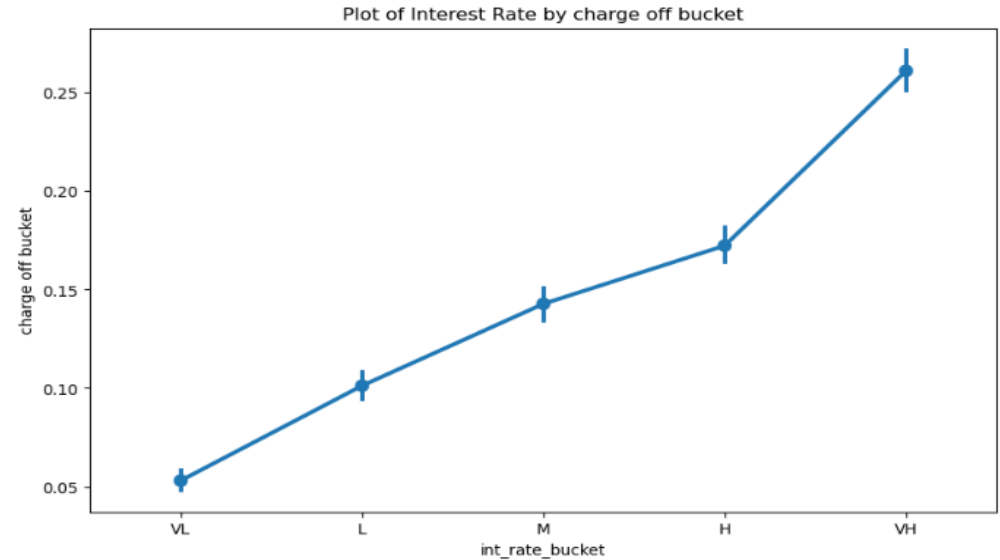
- Consider increasing interest rates for higher loan amounts to compensate for the increased risk of default.

This becomes a driver for the business in several ways.

Interest Rate – Key Insights

	min	max
int_rate_bucket		
VL	5.42	7.90
L	7.91	10.99
M	11.03	12.98
H	12.99	15.27
VH	15.28	24.40

- ❑ Higher Charge Off % for Higher Interest Rates -> ***Increasing trend observed***
- ❑ Interest Rates show a positive correlation with Loan Amount.



Recommendation:

- ✓ Reevaluate Interest Rate Policies
 - Consider reassessing the interest rate structure to balance risk and return, particularly for high-interest loans.

This becomes a driver for the business in several ways.

Annual Income

Higher Annual Income Reduces Default Risk:

- ✓ Borrowers with annual incomes over \$83,000 have a charge-off percentage of only 10.06%, compared to ~18% for those with incomes between \$4,000 and \$36,000.

Clear Trend in Default Rates:

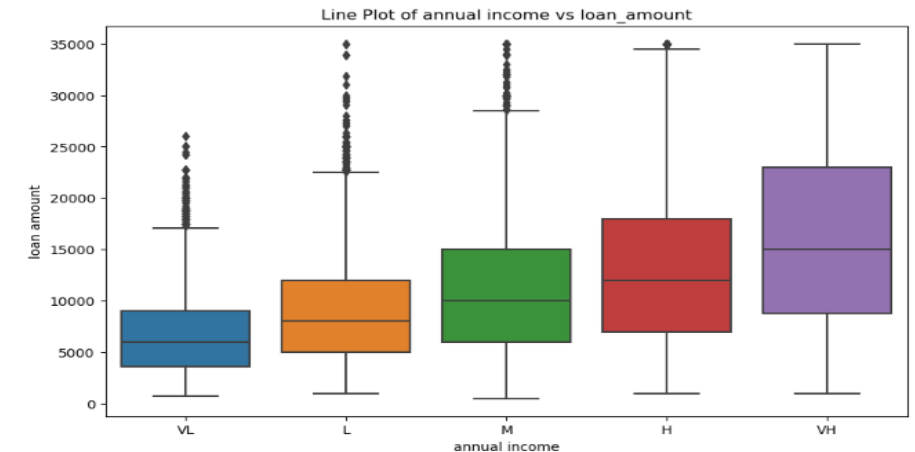
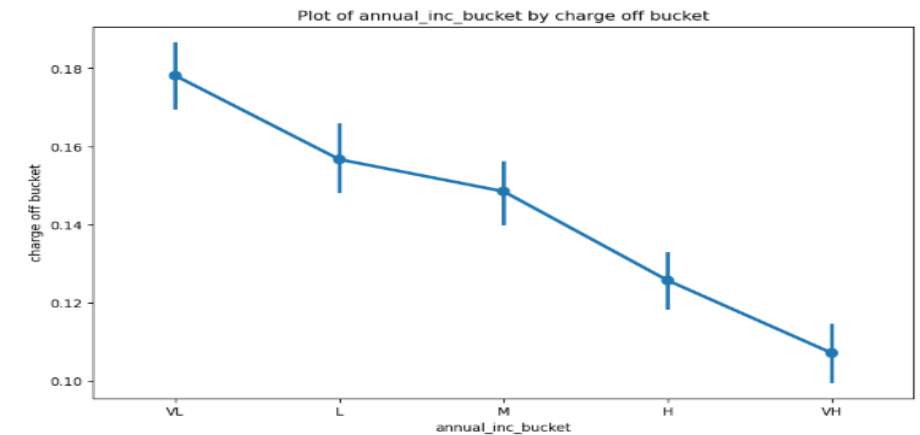
- ✓ As annual income increases, the % of fully paid loans rises and the % of charged-off loans decreases, suggesting a strong correlation between higher income and lower default risk.

Recommendation:

- ✓ *Prioritize High-Income Applicants*
 - Consider focusing on or offering better terms to borrowers in higher income brackets to potentially reduce default rates.
- ✓ *Reevaluate Risk Assessment Criteria*
 - Adjust loan approval and risk assessment processes to account for income levels more effectively.

