Lending Club Case Study

By Mamatha.K & Mallika Bera

Agenda

- Introduction
- Business Objective
- Exploratory Data Analysis and Data Cleanup
- Data Visualization and Actionable Insights.

Background

- **consumer finance company** which specializes in lending various types of loans to urban customers.
- The company must decide for loan approval based on the applicant's profile which involves risks if the applicant will likely be paying loan or will he be defaulting it.
- Analyzing such risks will help filter out risky applicants and take actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Business Objectives

Lending club case study

Business Objectives

- Providing the company **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.
- So that the company can utilize this information for its portfolio and risk assessment.

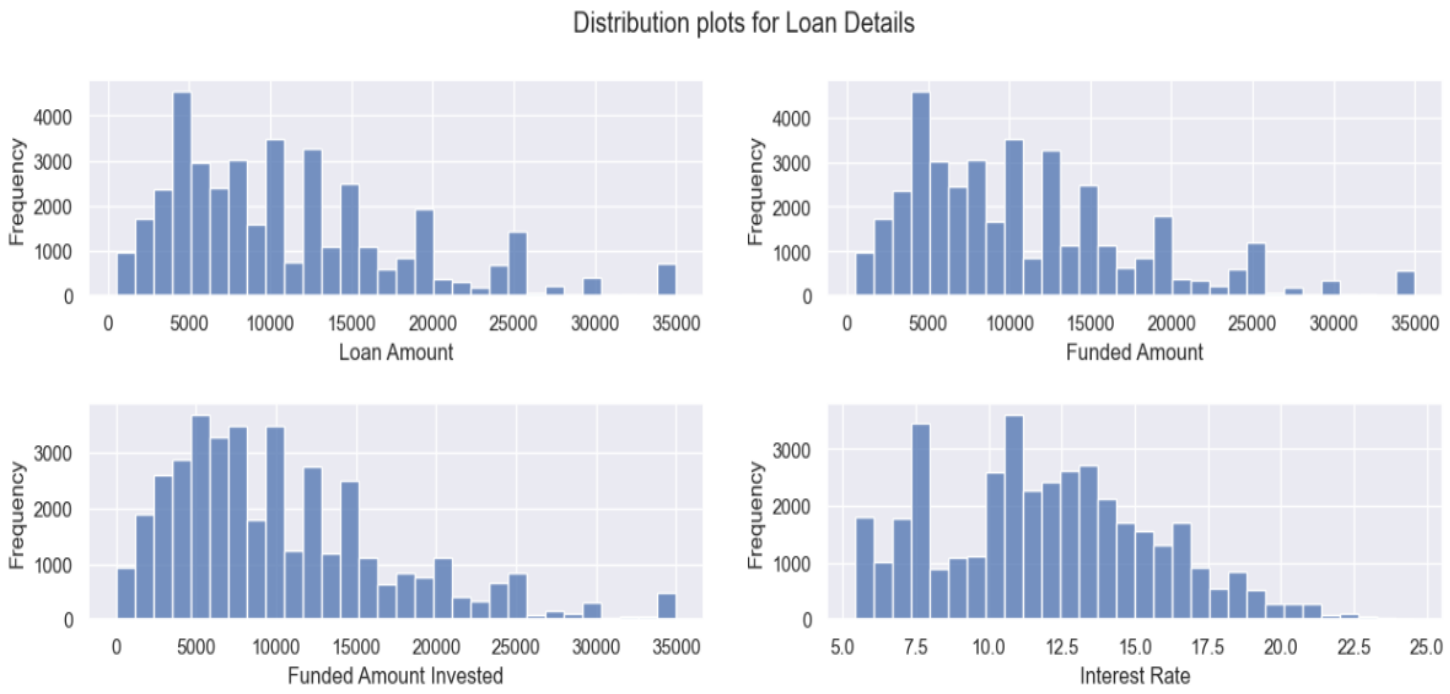
Exploratory Data Analysis - Data Cleanup

During our initial analysis of the datasheet, we could identify the following inconsistencies which need to be corrected before we start our EDA.

1. Numerical Columns are having special characters
2. Certain Columns are having too many missing values
3. Certain Columns having constant values such as 0,1,n etc.
4. Some of the columns are having incorrect data types

Univariate Analysis on Numerical variables.

