# The Lending Club Case Study

By

Projesh Kumar Mondal

Rajesh Kumar Nakka

# Problem Statement

**Business Understanding**

Lending Club is 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

## Objective

Borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'. If risky loan applicants are identified early, then such loans can be reduced thereby cutting down the amount of credit loss.

Identification of such risky applicants using Exploratory Data Analysis (EDA) is the aim of this case study.

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 utilise this knowledge for its portfolio and risk assessment.

# Data Understanding

The dataset has 3977 rows and 111 columns. The Columns in the dataset can be broadly segregated in categories as follows:

**Loan Information:**
id, member_id: Unique identifiers for the loan and borrower.
loan_amnt: The total amount of money borrowed.
funded_amnt: Amount approved for funding by the lender.
funded_amnt_inv: Investor-contributed amount for the loan.
term: Loan term in months (e.g., 36 or 60 months).
int_rate: Interest rate charged on the loan.
installment: Monthly installment payment amount.

**Borrower Details:**
emp_title: The borrower's job title.
emp_length: Years of employment (e.g., "10+ years", "<1 year").
home_ownership: Type of homeownership (e.g., OWN, MORTGAGE, RENT).
annual_inc: Borrower's self-reported annual income.
addr_state: Borrower's state of residence.

**Derived Metrics:**
dti: Debt-to-income ratio (borrower's monthly debt payments as a percentage of income).
fico_range_low, fico_range_high: Borrower's FICO credit score range.
Lender and Investor Details:
initial_list_status: Initial listing status of the loan.
out_prncp: Remaining outstanding principal.
total_pymnt: Total amount paid by the borrower to date.

**Purpose and Description:**
purpose: The purpose of the loan (e.g., debt consolidation, home improvement).
title: Free-text field for the loan purpose.
desc: Description of the loan provided by the borrower.
Repayment Details:
last_pymnt_d, last_pymnt_amnt: Date and amount of the last payment.
next_pymnt_d: Expected date of the next payment (if applicable).
total_rec_prncp, total_rec_int: Total principal and interest received to date.

**Credit and Payment History:**
revol_util: Revolving credit utilization percentage (percentage of available credit used).
earliest_cr_line: Date of the borrower's earliest credit line.
open_acc: Number of open credit accounts.
total_acc: Total number of credit accounts.
pub_rec: Number of derogatory public records (e.g., bankruptcies).
Loan Status:
loan_status: Status of the loan (e.g., Fully Paid, Charged Off, Current).
pymnt_plan: Indicates if the borrower is on a payment plan.
Purpose and Description:

# Data Cleaning

**Row Analysis**
- No duplicate rows found
- No Extra Columns like Header and Footer
- Rows with "loan_status" of "Current" Dropped

**Handling Missing Values**

After doing column deletion, three columns were still having blanks (<3%).
- pub_rec_bankruptcies: Imputed with mode.
- revol_util: It has no outliers. Imputed with mean.
- emp_length: We assume that they are self. Imputed with node.

**Derived Columns**
- issue_d broken down into month and date.
- loan_amnt, int_rate, annual_inc, dti, revol_unit segmented using buckets.

**Column Deletion**
- 57 columns with more than 60% blank data deleted
- 9 Columns with single unique values deleted
- 24 Columns deleted because, either they are not relevant for the analysis, or they don't have proper data.

**Summary: Out of 111 columns, 90 are deleted**.

**Converting Column format and data types**
- Remove % from int_rate, revol_util and convert it to float type.
- Convert the column issue_d from string to DateTime.
- Change dtype of emp_length from Object to Category

**Outlier Handling**
- annual_inc: top 5 percentile rows removed.

# Univariate Analysis



Count plot for grade



Count plot for sub_grade



Count plot for emp_length



Count plot for issue_d_year
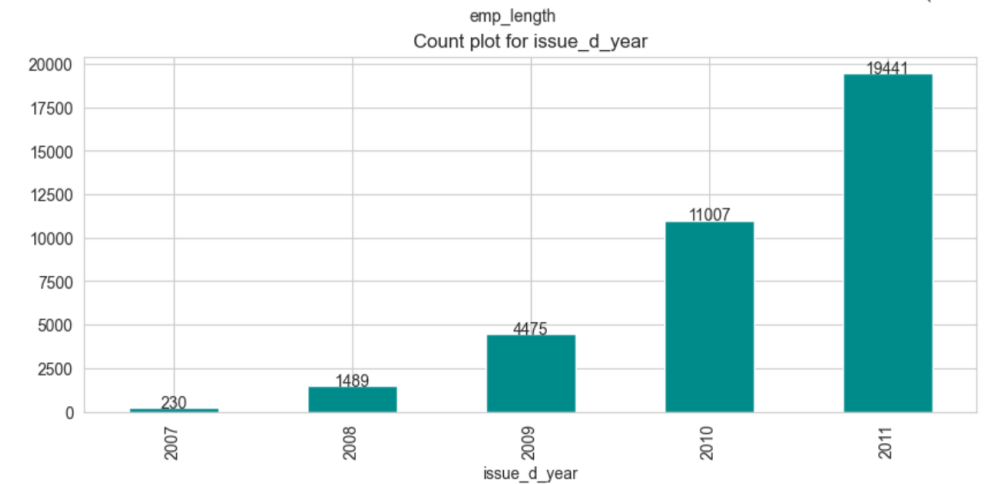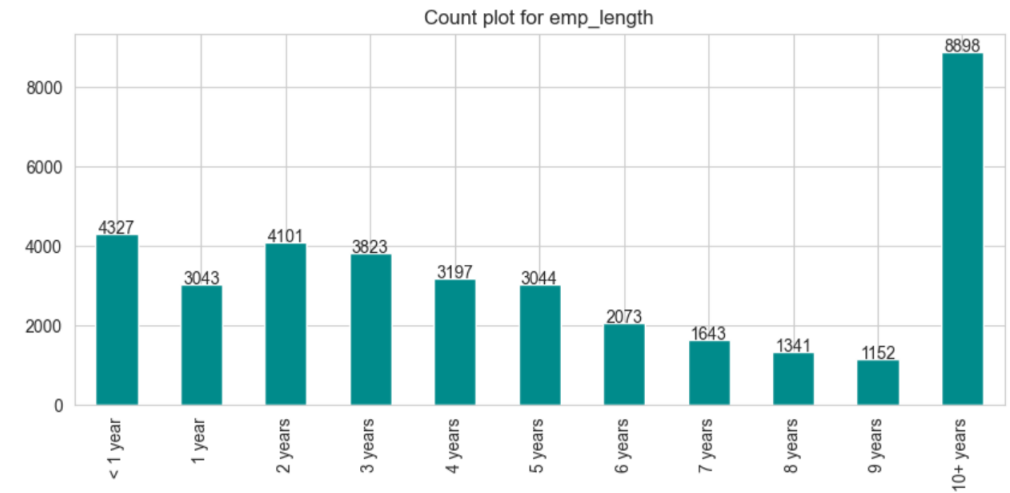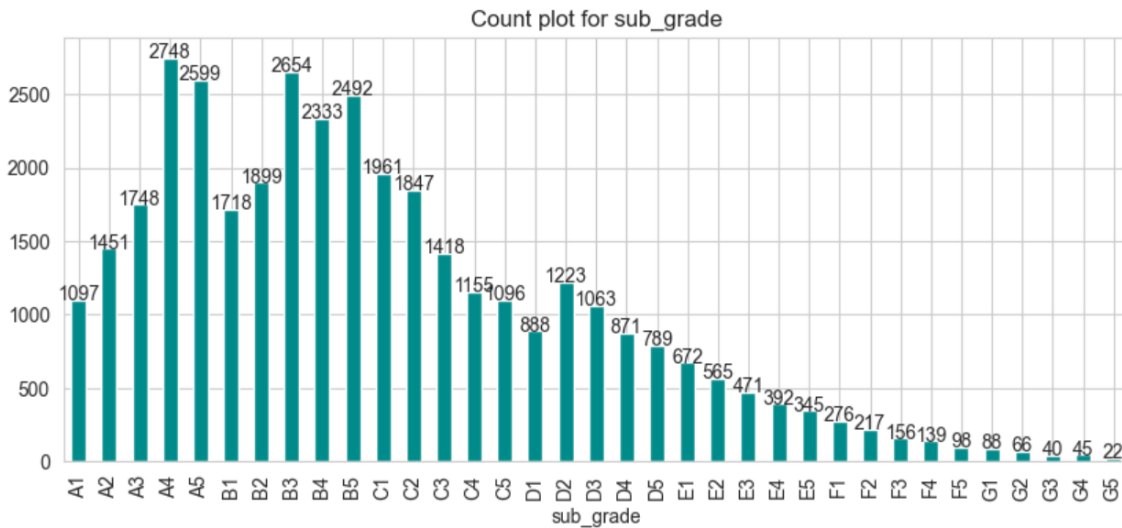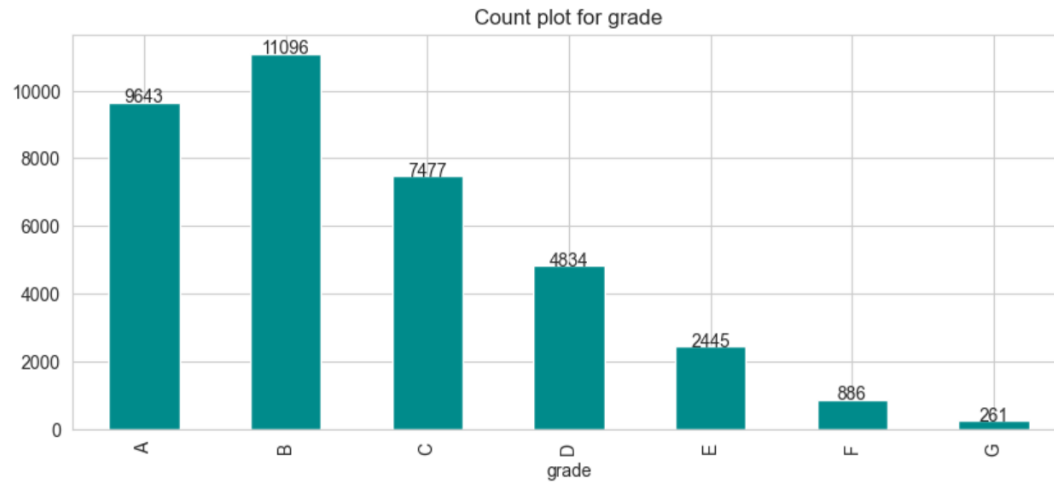
**Observations**
- Highest Number of Customers are having Grade B followed by Grade A. Followed by C,D, E, F and G. G having the lease Number of Customers. This maybe because the company is cautious towards providing loan to persons having lower Grades.
- sub_grade also have similar trend, except that there is a sudden rise in A4, A5 and a dip in B2, B3. For the Employment length, 10+ years bracket has the highest customers.
- No of loans issues per year has increased considerably year on years from 2007 to 2011

