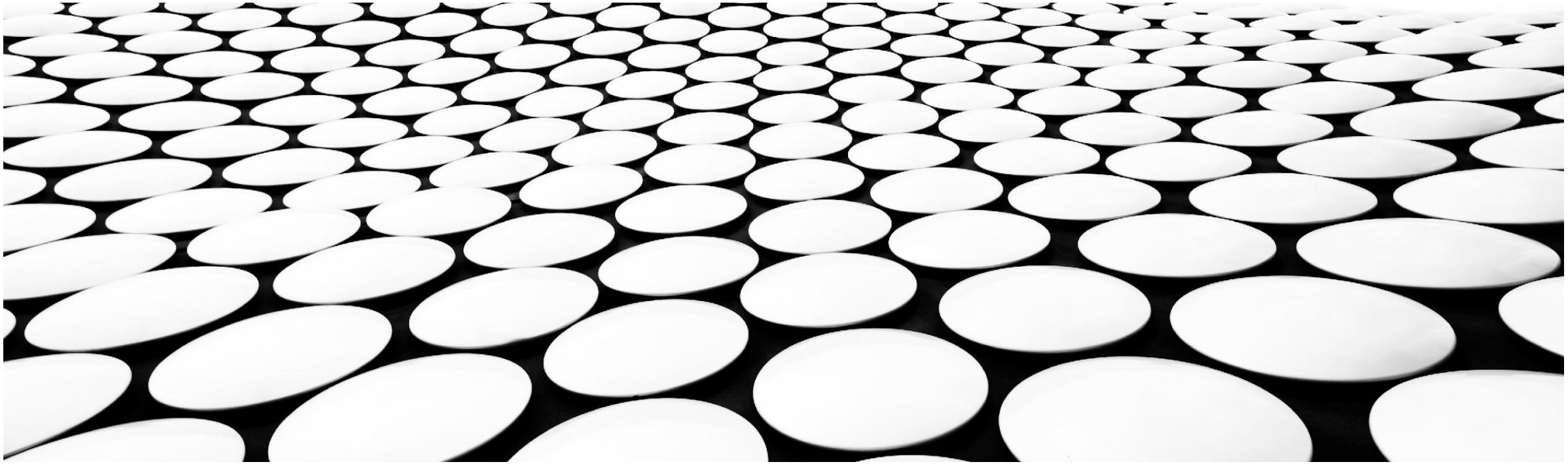
LENDING CLUB CASE STUDY



OBJECTIVE:

- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment

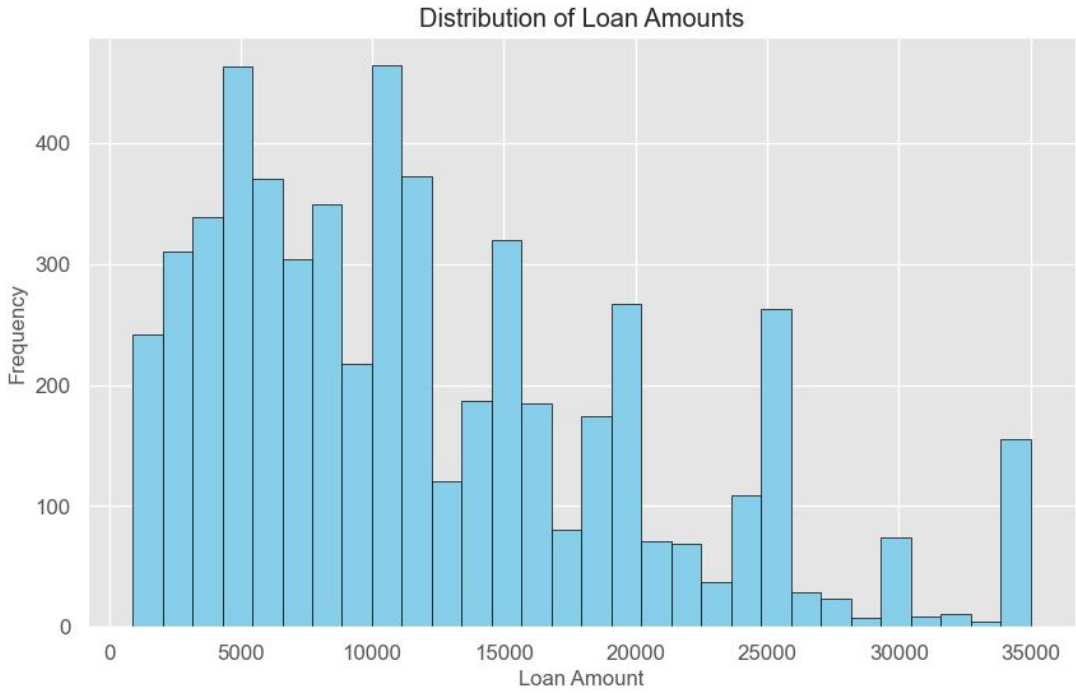
APPROACH:

We will be following the below standard approach to draw insights with respect to data provided:

- Data Cleaning
- Univariate Analysis
- Segmented Univariate Analysis and Bivariate Analysis
- Summary