- We have segregated the columns into two groups - **Numerical columns** and **Categorical Columns** based on the columns' datatypes.
- We are analyzing **numerical columns** into sub-groups based on the functional definitions



Observations

•Loan Schemes of moderate interest rates are more popular, however there is a sudden peak at interest rate 7.5% which needs further investigation

•Loan Amount, Funded Amount and Funded Amount invested has very similar distribution pattern(ignoring the outliners)

•Most loaned and funded amount is around Rs 5000

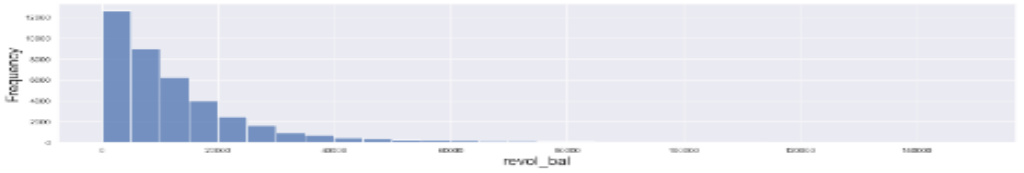
•Loan Amount and interest rate should be analyzed against 'Charged off' Loan types as they reflect meaningful patterns.

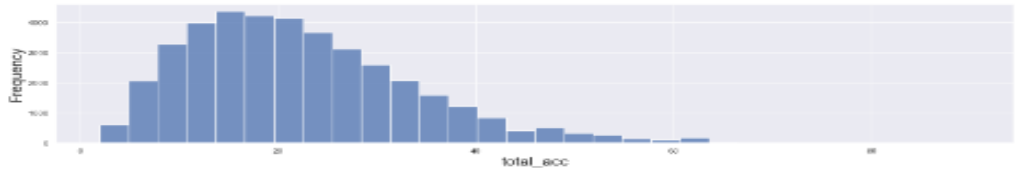
Univariate Analysis on Loan Details.

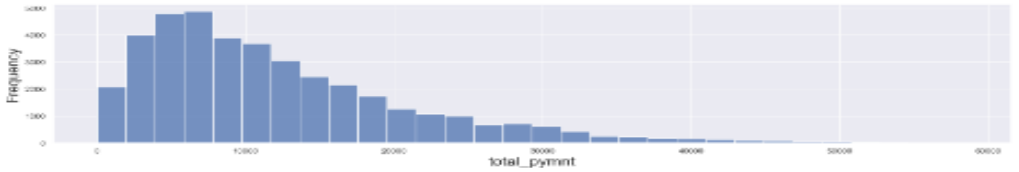
Observations:

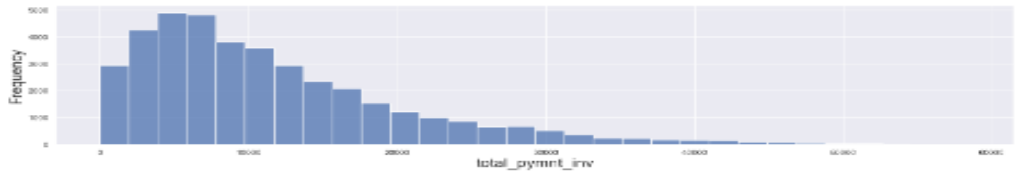
- Installments of lower amount is more in count.
- Customers with employment length > 10 years is more in number, we should investigate the charged off rate for these customer base.
- For majority annual income is around 50K. We can check if it has any specific trend for defaulting customers.
- Loan Schemes of moderate interest rates are more popular.
- Most of the customers have higher DTI which can be a risk factor. This should be further investigated.

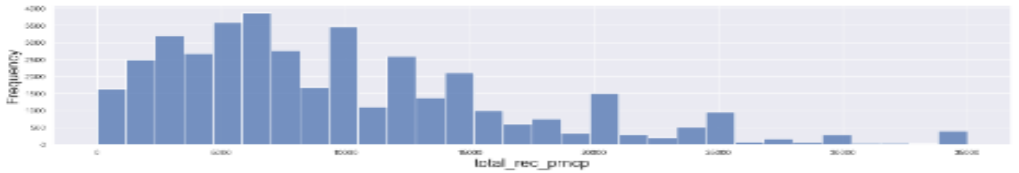
Univariate Analysis on Customer's Financial Capacity