# Univariate Analysis
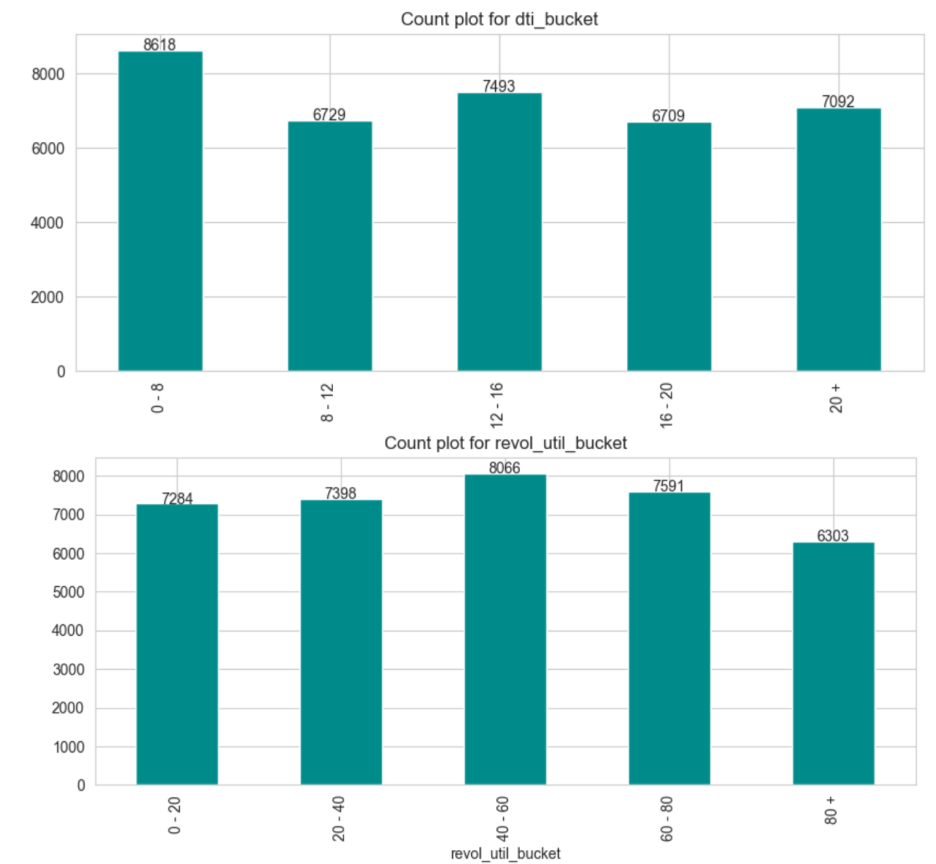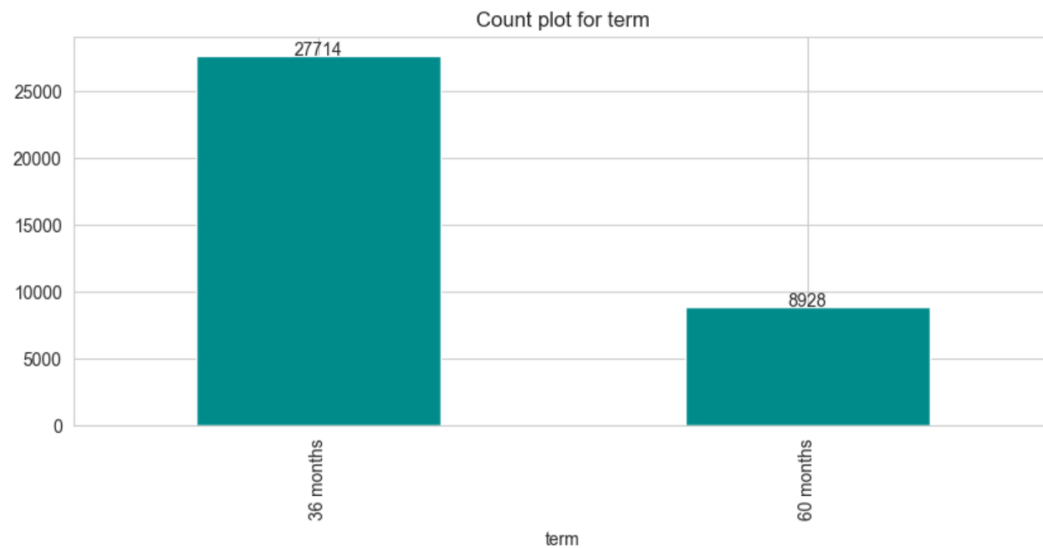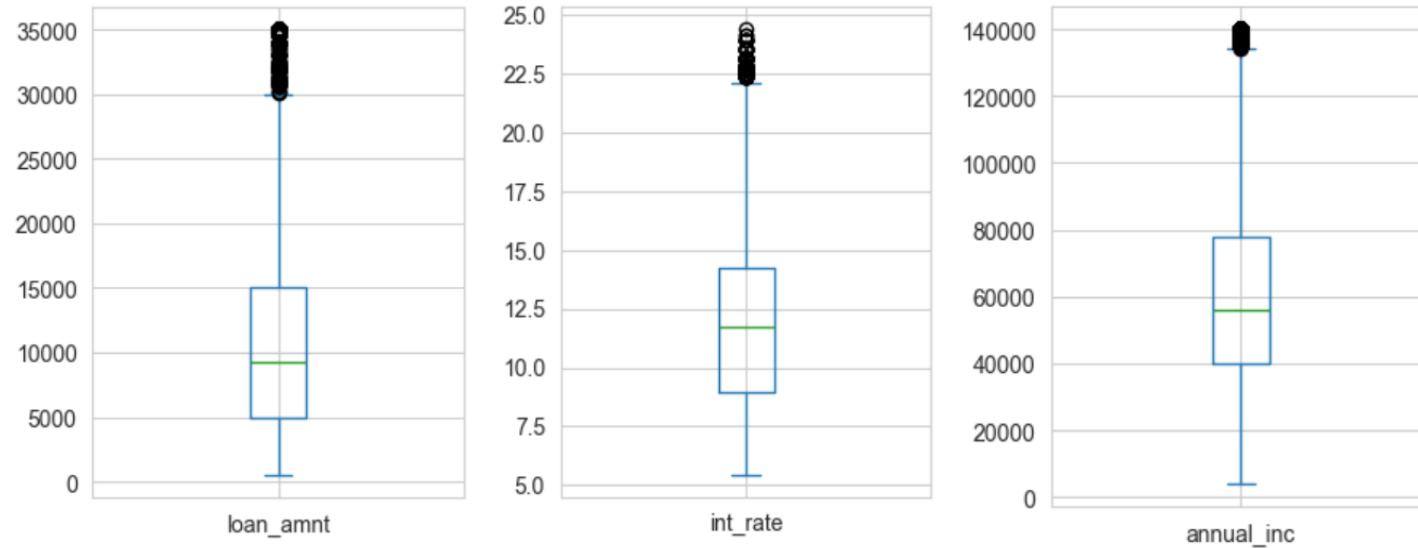


**Observations**
- Most customers takes loan between 5000 - 15000 USD.
- Most customer has interest rate between 9-14%
- Income of most customers falls between 40000 and 80000 USD.
- dti and revol_util are spread evenly along the range
- Most of the customers opt for 36-month loan. .