This becomes a driver for the business in several ways.

	min	max
annual_inc_bucket		
VL	4000.00	38000.0
L	38000.04	50500.0
M	50508.00	66000.0
H	66072.00	90000.0
VH	90050.00	600000.0



Installments - Insights

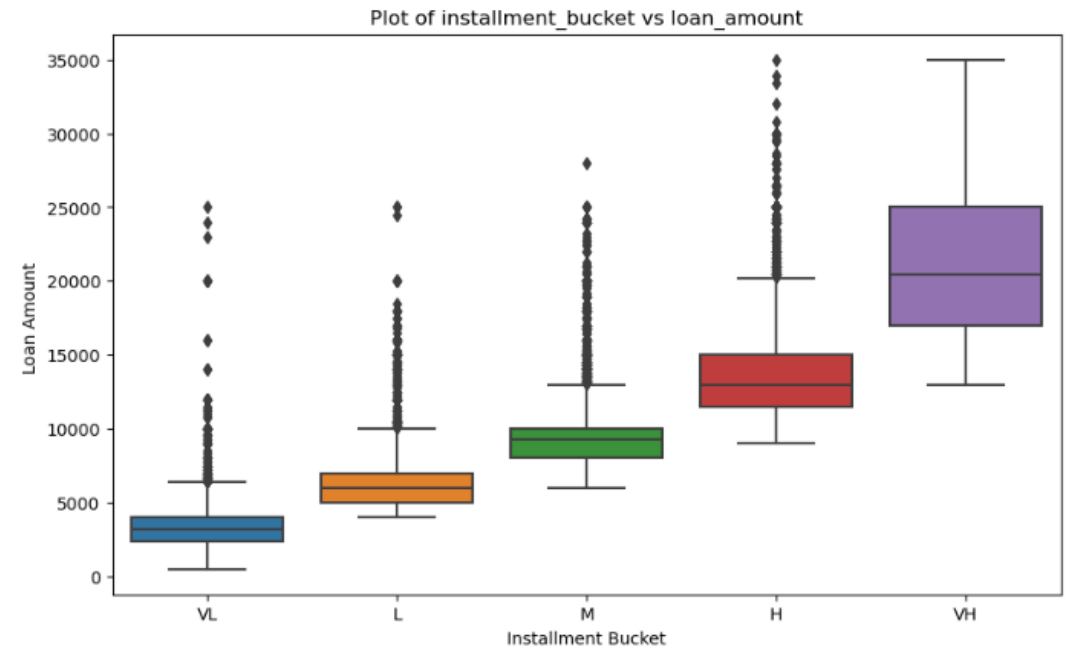
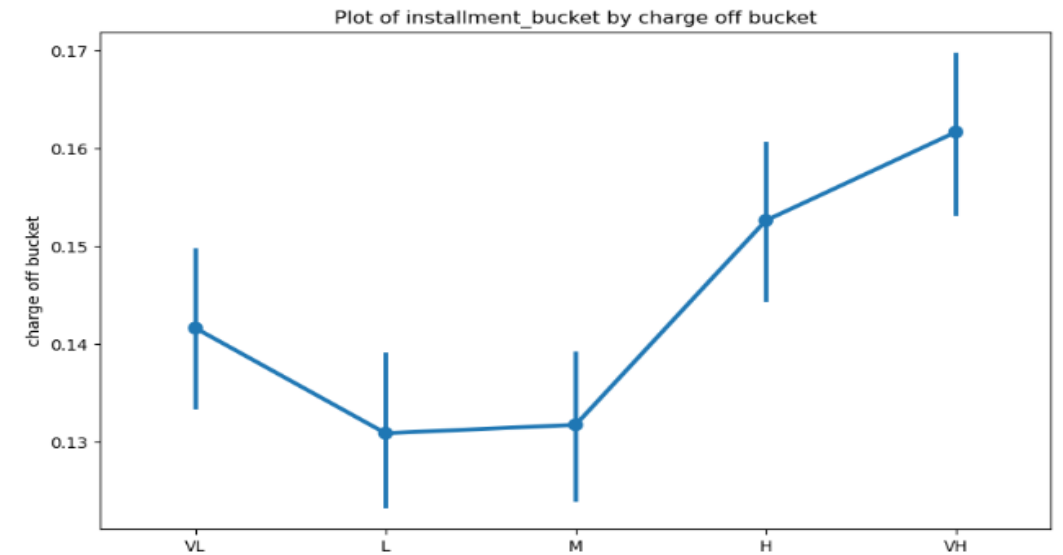
	min	max
installment_bucket		
VL	16.08	150.42
L	150.48	229.14
M	229.15	328.10
H	328.14	480.29
VH	480.30	1305.19

- ❑ Instalment on average above \$328/month has higher % charge off compared to lower buckets.
- ❑ Instalments show positive correlation with Loan Amounts and Interest Rates.

Recommendation:

- ✓ *Since the lower buckets are showing more stable results, the business can use this as a driver and reduce the probability of getting defaulted.*

This becomes a driver for the business in several ways.