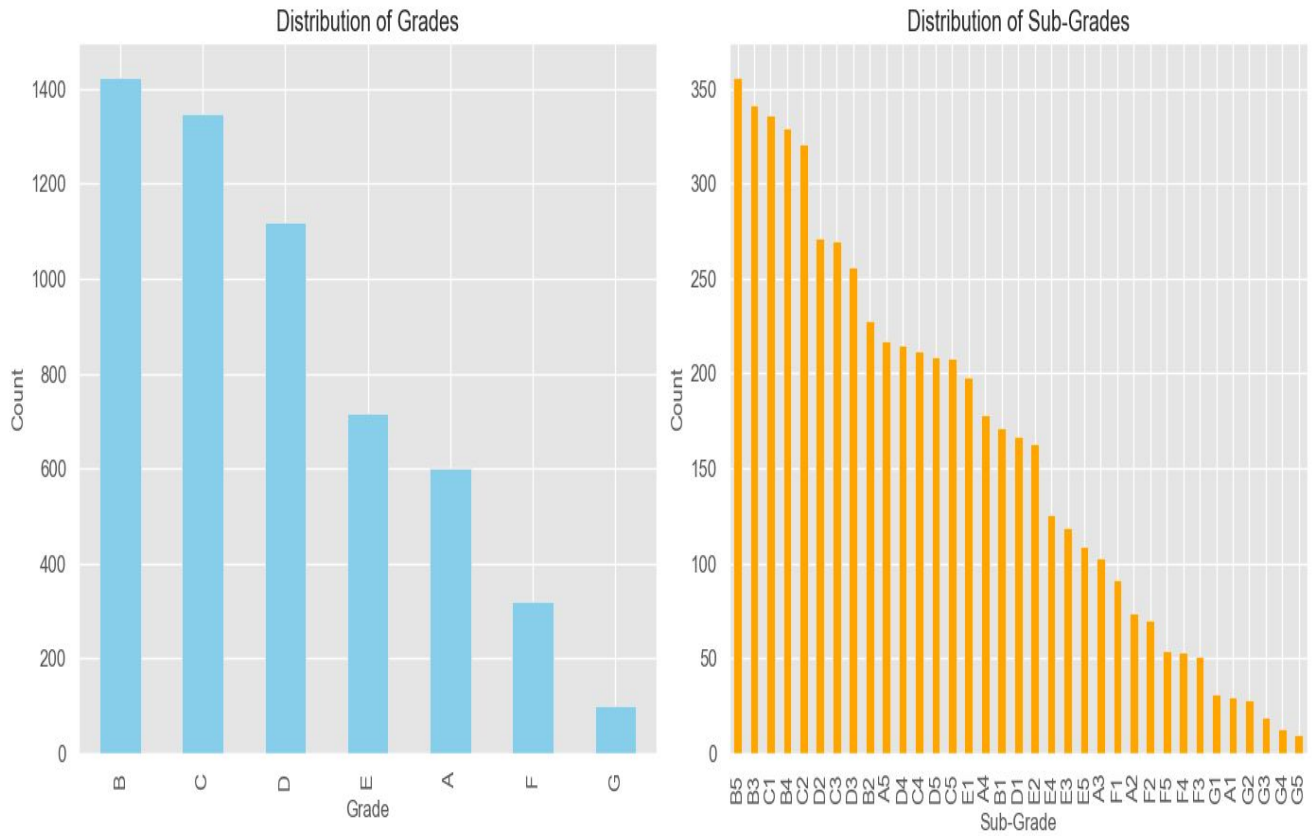
Distribution of Loan Amount By Charged Off status

- Most of the charged-off loans fall within the 5,000 to 10,000 range.
- Borrowers within this loan amount range are more likely to default, indicating a higher risk category.



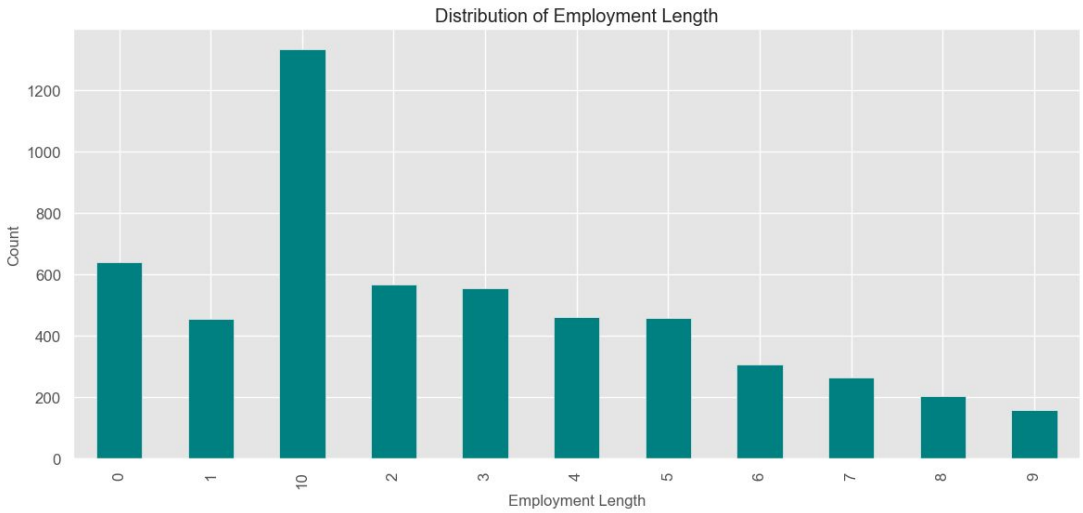
Distribution of Grades and Subgrades by Charged Off loan status

- Grade B had the most number of charged-off loans.
- Sub-grades B5, B3, B4, C1, and C2 had the most number of charged-off loans.
- Grade G had the least number of charged-off loans.