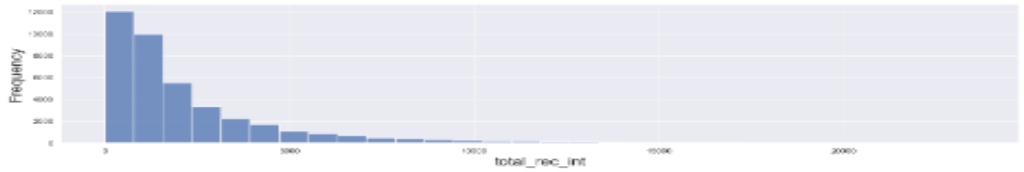


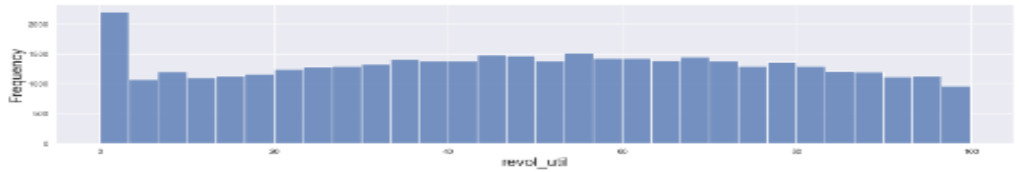


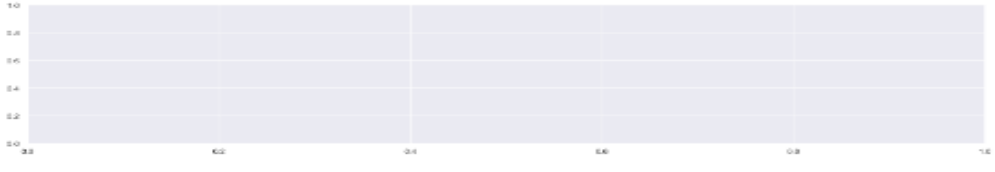












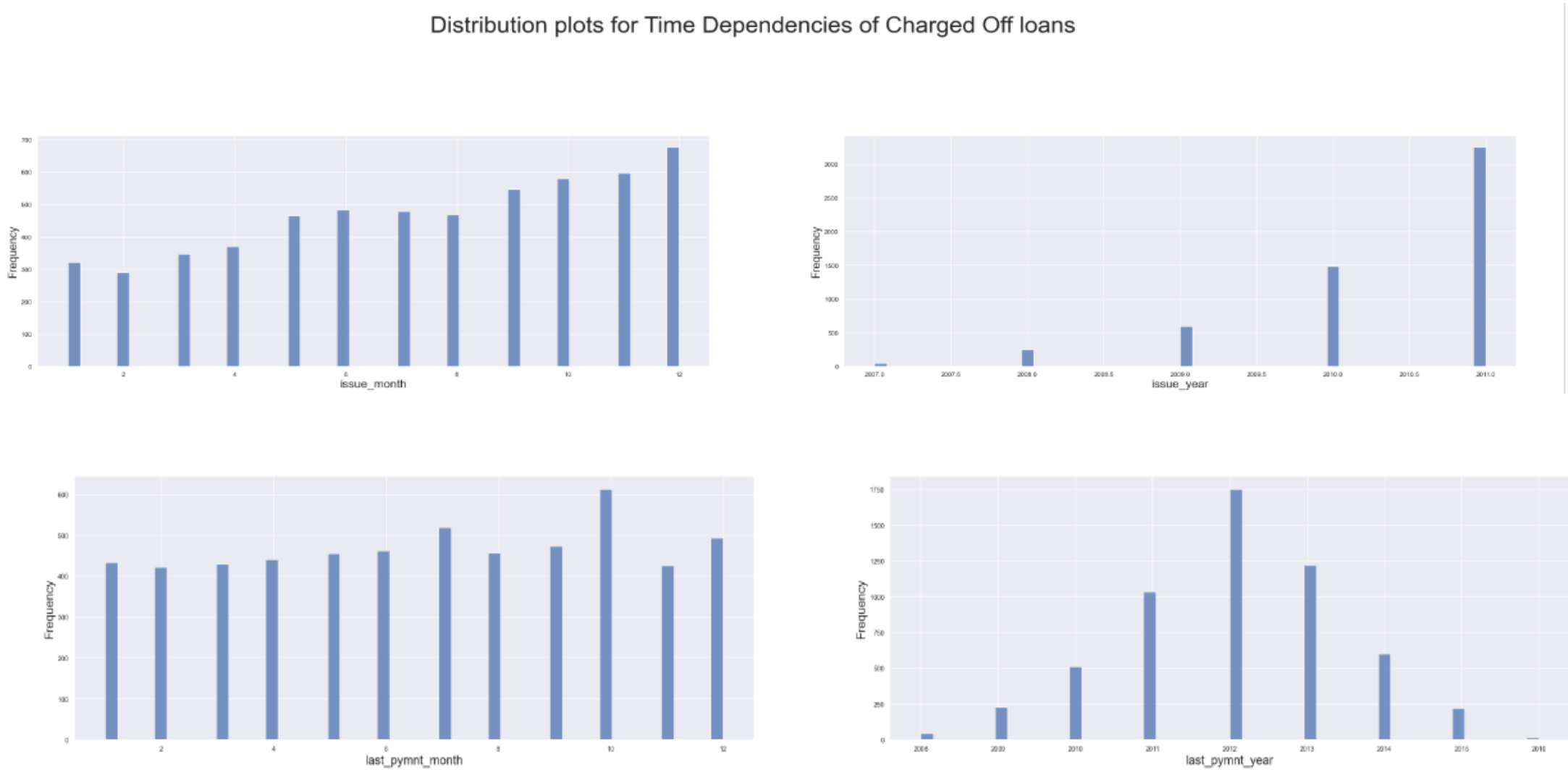
Univariate Analysis on Payment Status

....(Continued)

Observations:

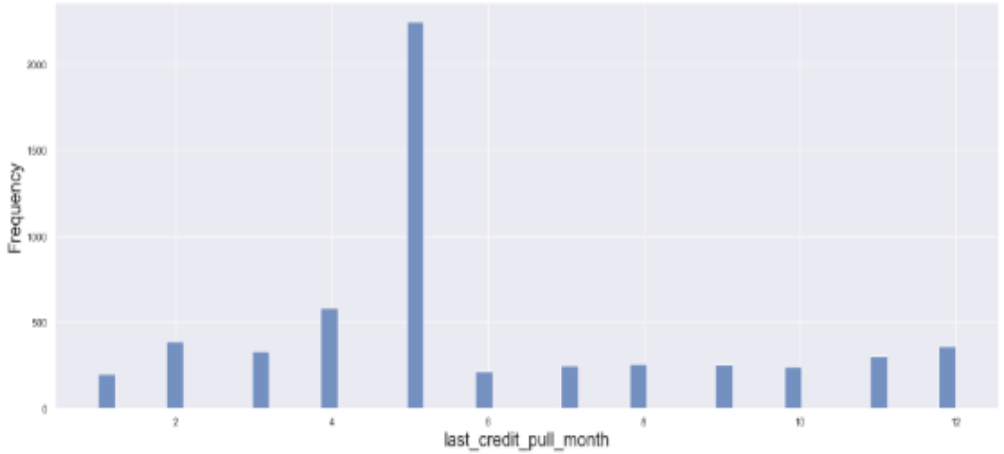
- Total payment and total payment invested shows similar data distribution.
- Average number of open accounts for customers is 9. We need to check the number of open accounts for defaulters.