Loan-to-Income (LTI) - Key Insights

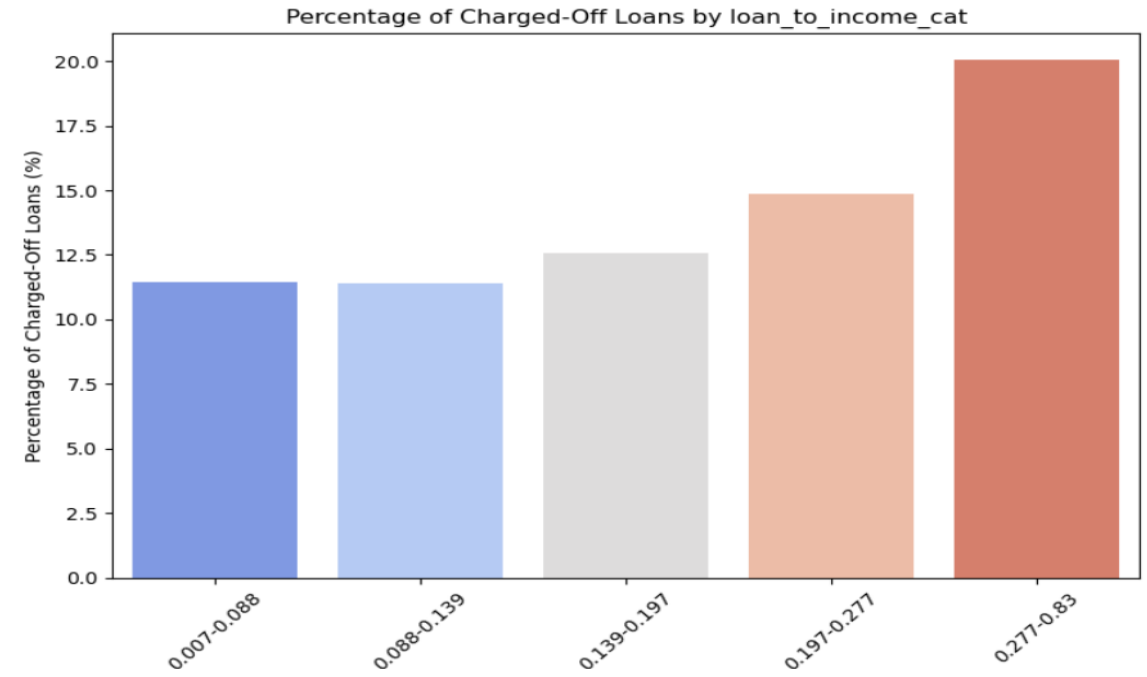
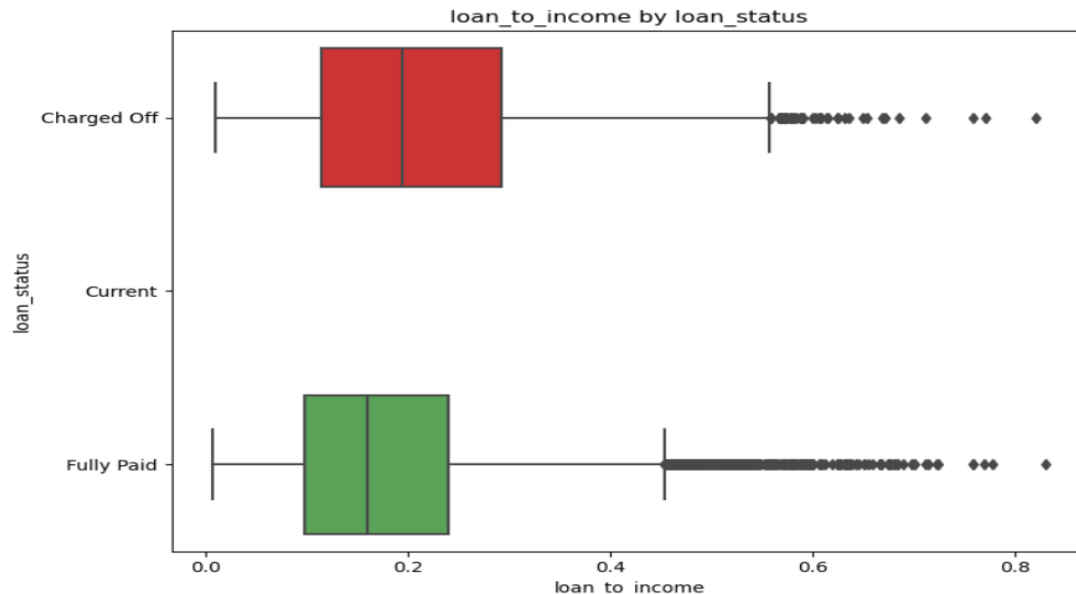
$$LTI = (Loan\ Amount / Annual\ Income)$$

Higher Loan-to-Income Ratio & Default Risk:

- ✓ Borrowers with a loan-to-income ratio above 0.25 have a default risk exceeding 18.30%.

Clear Trend in Default Rates:

- ✓ A direct relationship exists between loan-to-income ratio and default rates; higher ratios are linked to increased default percentages.



Recommendations:

- **Monitor High Loan-to-Income Ratios:**
 - ✓ Consider implementing additional scrutiny or higher interest rates for the borrowers with loan-to-income ratios above 0.25.
- **Adjust Loan Issuance Policies:**
 - ✓ Prioritize lending to those with lower loan-to-income ratios to minimize default risks.

This becomes a driver for the business in several ways.

Credit-Line-Age

The duration of a borrower's credit history at the time of loan issuance = **(Loan Issue Date - Earliest Credit Line Date)**

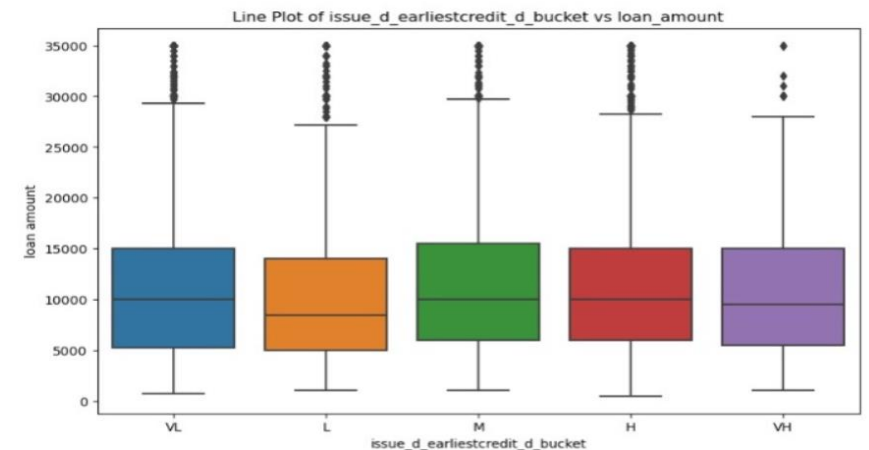
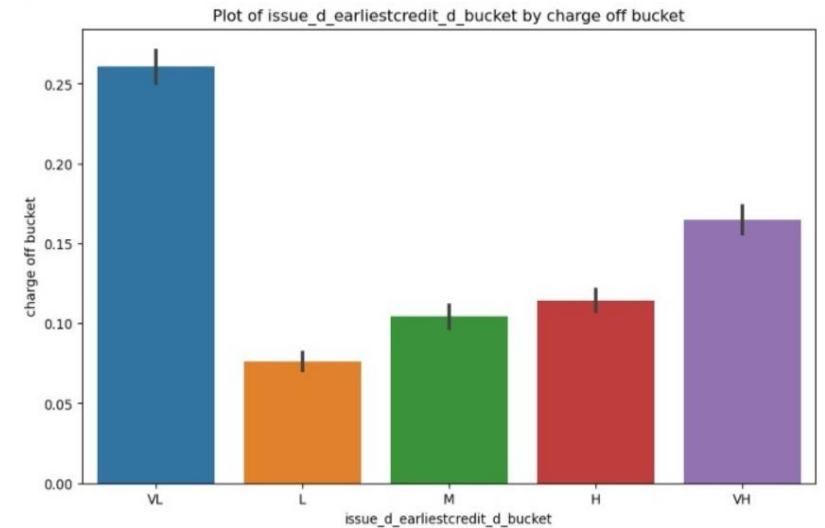
	min	max
issue_d_earliestcredit_d_bucket		
VL	-884 days +00:00:00	882 days 00:00:00
L	883 days 00:00:00	1127 days 00:00:00
M	1155 days 00:00:00	1643 days 00:00:00
H	1644 days 00:00:00	1888 days 00:00:00
VH	1915 days 00:00:00	3135 days 00:00:00