# Univariate Analysis



**Observations**
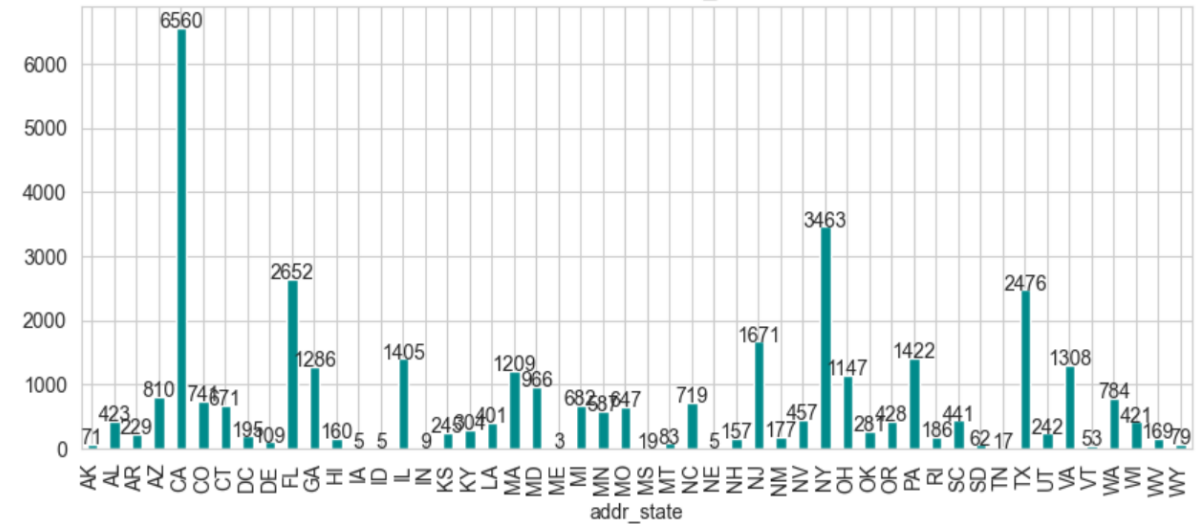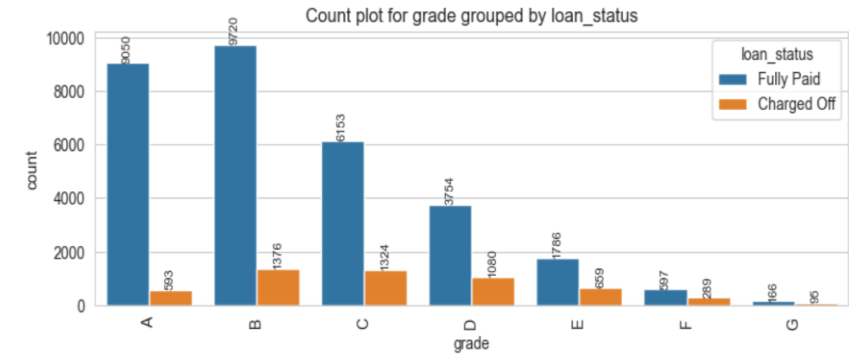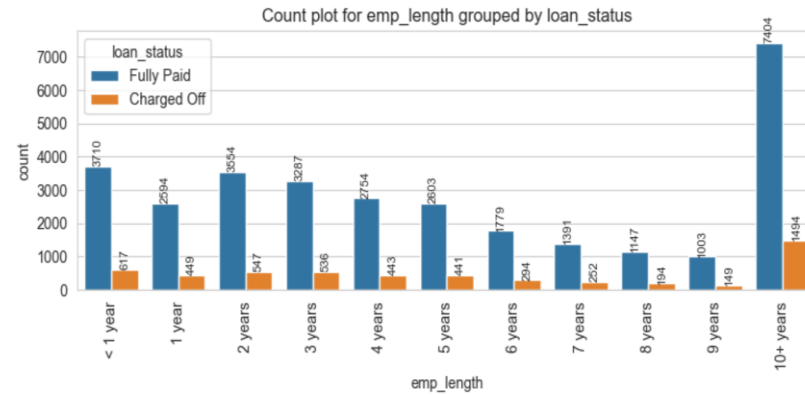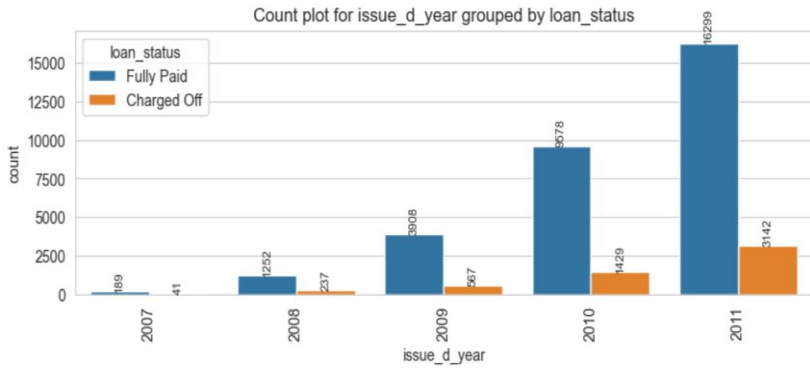- Customers having Own home is very low.
- Most of the customers are not fully verified.
- Most of the loans are for debt consolidation
- CA, NY and FL are the three highest customer states.

# Segmented Univariate Analysis
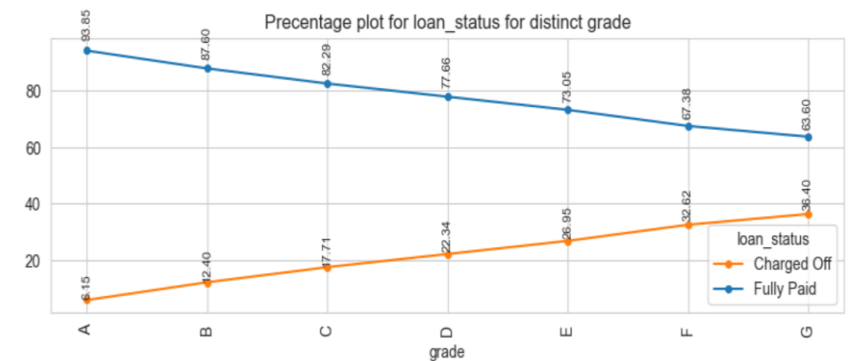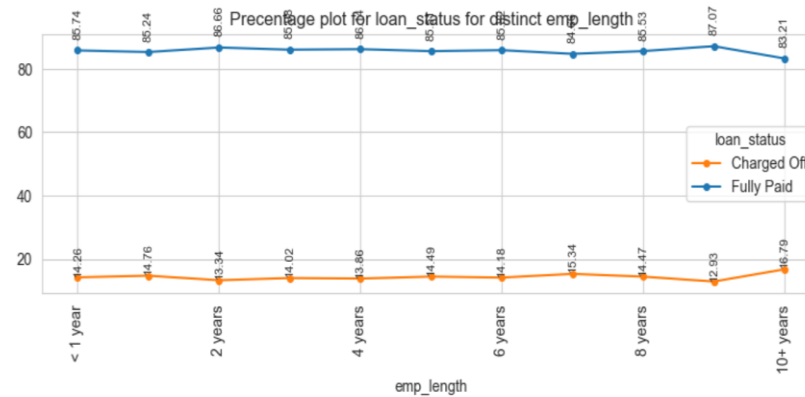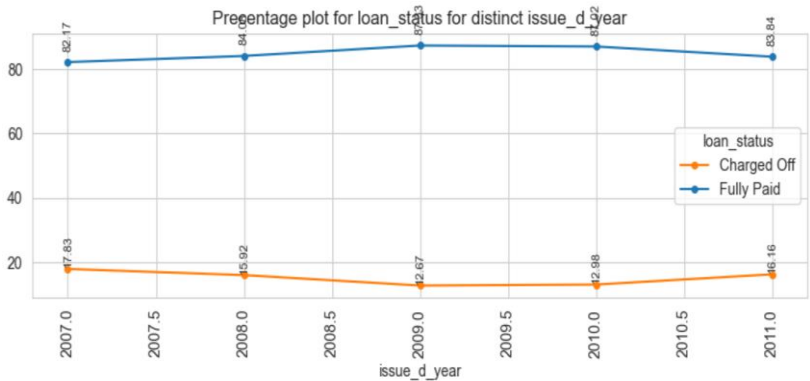


**Observations**
- Grade B consist of the highest number of Charged Off customers. But the percentage of charged off customers increases as the grade decreases indicating a stong correlation of percentage of charged off customers with grades. Customer with higher grades are less likely to default.
- For employment length, percentage of customer defaulting has similar trend across employment length buckets but increases a but increases a but for 10+ years category. Company should be more vigilant while providing loans to 10+year and may increase the interest rate a bit for them.
- Percentage of charged off customer has increased a bit for loans issued in 2011.