- Revolving Line Utilization and revolving balance peaks at lower value, which is good. For higher values the distribution is uniformly spread.
- Total recovered principal shows similar trend with loan amount, with peaks at 5000, 10000, 15000 and so on. Co-relation between these columns need to be verified.

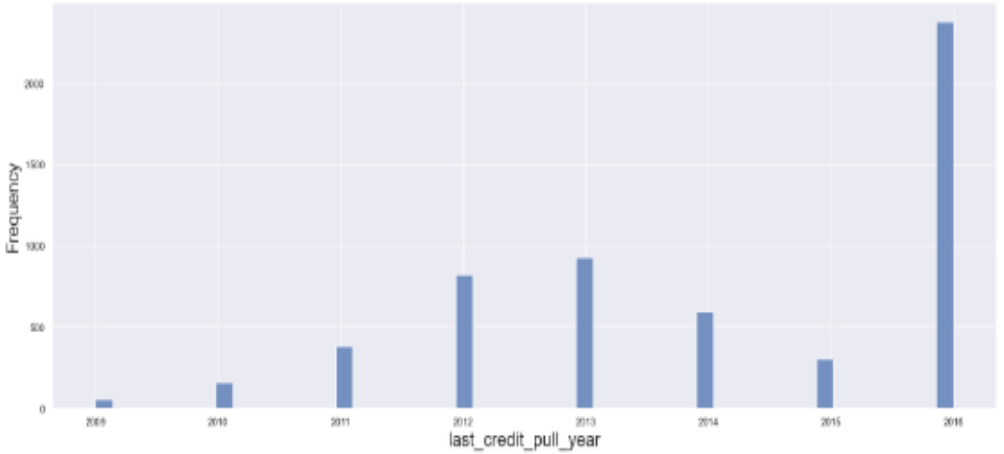
Univariate Analysis on Payment Status



Univariate Analysis on Time Dependencies

Continued....