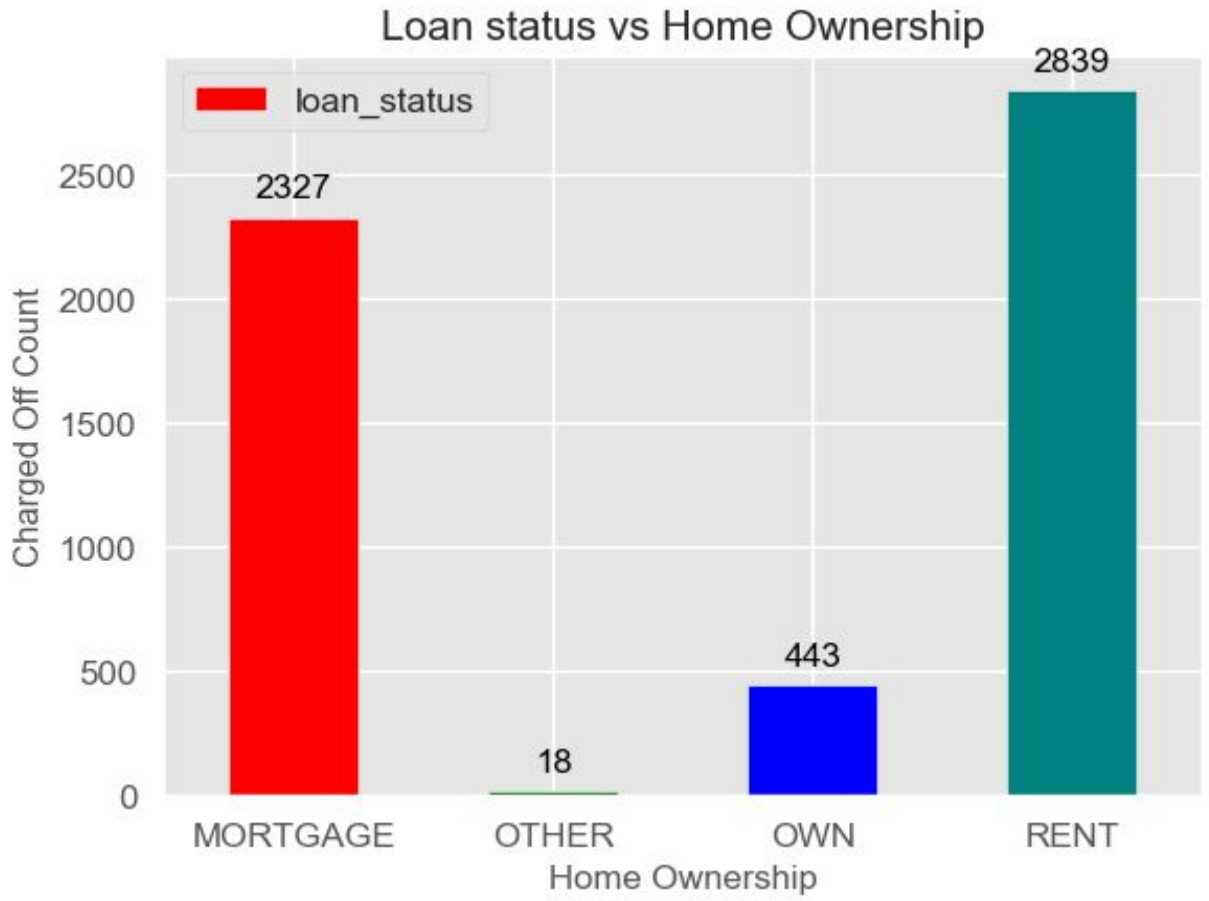
Employment Length Distribution by Charged Off Status

- Applicants having more than 10 years of employment has highest number off charged off loans
- Applicants having less than 1 year of employment are moderately risky loans