# Segmented Univariate Analysis
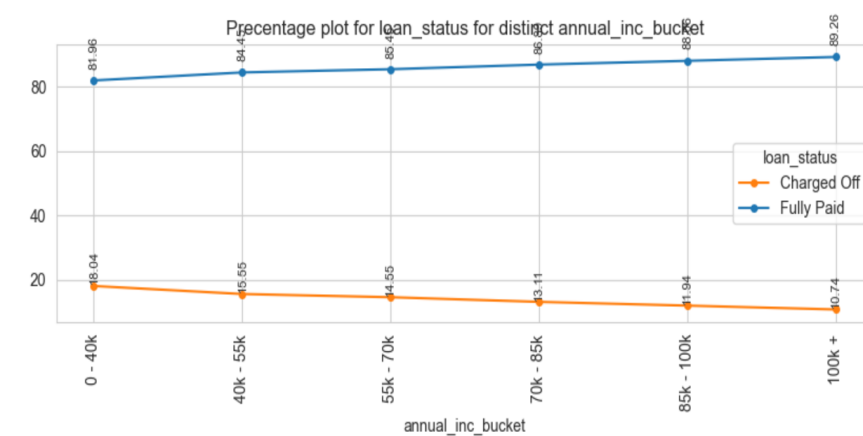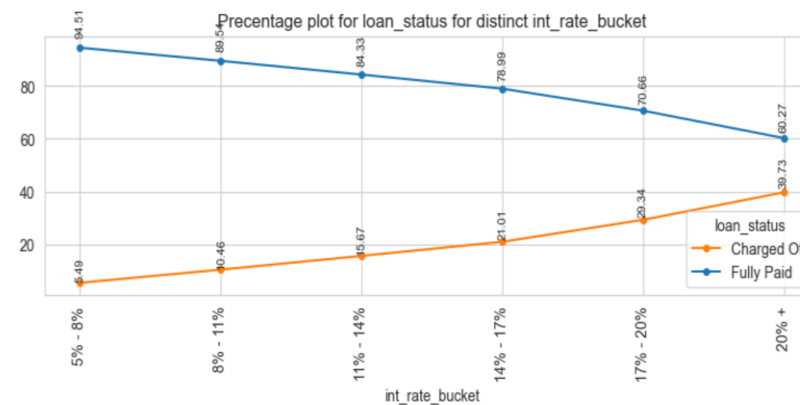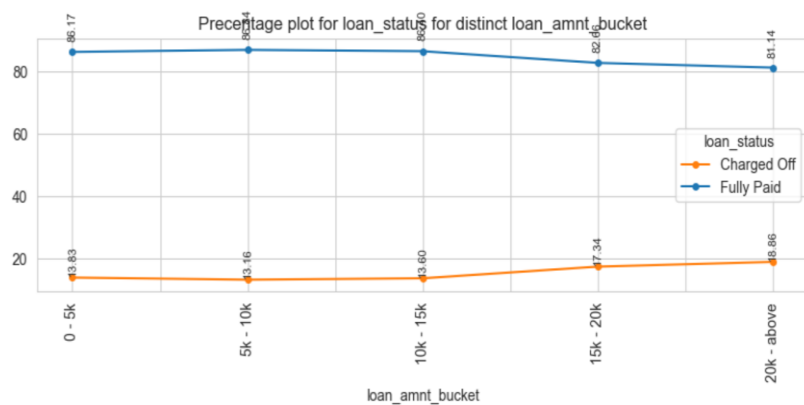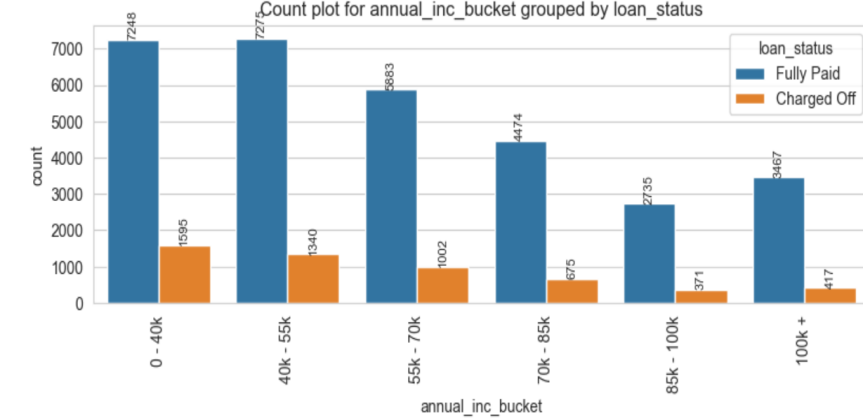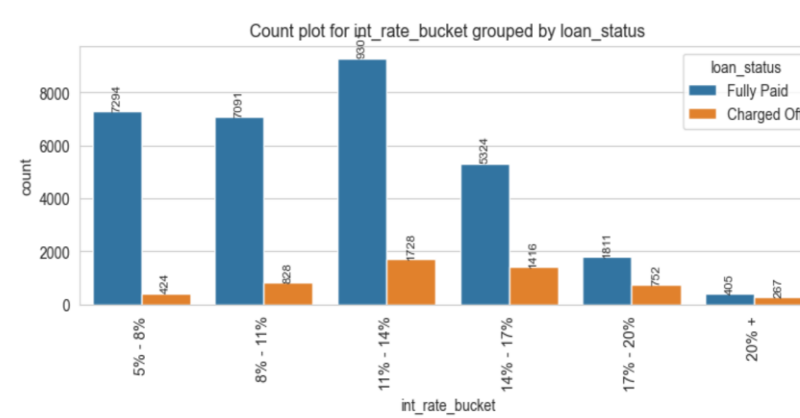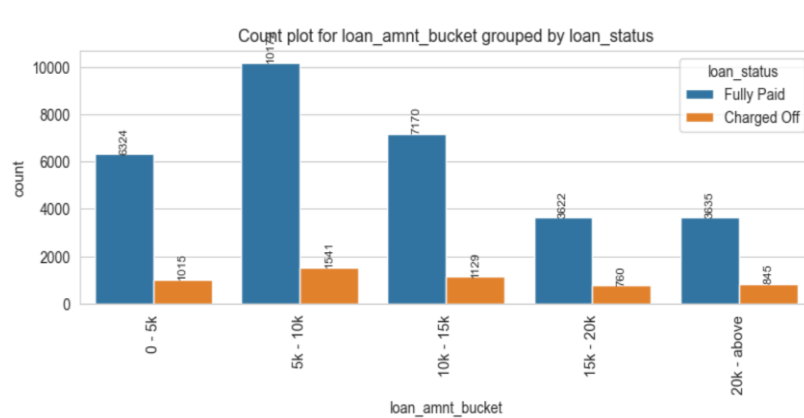