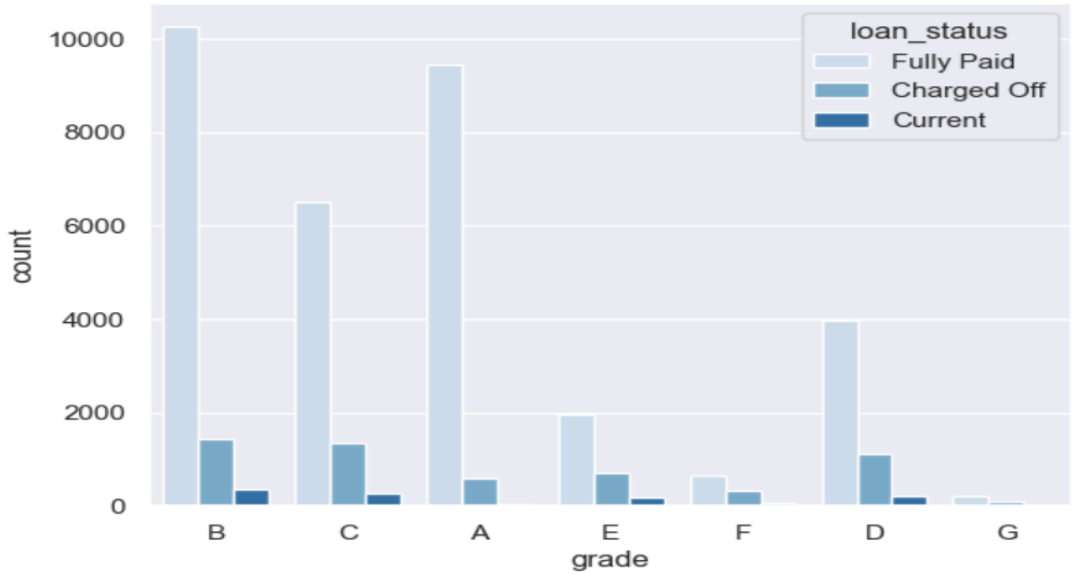
Observations:

- The number of Charged Off Loans has been increasing every year considerably. It is evident that identification of risky applicants is not being done currently.
- For most of the Charged Off loans, payments were not received after 2012.
- Credit pull records are more in recent years probably because the repayment rate has lowered down.

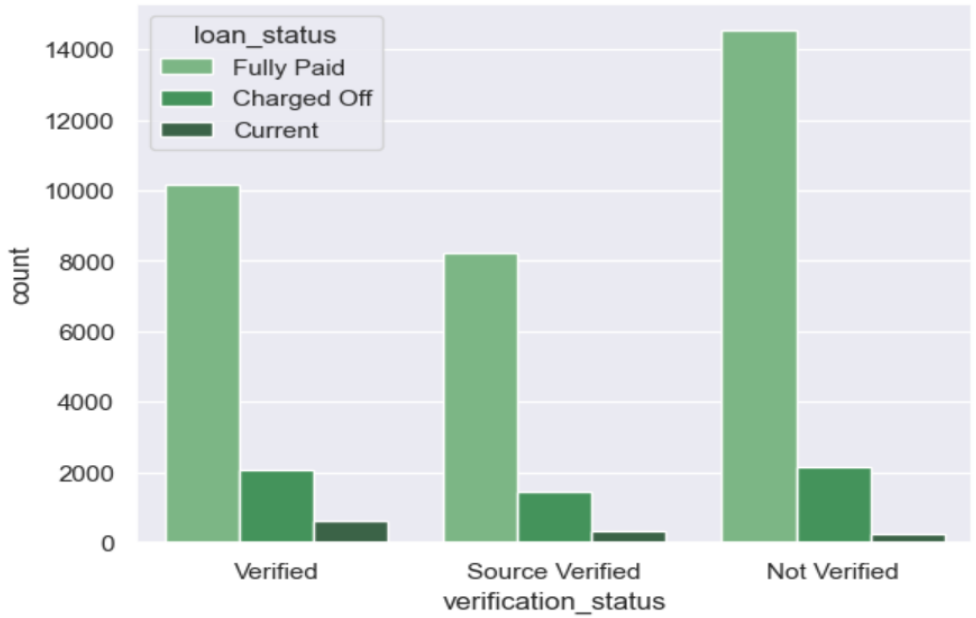
Univariate Analysis of Time Dependencies