❑ Trend in Default Rates:

- ✓ Very low credit line age (< than 2.5 years) are associated with highest default rates of more than 25% (while avg loan amount in each bucket is in same range)

- ✓ Rest all buckets are either below average or around the average of ~ 15%

Note: There are 5 records with negative credit line ages, which are almost of negligible percentage and hence retaining the same for analysis as they wont impact insights.



Recommendations:

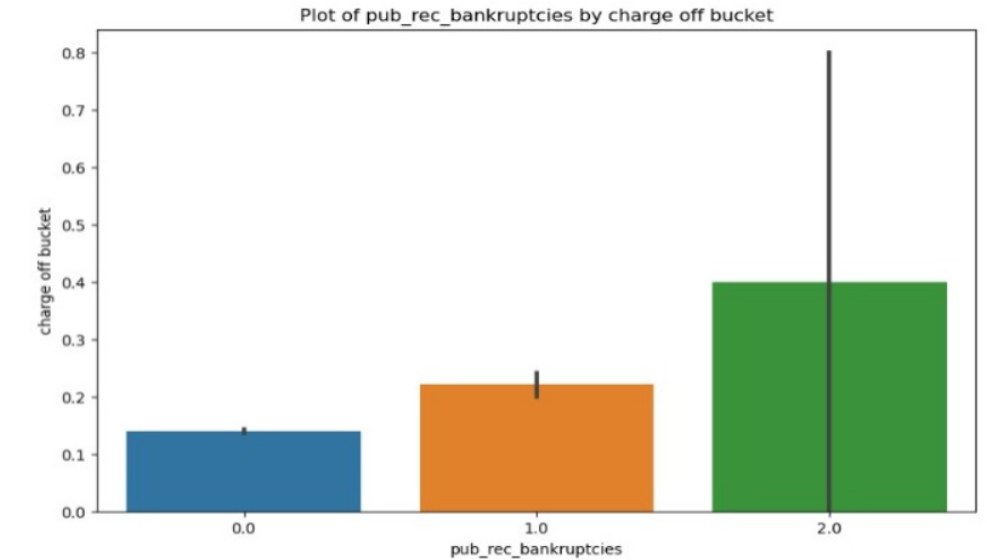
➤ Monitor Credit Line Trends

- ✓ Make sure to have a check on this variable to take informed decisions about the loan approval.

This becomes a driver for the business in several ways.

Public Record Bankruptcies

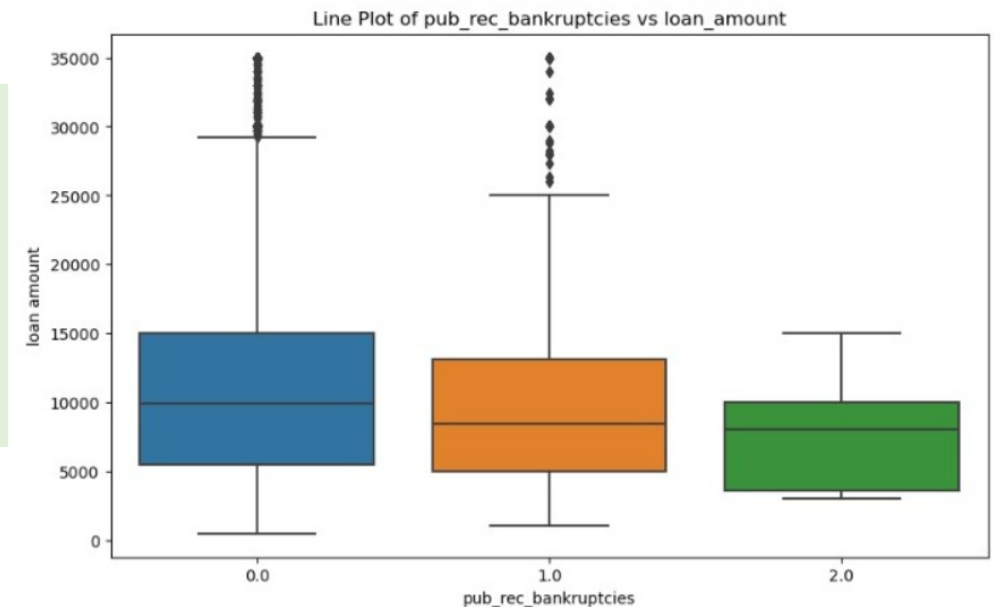
- ❑ Higher Public Rec Bankruptcies, higher the Charge Off
 - ✓ . If it's greater than 0, there are more high chances of getting defaulted
- ❑ Clear Trend in Default Rates:
 - ✓ A **direct relationship** exists between pub_rec_bankruptcies and default rates; higher ratios are linked to increased default percentages.



Recommendations:

- Monitor the trends
 - ✓ Make sure to have a check on this variable to take informed decisions about the loan approval.

This becomes a driver for the business in several ways.