**Observations**
- The loan amount 5k-10k sees the largest number of defaulters. But probability of defaulting increases with loan amount.
- Percentage of defaulters increase with increase in interest rate.
- Percentage of defaulters decreases with increasing annual income. Company should push for customer with higher income.
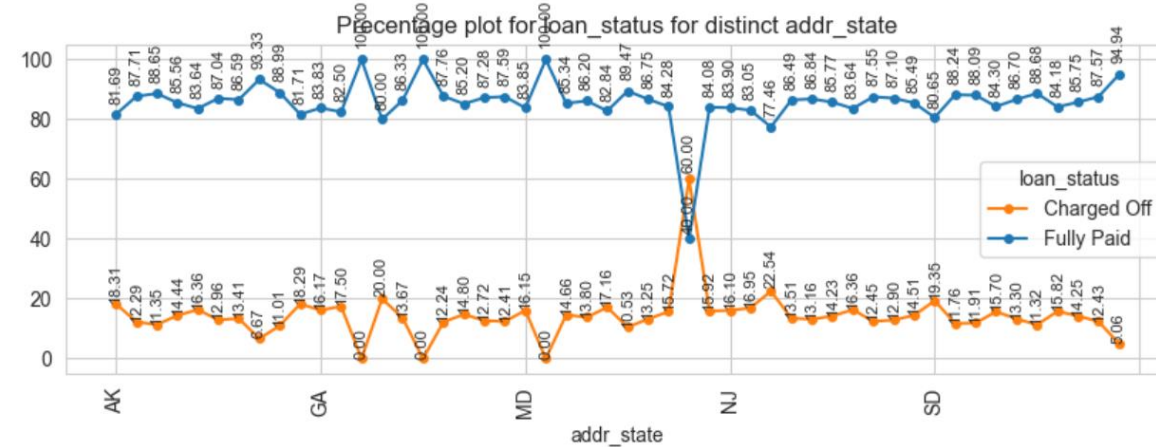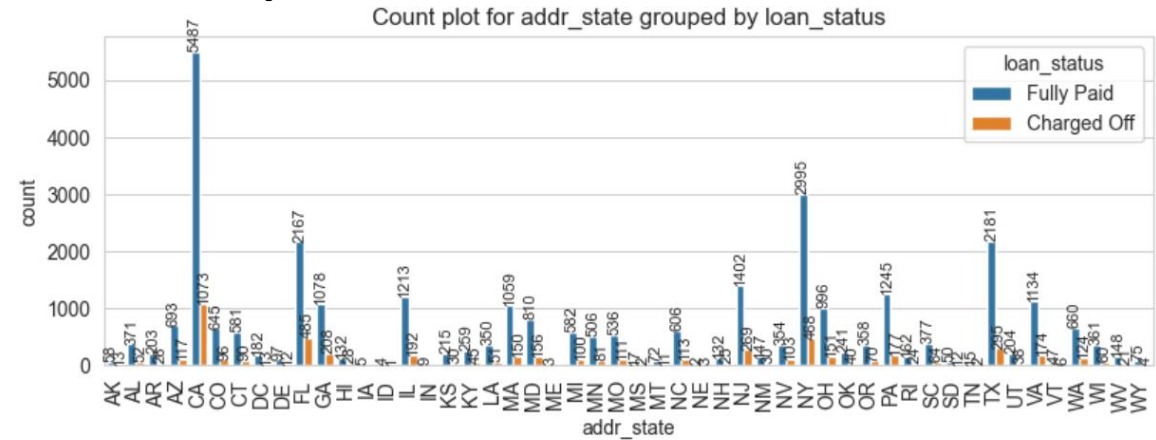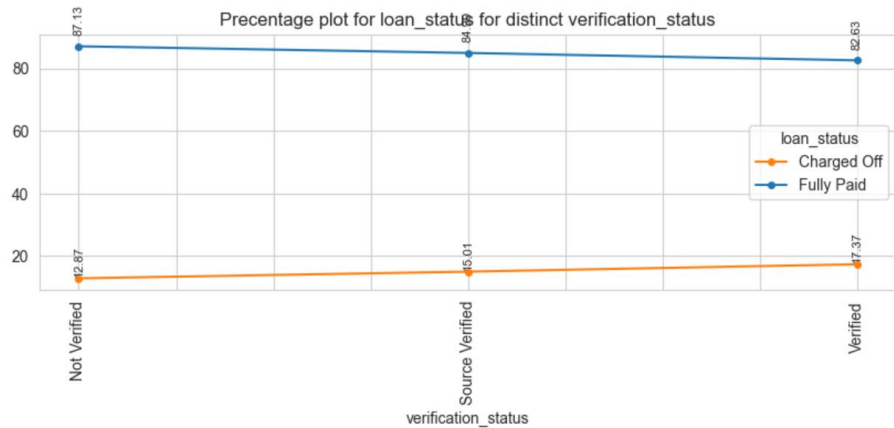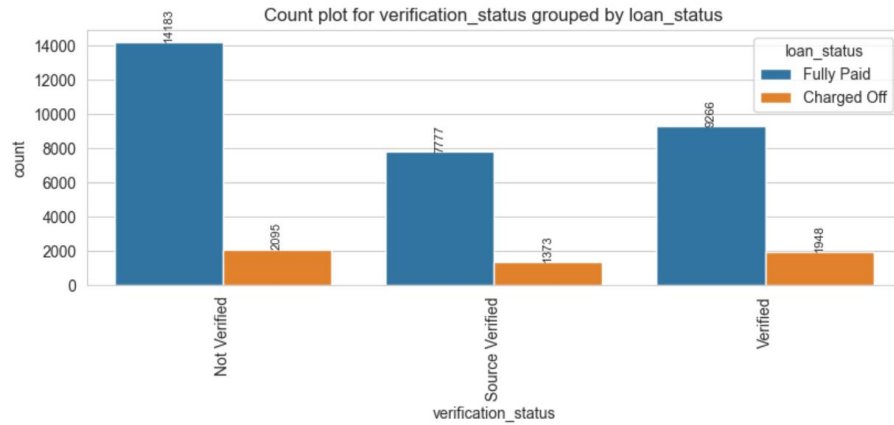
# Segmented Univariate Analysis
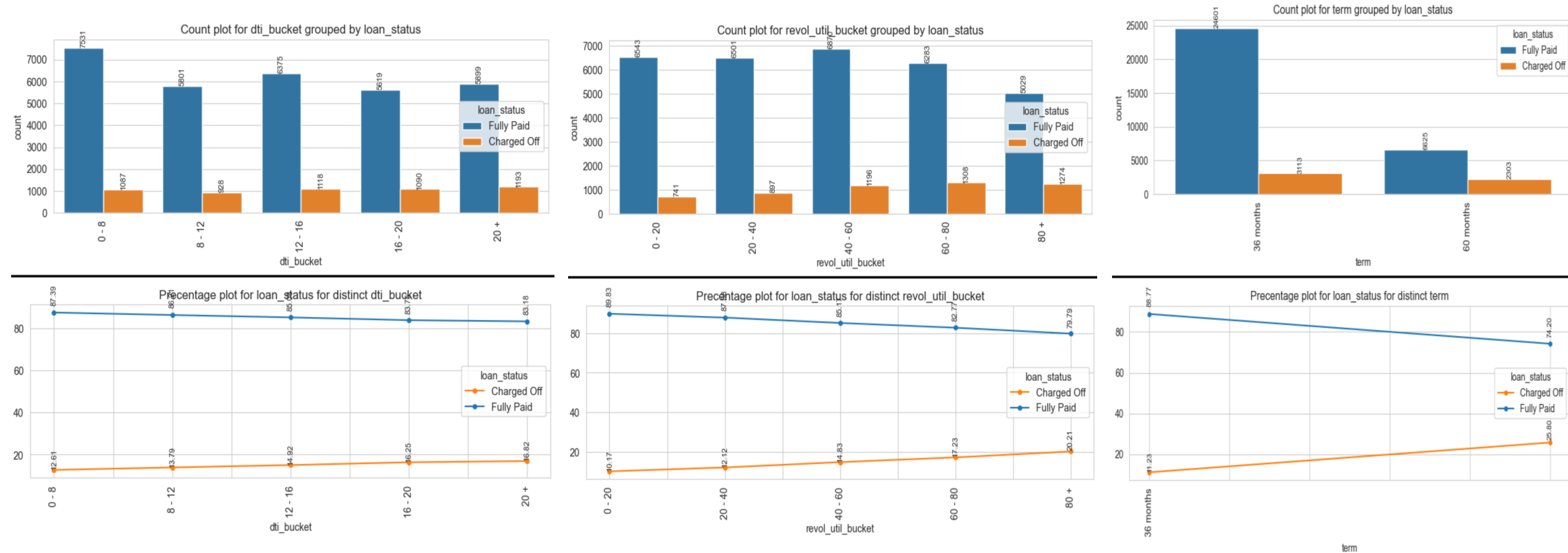


**Observations**
- Percentage of defaulters increase with increase in debt to income ratio (dti)
- Customer with higher revol_util has more probablity of defaulting. Company should decrease the revolving credit limit in that case.
- Customer opting for 60 months term are more likely to default.

# Segmented Univariate Analysis



Count plot for verification_status grouped by loan_status



Count plot for addr_state grouped by loan_status



Precentage plot for loan_status for distinct verification_status



Precentage plot for loan_status for distinct addr_state
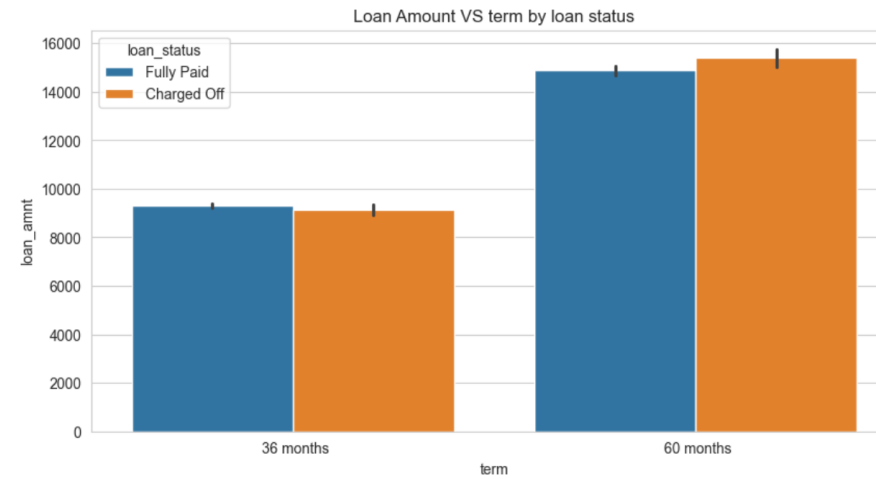
**Observations**
- verified customers are actually more likely to default.
- Debt consolidation has the highest number of defaulters. But percentage of defaulters are highest for small businesses.