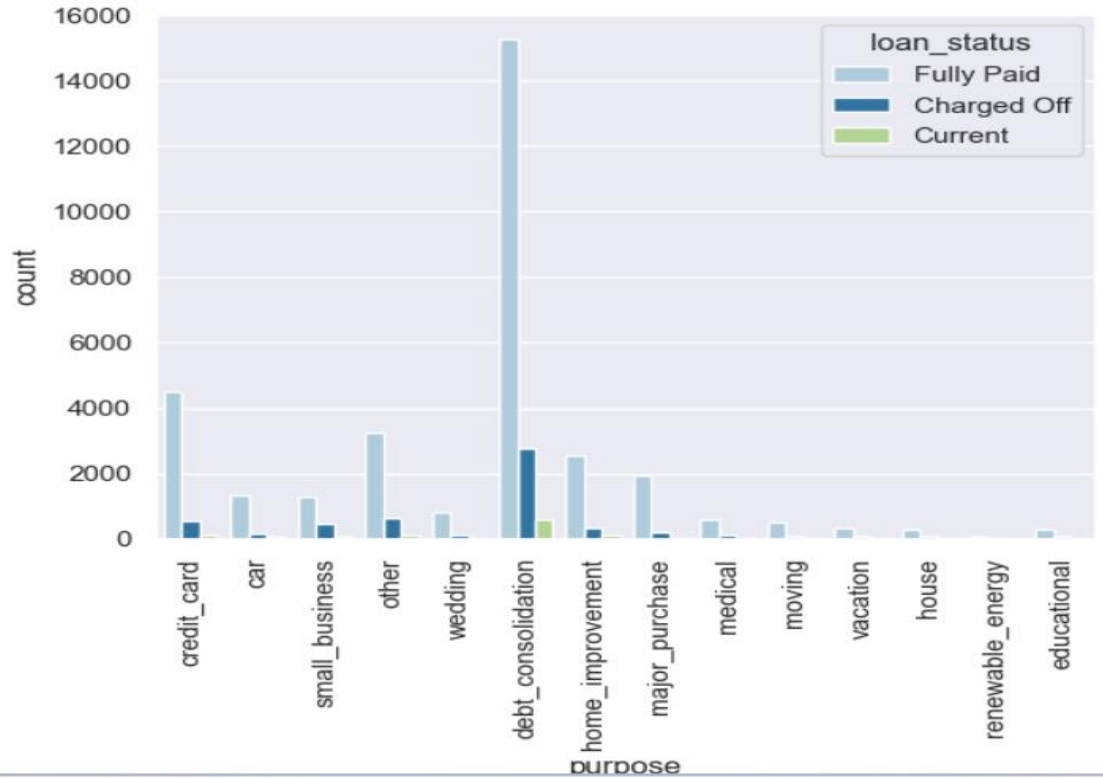
Univariate Analysis on Categorical variables.