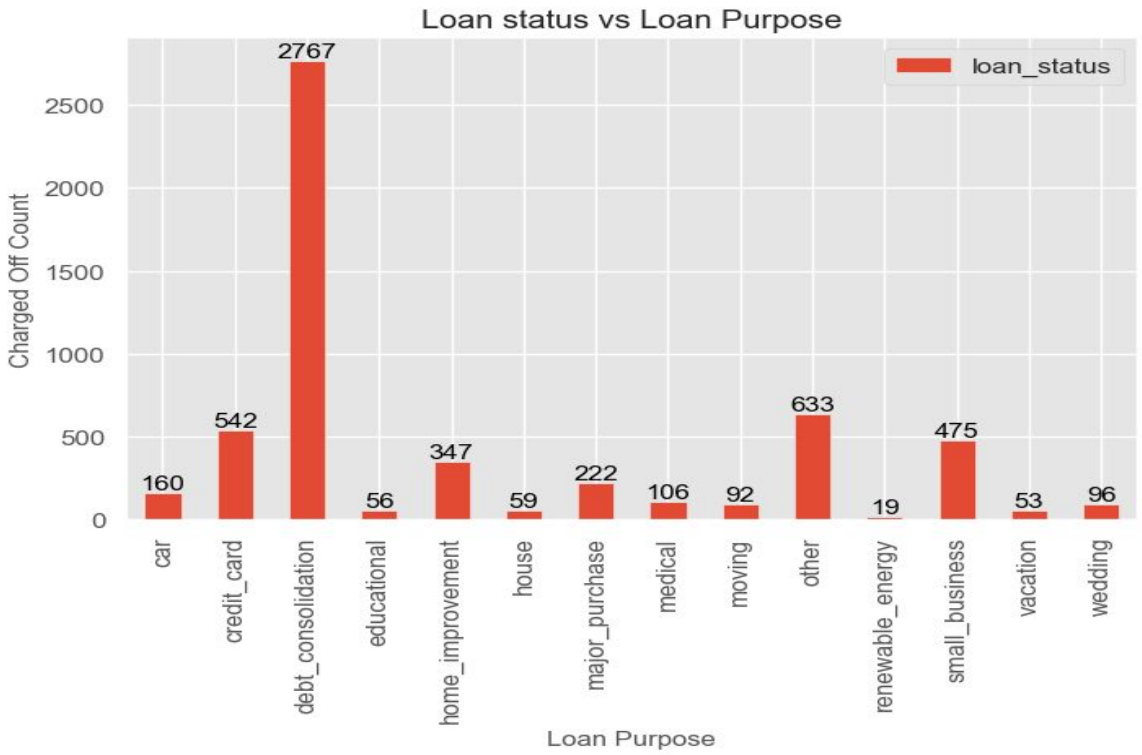
Charged off Loans by Home Ownership

- The applicants living on Rent or having Mortgages are more likely to default

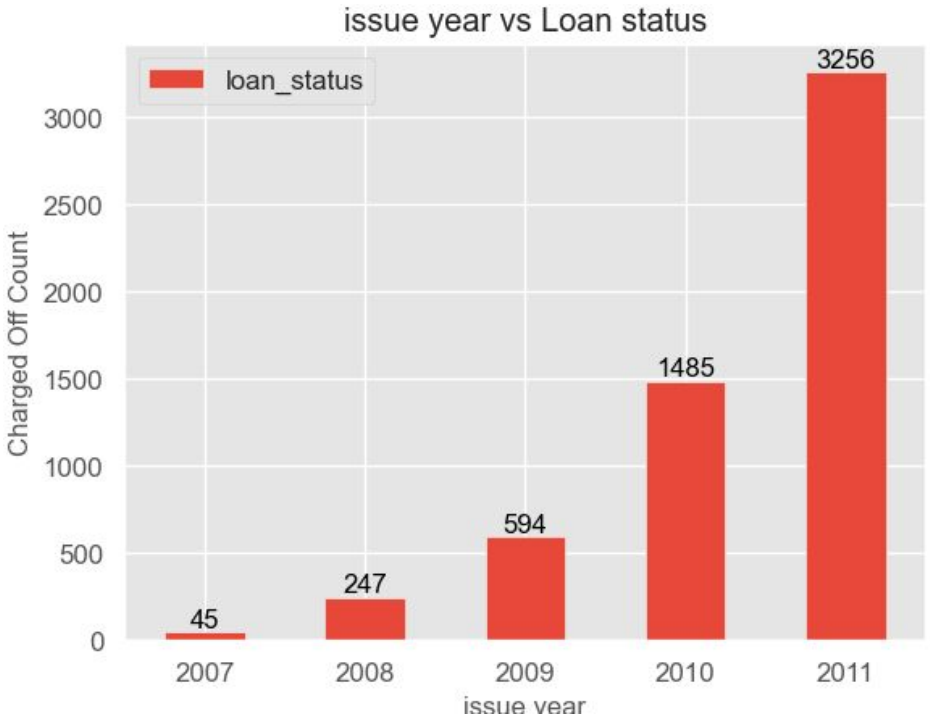


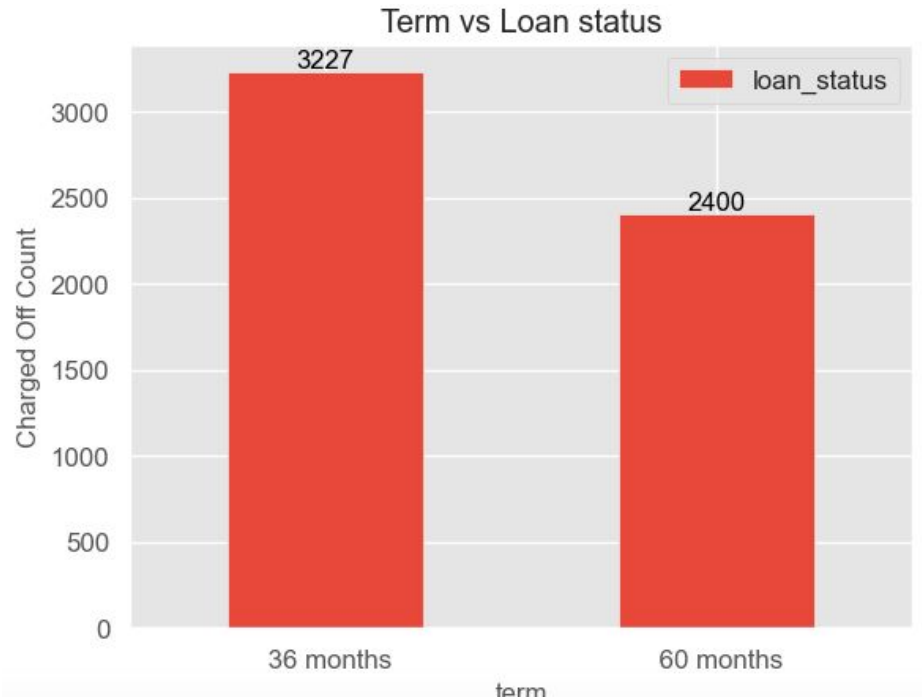
Charged off loans by Loan Purpose

- Applicants taking loan for Debt Consolidation are more likely to default.
- Applicants taking loan for Miscellaneous purposes are more likely to default.
- Loan to repay credit card bill is also a moderate indicator that the applicant will default.