# Bivariate Analysis
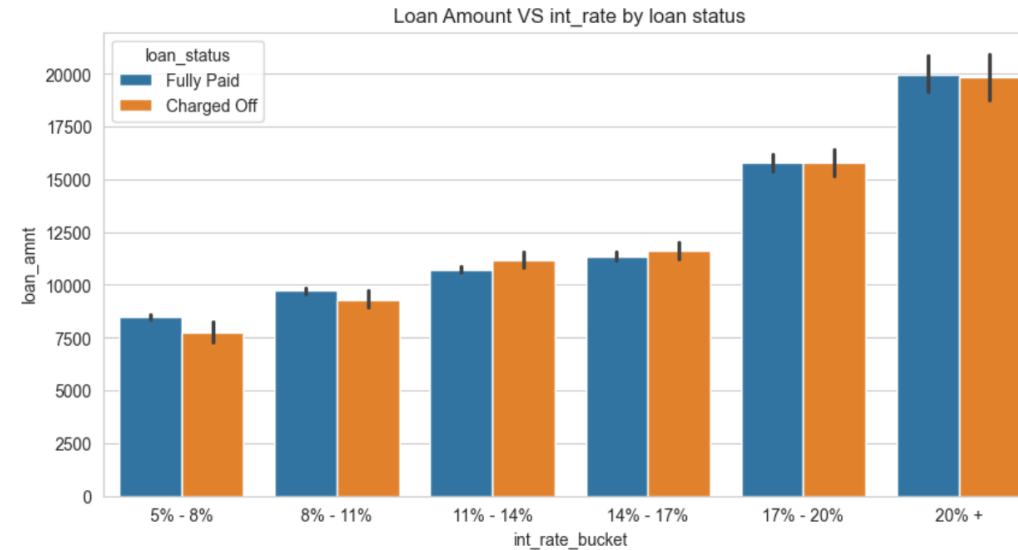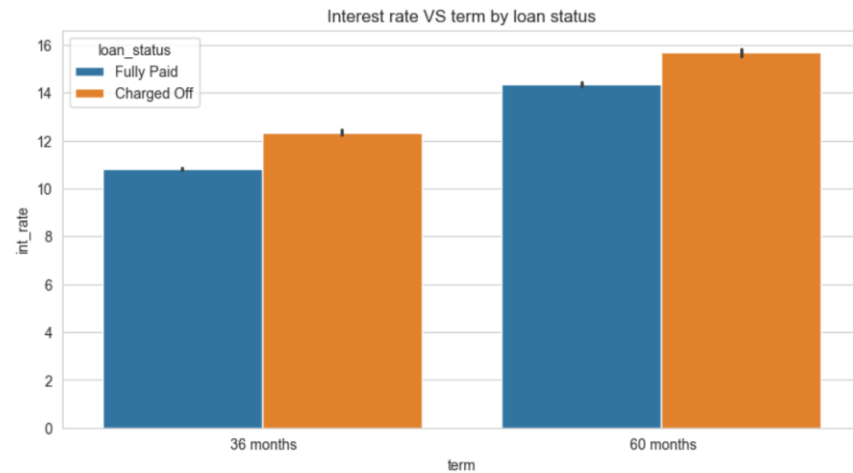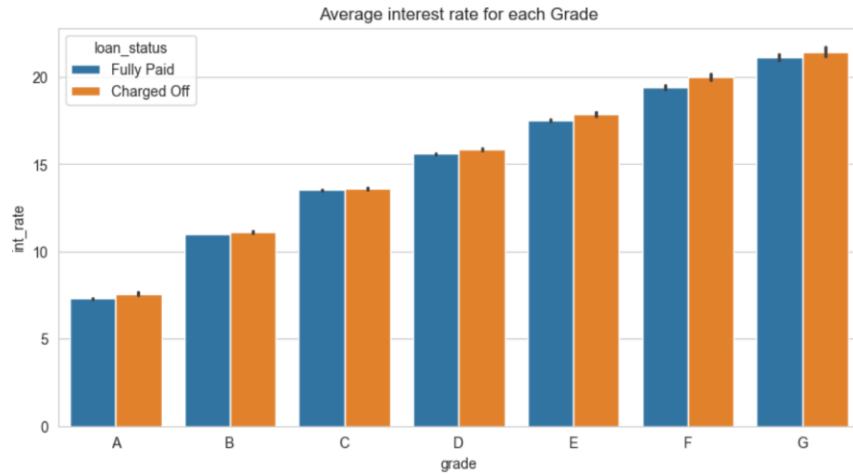


**Observations**
- Loan provided with an average of 12.5% interest rate for 36months of tenure is more likely to Default- Loan provided with an average of 16.5% interest rate for 60months of tenure are likely to default
- Customer with 60 months term has larger loan amount
- Interest Rate has a positive Correlation loan amount

# Bivariate Analysis
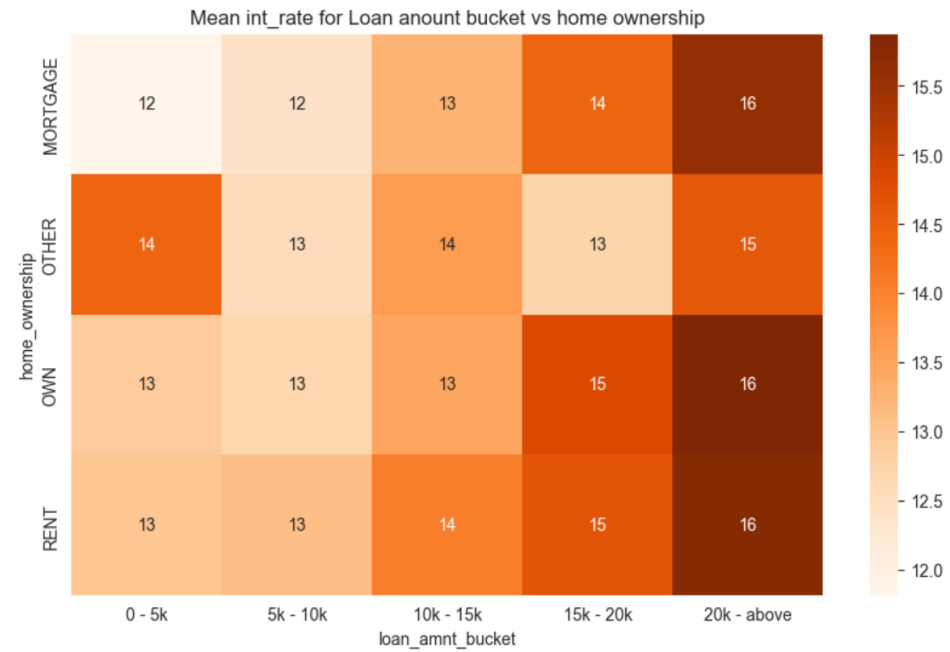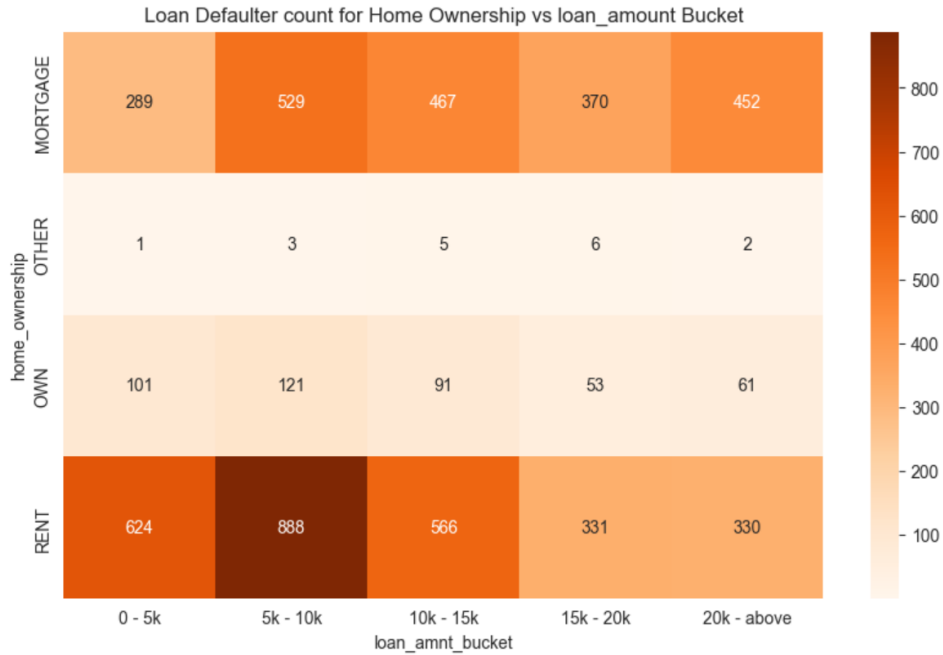


Loan Defaulter count for Home Ownership vs loan_amount Bucket



Mean int_rate for Loan anount bucket vs home ownership



Loan Defaulter count for Home Ownership vs annual income Bucket
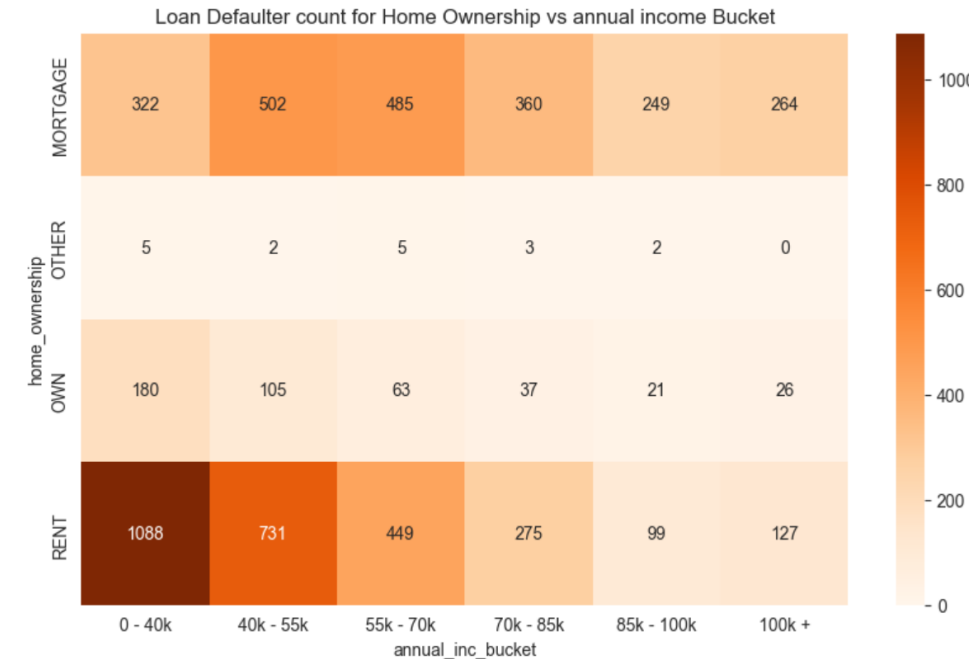
**Observations**
- Customer living in Rent and loan amount less than 15K are highly like to default
- Interest Amount Increases for higher loan bucket.
- Customer living in rent or having mortgage properties and having Annual income <80K is very likely to Default.