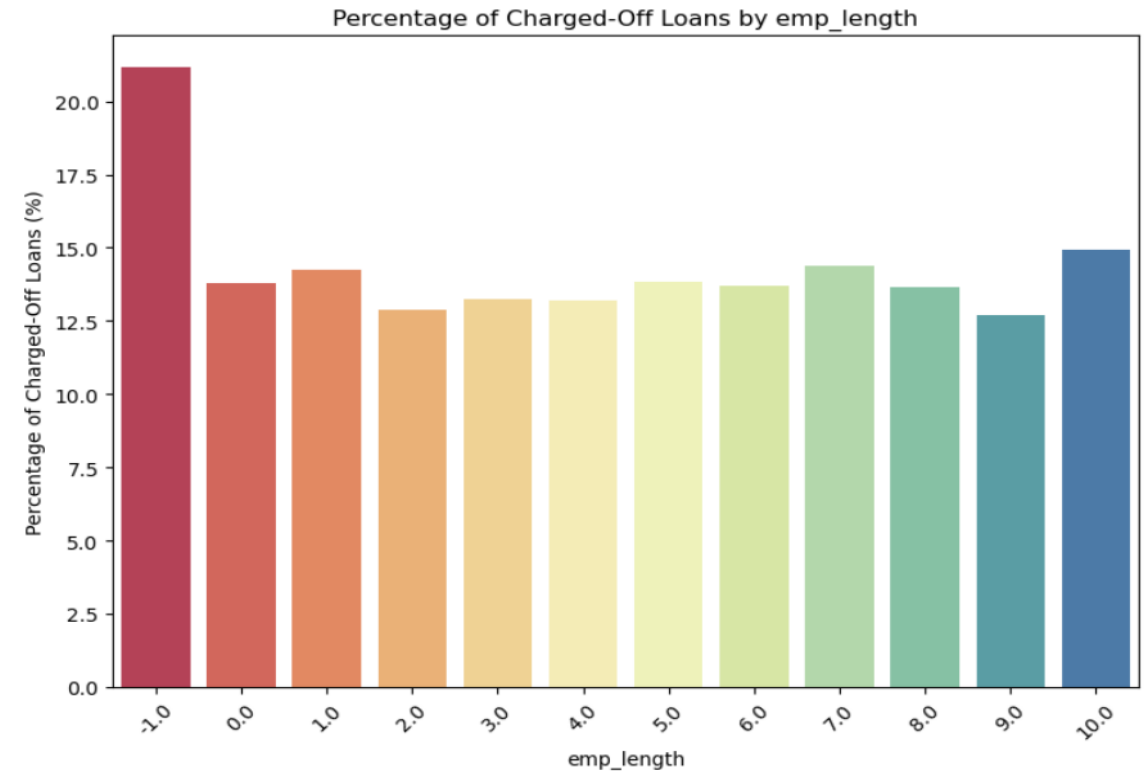
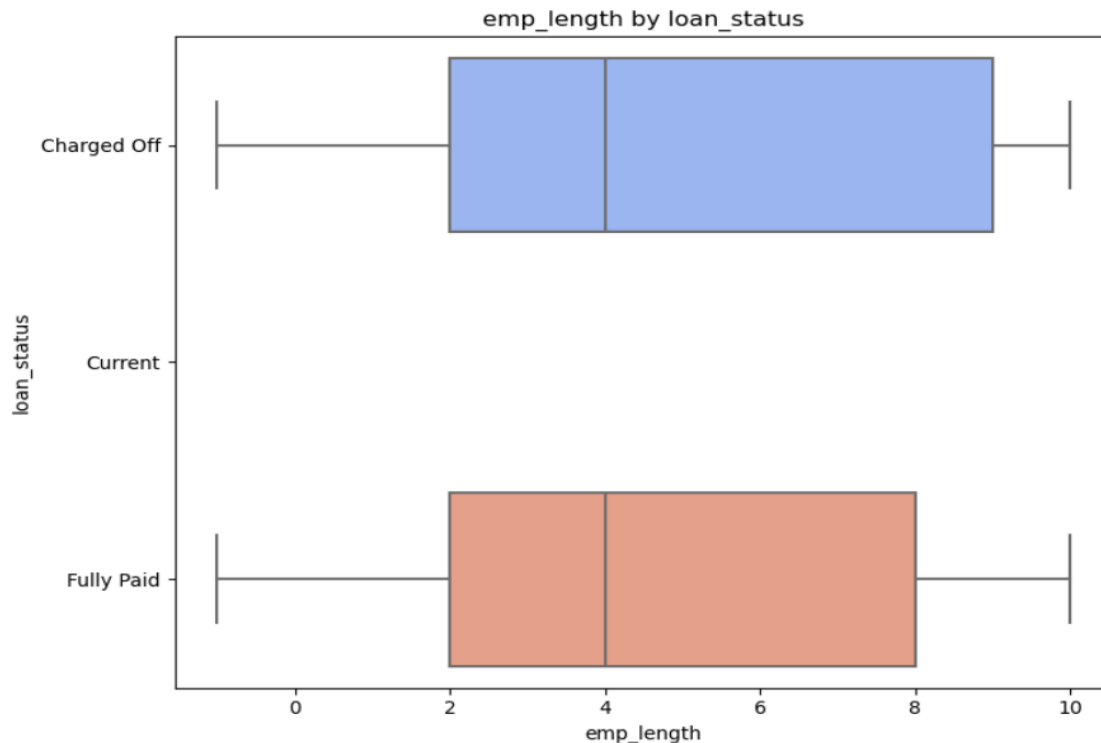


Employment Length – Key Insights

❑ High Default Rate with Missing Employment Data

- Loans with **missing employment length data** have a notably higher default rate (**21.26%**).

❑ Tenure **NOT** Always Indicative of Default



Recommendations:

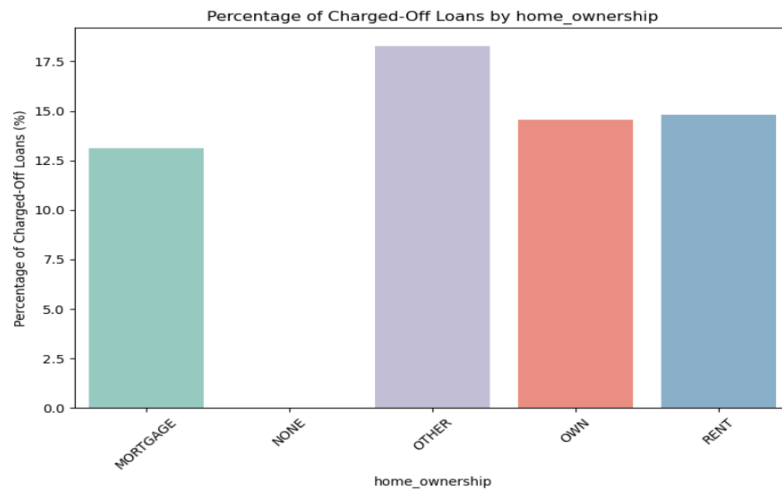
➤ Address Missing Data:

- ✓ Implement measures to make employment length a mandatory field in loan applications to ensure complete data collection, thereby proceed with further verifications.