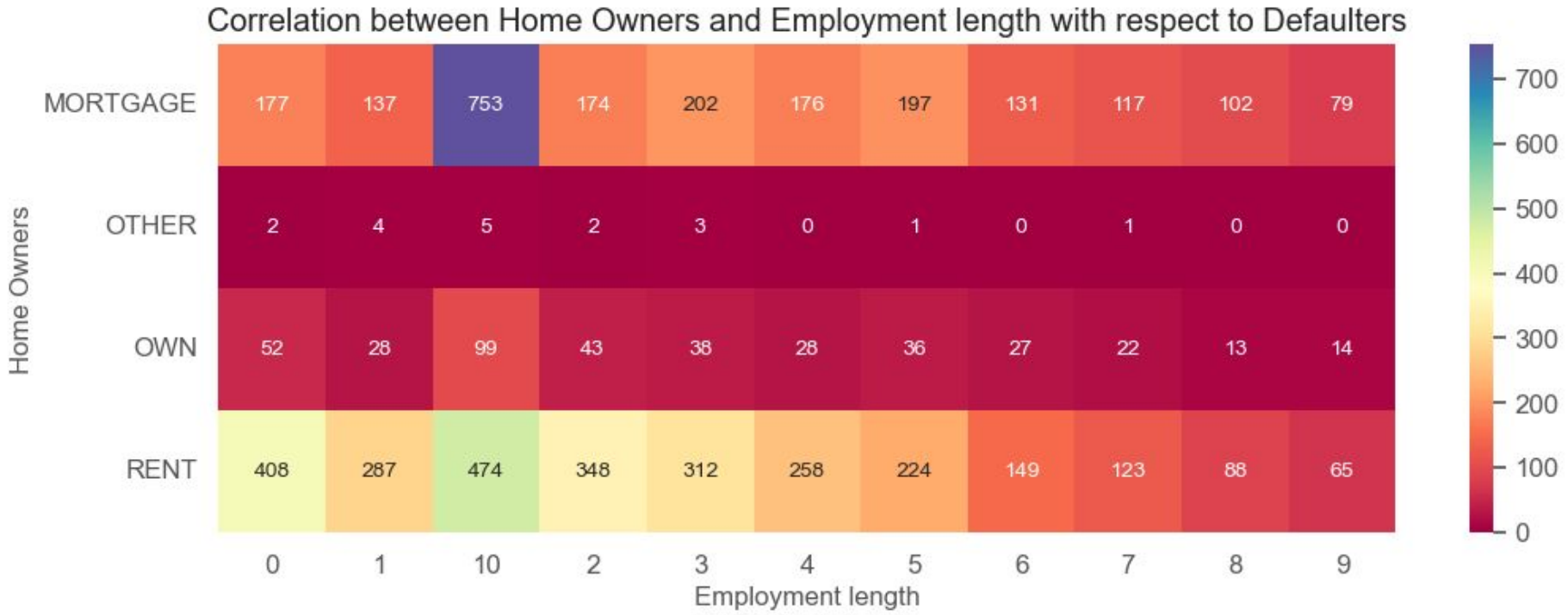
Loans status growth trend over the years and term





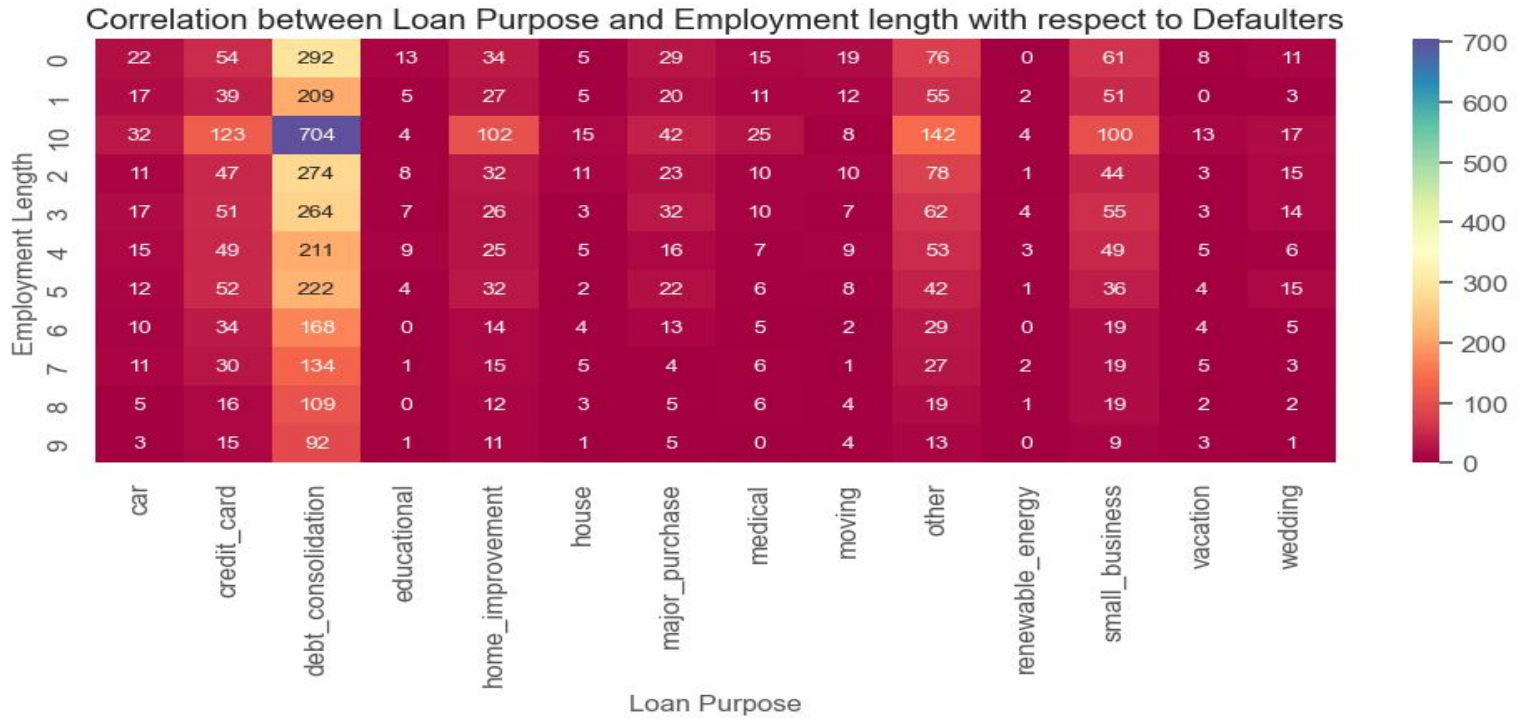
- charged off loans increased from 2007 to 2011
- Loans with a 36-month term accounted for most of the charged-off loans.