# Bivariate Analysis



Distribution of loan amount vs Verification Status by loan status



dti vs purpose by loan status
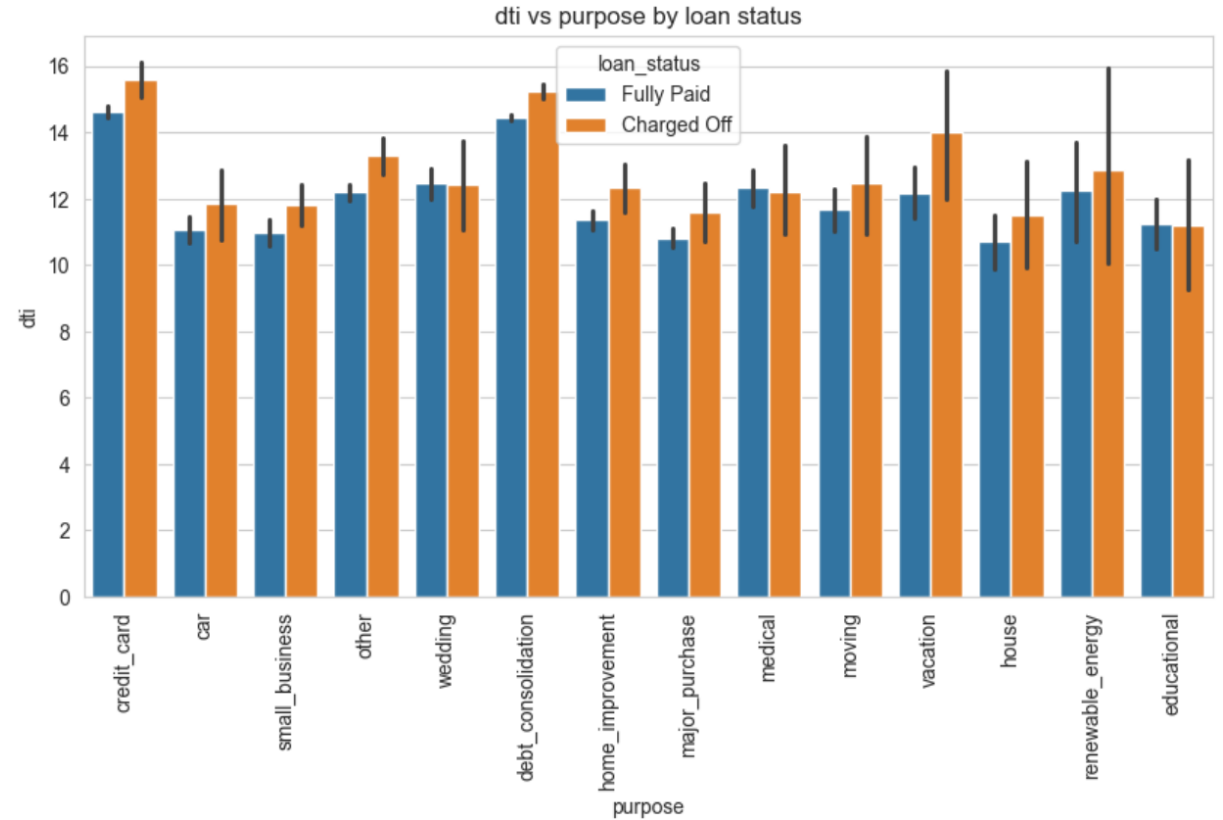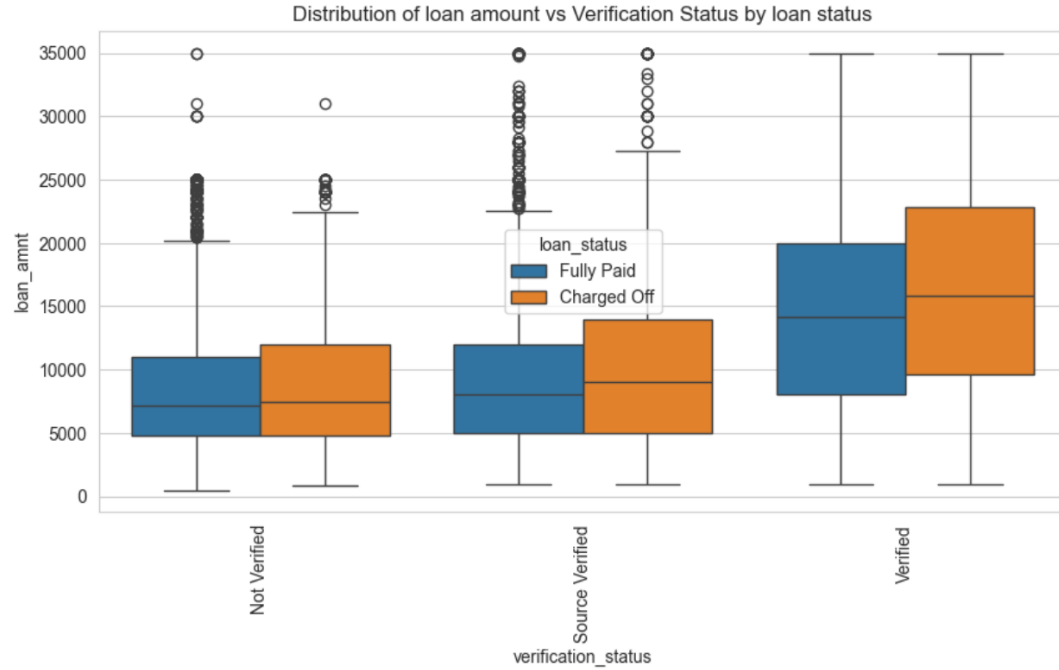
**Observations**
- Unverified/Source Verified Customer can generally take loan up to 14k.
- Customer opting loan for Credit card and debt consolidation has higher dti value, hence more porbable to default