Other Notable Insights – 1/3

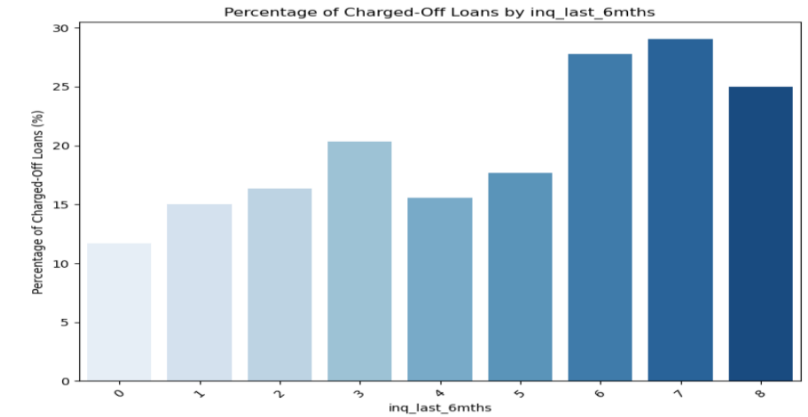
❑ Inquiries in past 6 months:

- ✓ Borrowers with more than 6 credit inquiries in the last 6 months show a significant increase in default risk, with the percentage of charged-off loans rising sharply from 11.71% to 25%.
- ✓ No Clear Trend in the pattern.



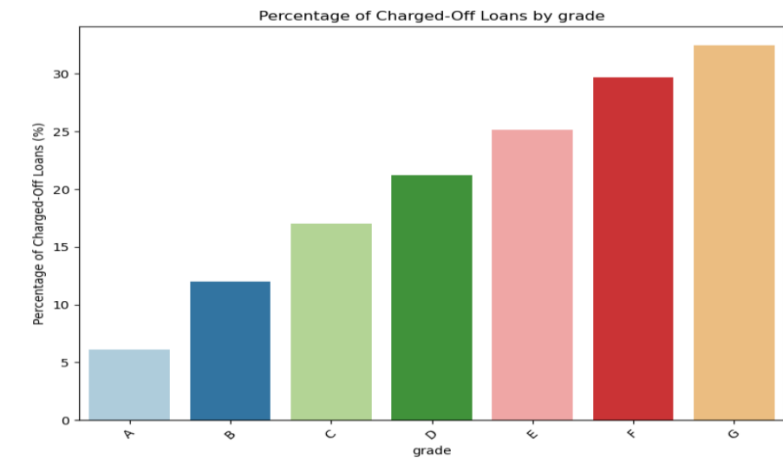
❑ Home Ownership

- ✓ Higher charge-off percentages identified in “Other” category
- ✓ No Clear Trend in the pattern.



❑ Grade / Sub – grade

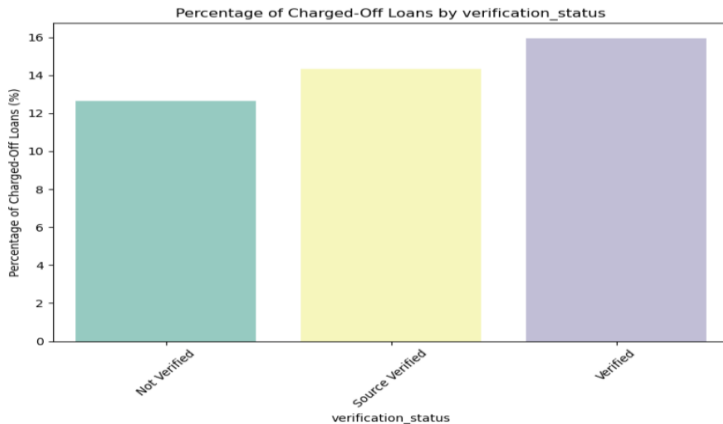
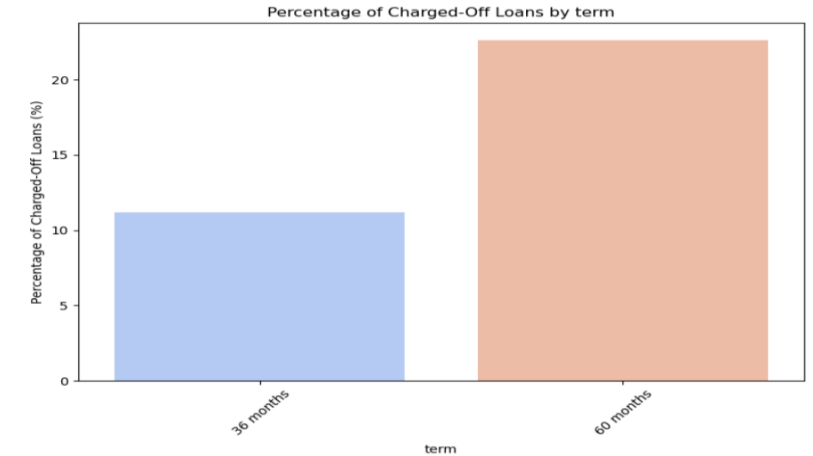
- ✓ As grades progress from A to G, the percentage of defaults increases directly. Subgrade is the further categorization of grade.
- ✓ Though we have a clear trend, grade could have been derived from loan amount and interest rate



Other Notable Insights – 2/3

❑ Term:

- Loans with a term of 60 months have a significantly higher charge-off percentage (22.62%) compared to those with a 36-month term (11.21%).