Correlation between Homeowners and Employment length with respect to Defaulters



- Applicants with more than 10 years of employment and who are either living on rent or have mortgages have the most risky profiles for defaults

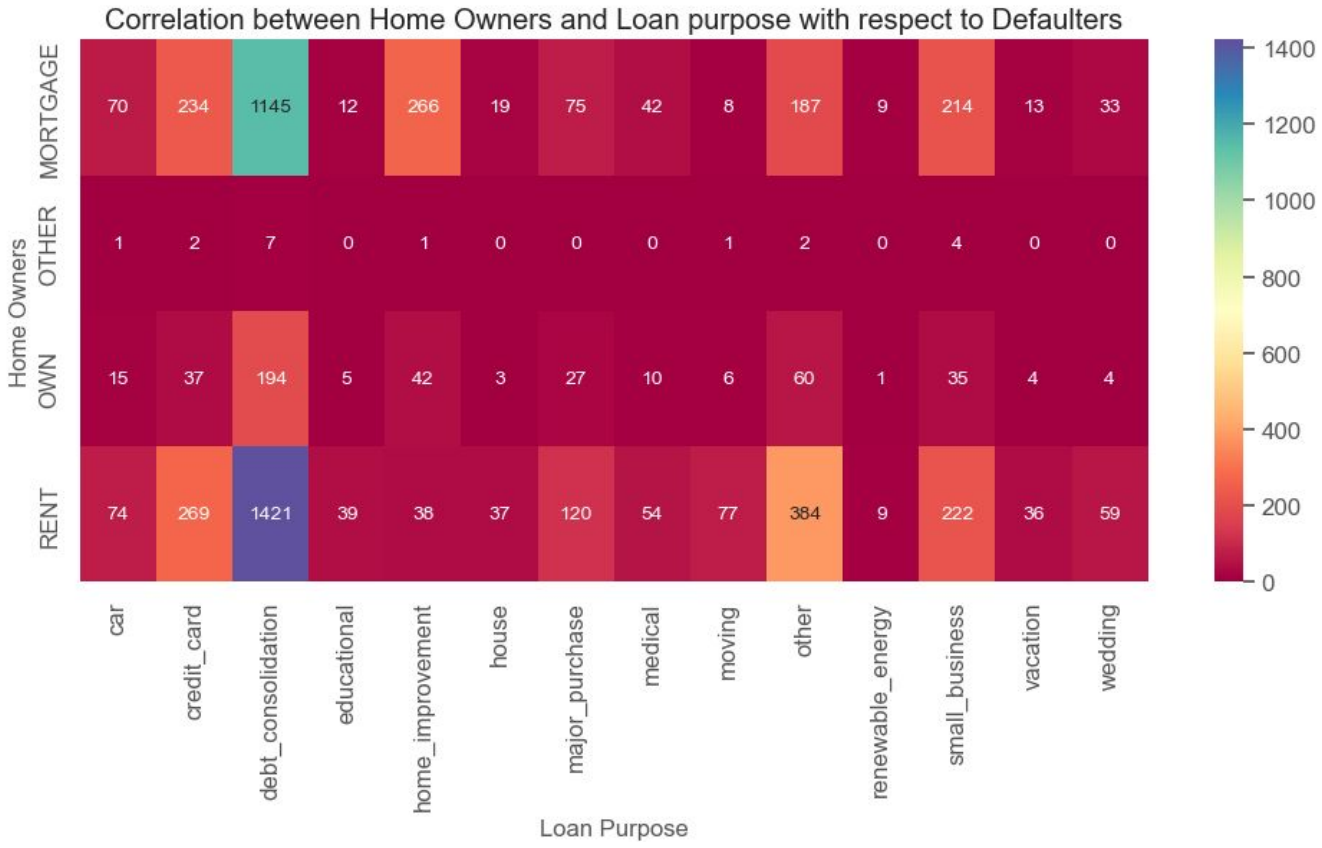
Correlation between Loan Purpose and Employment length with respect to Defaulters