# Bivariate Analysis



Interest Rate vs purpose



Loan Defaulter count for Home Ownership vs annual income Bucket
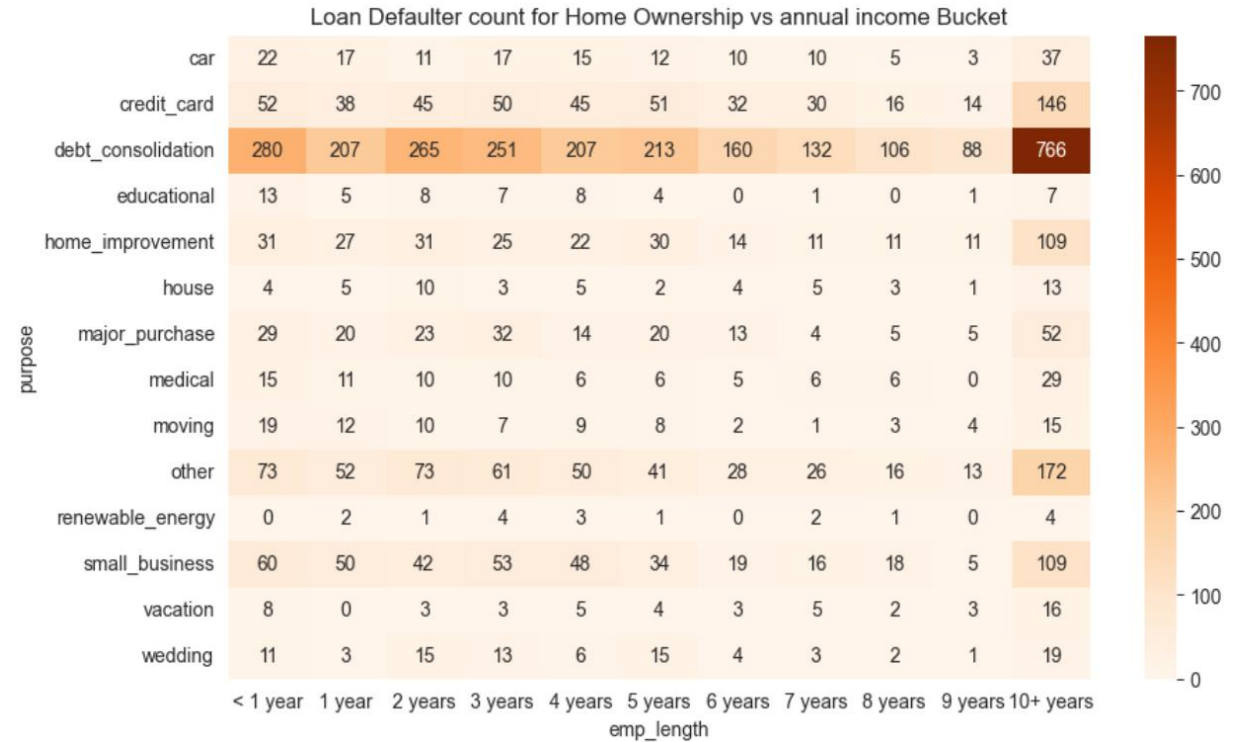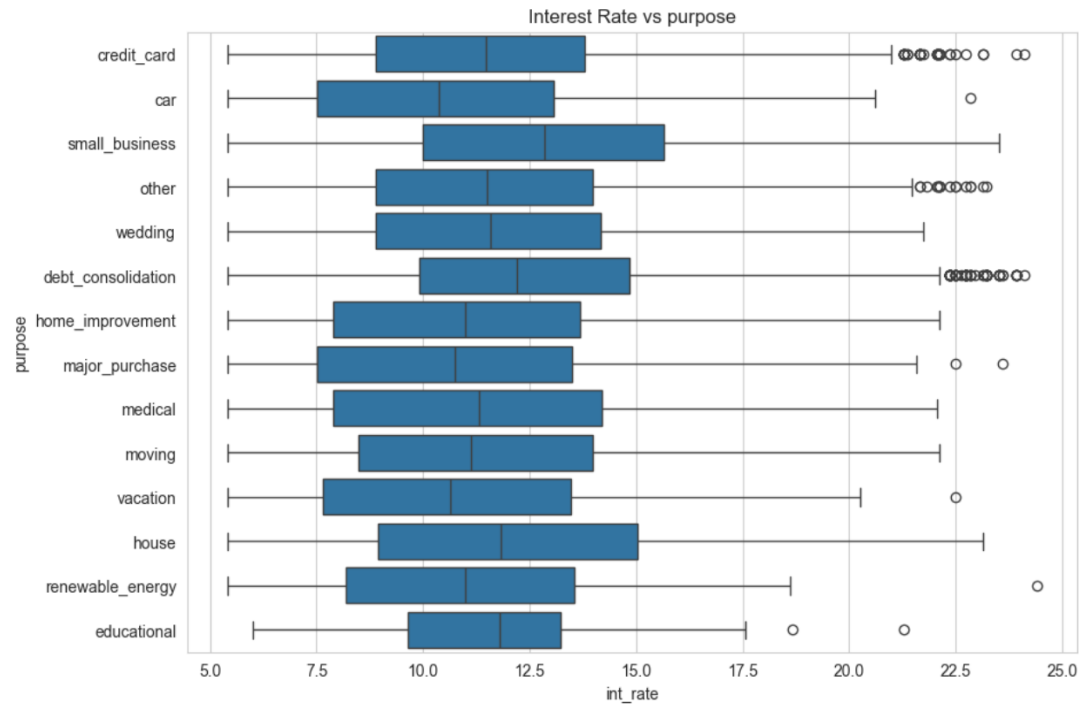
**Observations**
- Different loan purpose had different interest date distribution. The interest rate of small businesses being the maximum.
- Customers with employment period more than 10 years are taking most of the loans for Debt consolidation and are also the highest contributer in small business by a good margin

# Multivariate Analysis
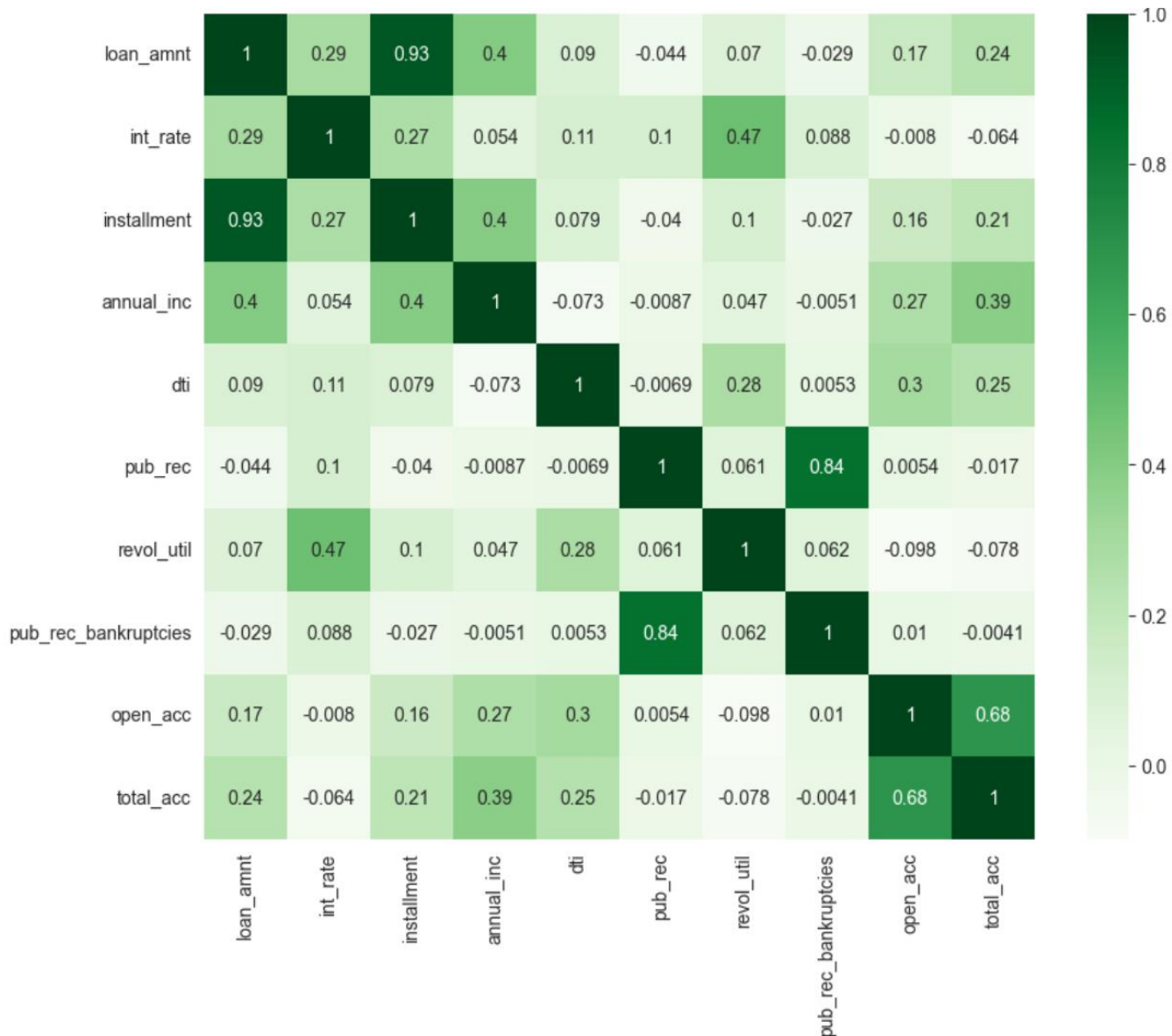


**Observations**

**Strong positive correlation:**
- loan_amnt, installment
- pub_rec_bankruptcies, pub_rec- open_acc, total_acc

**Moderately positive Correlation:**
- loan_amnt, lotal_acc, open_acc, annual_inc, int_rate
- int_rate, revol_util, installment, loan_amnt
- installment, open_acct, annual_inc
- annual_inc, total_acc, open_acc
- dti, open,acc, total_acc, revol_util

# Key Observations

- Customer living in Rent and loan amount less than 15K are highly like to default
- Debt consolidation has the highest number of defaulters. But percentage of defaulters are highest for small businesses.
- Customer opting loan for Credit card and debt consolidation has higher dti value, hence more porbable to default.
- Different loan purpose had different interest date distribution. The interest rate of small businesses being the maximum
- Verified customers are actually more likely to default.
- Customer opting for 60 months term are more likely to default.
- Percentage of defaulters increase with increase in debt to income ratio (dti)
- Higher dti value customers have probublity of defaulting more even if customer has good Grades
- Percentage of defaulters decreases with increasing annual income.
- Customers with employment period more than 10 years are taking most of the loans for Debt consolidation and are also the highest contributer in small business by a good margin
- Percentage of defaulters increase with increase in interest rate.
- Customer living in Rent and loan amount less than 15K are highly like to default

# Recommendations

- Average interest rate increases with increase in annual income. We have also seen that sustomer with more annual income default less. So, if interest rate can be reduced by 1-2% for customer with higher annual income, then we may see lesser defaulter among higher income group. Also, more customer from higher income group will opt for Lending club.
- Should keep the loan amount of unverified customers within 12K
- For customer living in rent, increase the interest amount by 2% for loans below 15k. Decrease the interest by 2% for loans above 15K.
- Given that applicants opting for 60-month loans are more likely to default, the company should consider evaluating the risk associated with longer-term loans and potentially limiting the maximum term or adjusting interest rates accordingly.
- Strict checking needs to be done in case of Customers with employment period more than 10 years and taking loan for Debt consolidation, small business.
- Company should have a stricter verification procedure, especially for higer loan amounts (>12k).

# Thank You