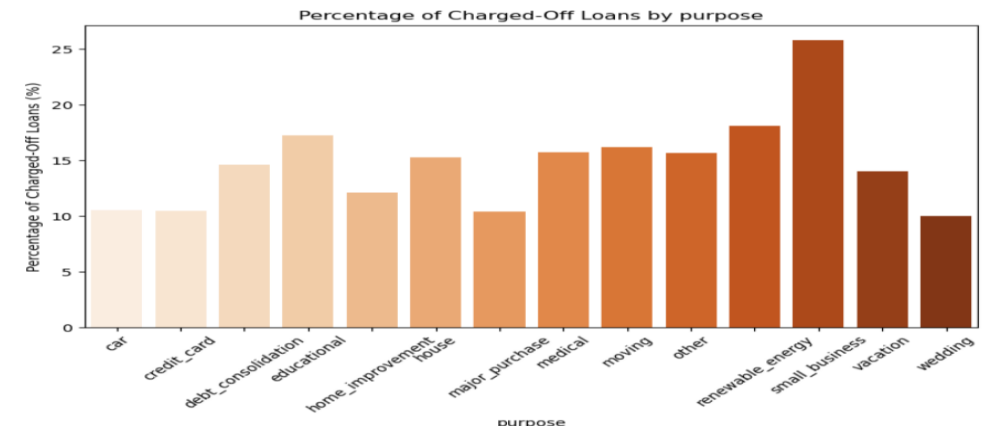


❑ Verification Status:

- No much valuable insights coming out from this since verified status has more defaulter %, which is not adding any value to our insights.

❑ Purpose:

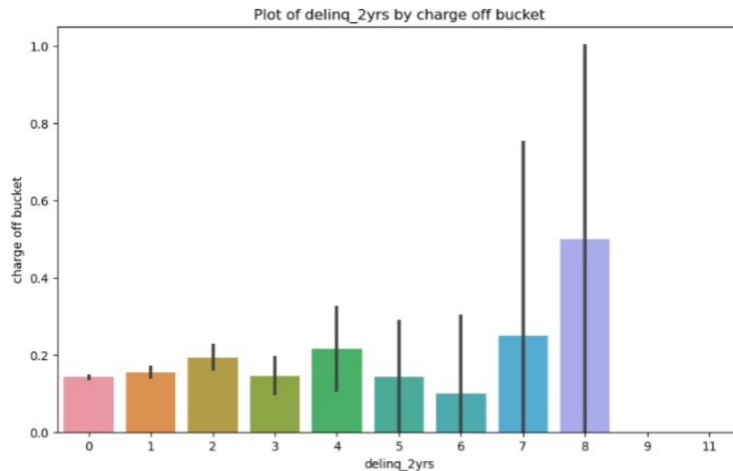
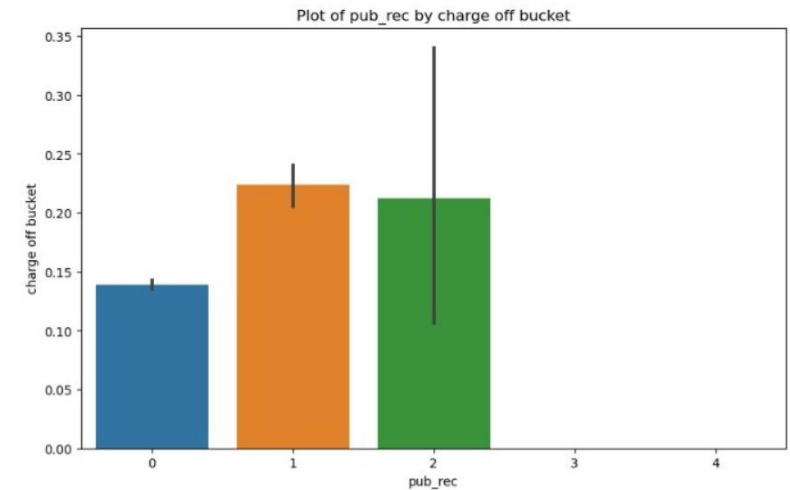
- Small Business has the highest defaulting % of 25.82 followed by Renewable Energy with 18.09%



Other Notable Insights 3/3

❑ Public Record Bankruptcies:

- Though there is a clear spike from 0 to 1, this has no much insights.

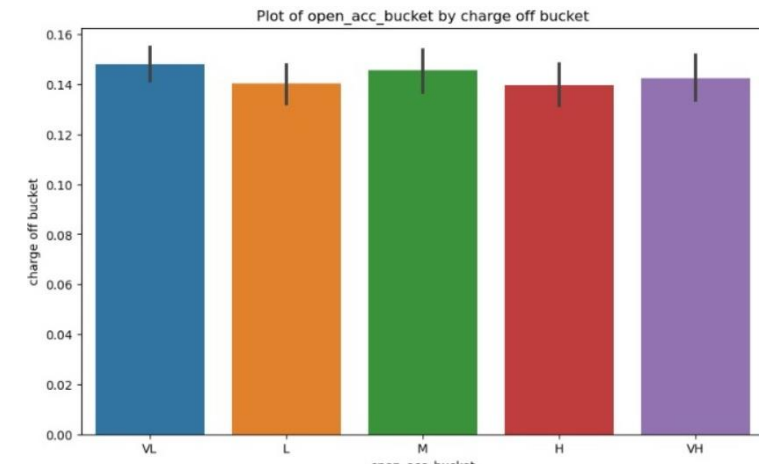


❑ Delinquent Last 2 years

- No much valuable insights coming out from this since there is no proper trend seen here.