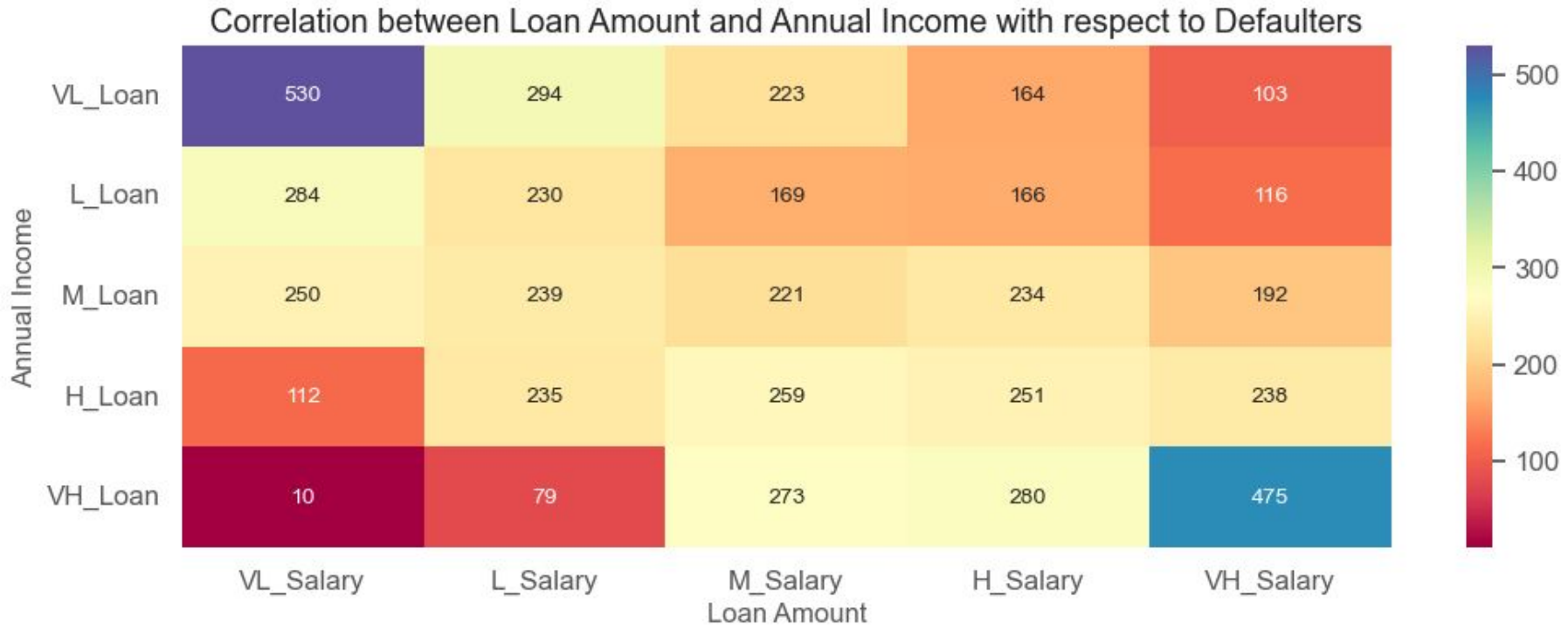
- Applicants with more that 10 years of employment who are taking loan for debt consolidation are the most risky profiles for defaults

Correlation between Homeowners and Loan purpose with respect to Defaulters

- Applicants living on rent or having mortgages who are looking to consolidate their debt are at higher risk of default