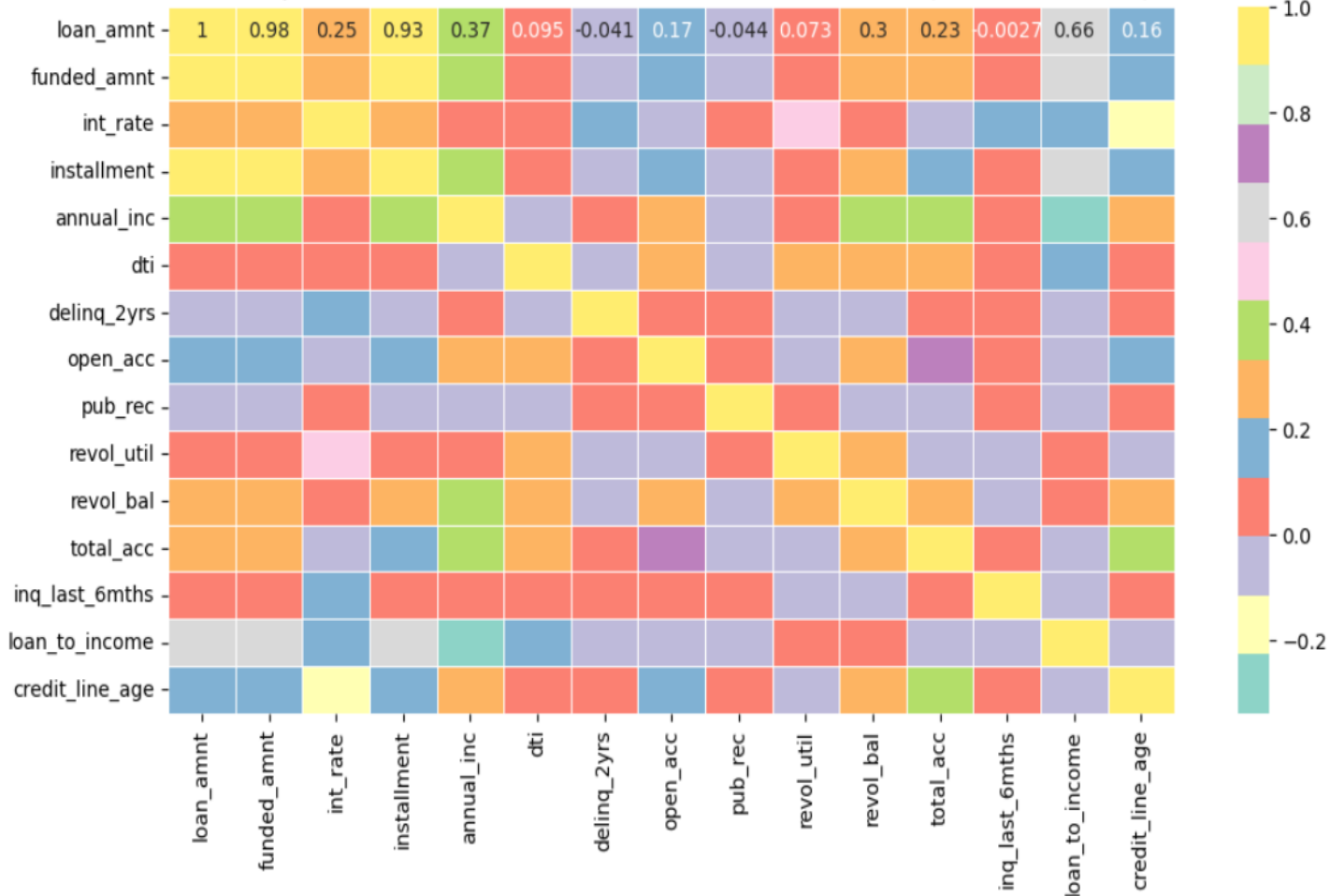
❑ Open Credit Lines

- No clear trend is identified as the charge off seems consistent.

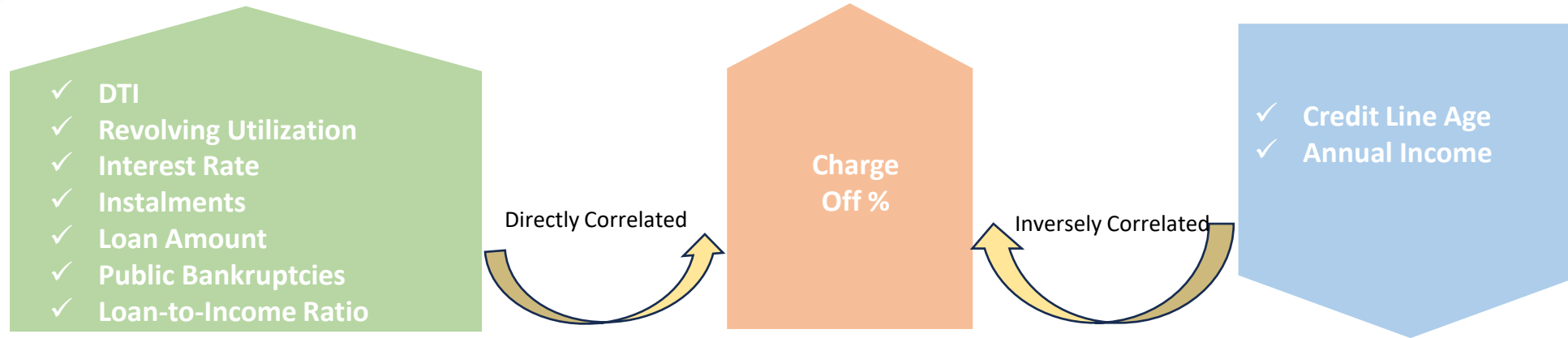


Heat Map

Heatmap of Correlation Matrix Inclusive of the Derived Variables (Numeric Variables)



- Plotted the heat map to understand the relationship between all the numeric variables.
- Analysed further to derive on insights and identify the drivers.



Recommendations:

1. **Missing Employee Length** : Address Missing Data issues – Try making the field mandatory for data completeness
2. **Loan Amount** : Scrutinize Large Loan Applications above 12k
3. **DTI** : Implement stricter approval criteria or higher interest rates for applicants with a DTI above 15% to mitigate default risk
4. **Revolving Utilization** : Additional scrutiny for Utilization above 0.50 or (50%)
5. **Interest Rate** : Consider reassessing the interest rate structure to balance risk and return, particularly for high-interest loans (above 15%)
6. **Instalments** : Lower buckets (< \$330) show more stable results, the business can use this as a driver and reduce the probability of getting defaulted
7. **Public Record Bankruptcies** : *Additional scrutiny for presence of public records bankruptcies (i.e., 1,2)*
8. **Loan – to – Income Ratio** : a) **Monitor High Loan-to-Income Ratios (> 0.25)** b) **Adjust Loan Issuance Policies**
9. **Credit Line Age** : Make sure to have a check on history of credit line (<2.5 years) to take informed decisions about the loan approval
10. **Annual Income** : a) Prioritize High-Income Applicants. b) Reassess risk rules for low annual income applicants