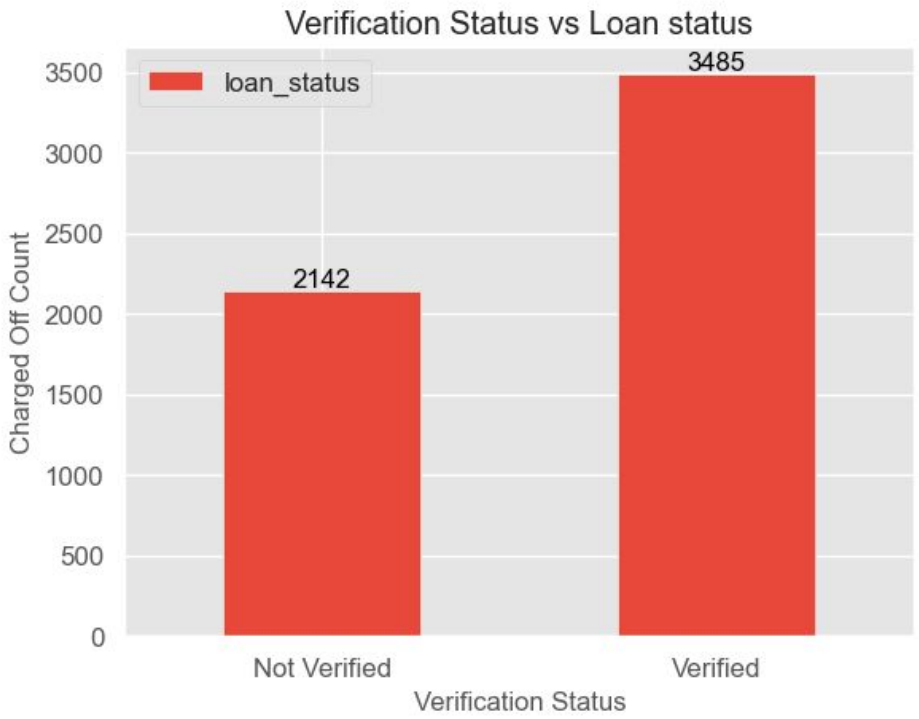


Correlation between Loan Amount and Annual Income with respect to Defaulters



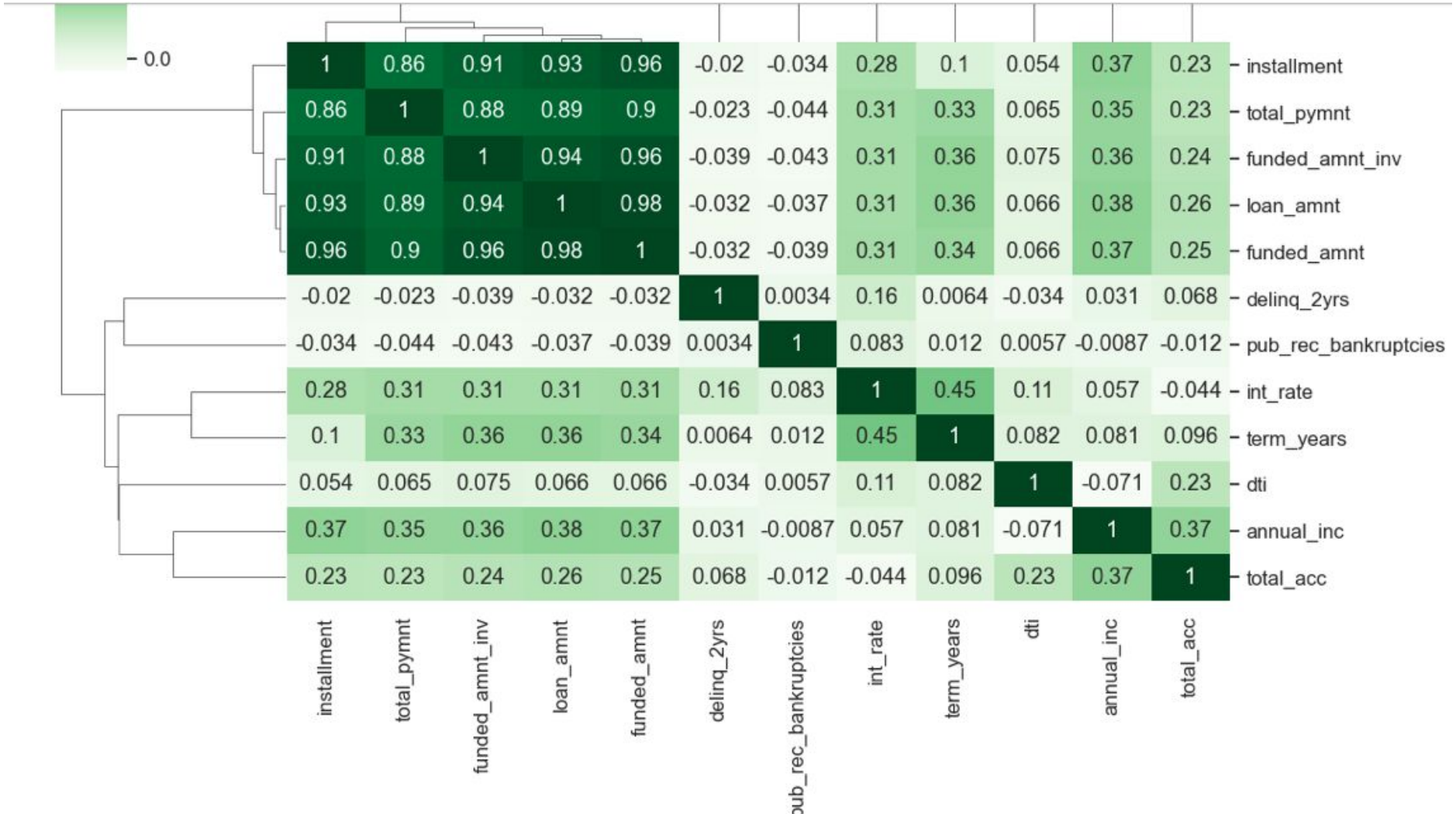
- applicants with salary less than 35000 and higher than 75600 are risky for default.
- applicants taking loan less than 5000 or higher than 19200 are risky for default.

Verification Status with respect to Defaulters



- There are more than 50% of the un-verified applicants that are defaulting on comparing with verified ones. So it's highly recommended to verify the loan applicants before approving the loan

Correlation Clustermap of Loan Attributes



Strong Correlation

- **Installment**: Shows a strong correlation with total_pymnt, funded_amnt_inv, loan_amnt, and funded_amnt, indicating that higher installment amounts are associated with higher total payments and larger loan amounts.
- **Term Years**: Has a strong correlation with int_rate, suggesting that longer loan terms are often associated with higher interest rates.

Weak Correlation

- **DTI (Debt-to-Income Ratio)**: Exhibits a weak correlation with all columns, indicating that variations in debt-to-income ratio do not significantly impact or are impacted by other variables.
- **Delinq_2yrs**: Shows a weak correlation with all columns, meaning the number of delinquencies in the past two years doesn't significantly influence other variables in your dataset.
- **Pub_rec_bankruptcies**: Also has a weak correlation with all columns, implying that public record bankruptcies do not have a strong relationship with the other factors in the dataset