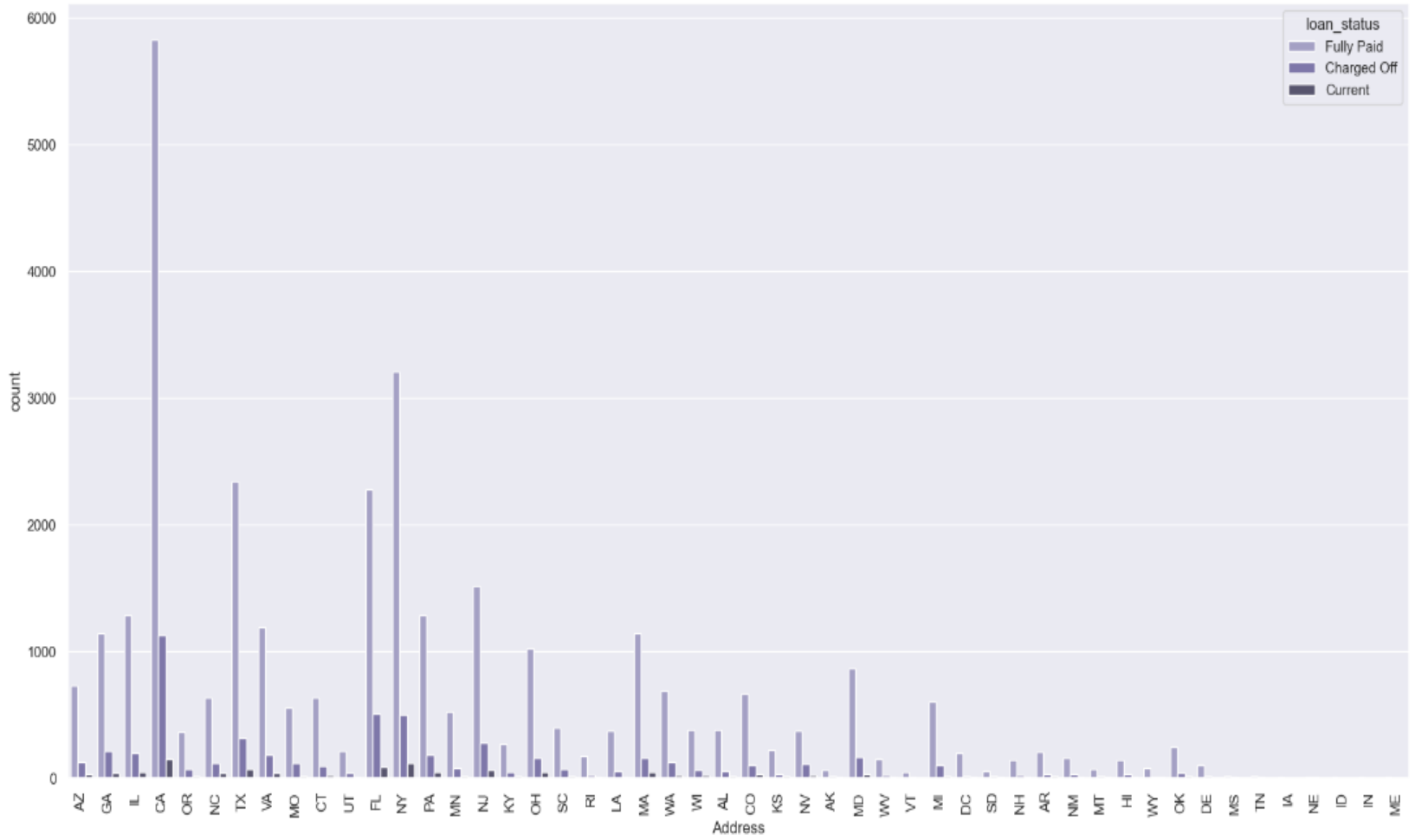
- We have categorized the **categorical columns** into sub-groups based on the functional definitions and have analyzed data.


loan_status
Fully Paid
Charged Off
Current
count
grade
B
C
A
E
F
D
G


loan_status
Fully Paid
Charged Off
Current
count
home_ownership
RENT
OWN
MORTGAGE
OTHER
NONE


loan_status
Fully Paid
Charged Off
Current
count
verification_status
Verified
Source Verified
Not Verified


loan_status
Fully Paid
Charged Off
Current
count
purpose
credit_card
car
small_business
other
wedding
debt_consolidation
home_improvement
major_purchase
medical
moving
vacation
house
renewable_energy
educational


loan_status
Fully Paid
Charged Off
Current
count
6000
5000
4000
3000
2000
1000
0
AZ GA IL CA OR NC TX VA MD CT UT FL NY PA MN NJ KY OH SC RI LA MA WA WI AL CO KS NV AK MD WV VT MI DC SD NH AR NM MT HI WY OK DE MS TN IA NE ID IN ME
Address

Observations on Categorical Variables

- Number of loans defaulted is very less for housing loan customers who own the house. But the number of applicants owning the house is very less, not providing much insights.
- Majority of the loans were taken for debt_consolidation, which means, loans were approved for customers already having other financial debts. This can be one of the major factors for defaulting the loan.
- Verification status shows similar pattern for all types of customers, so it is not very helpful for any conclusion.
- Canada has the highest number of customer base; hence number of defaulters are also higher in Canada.
- NY also has comparatively higher number of defaulters.

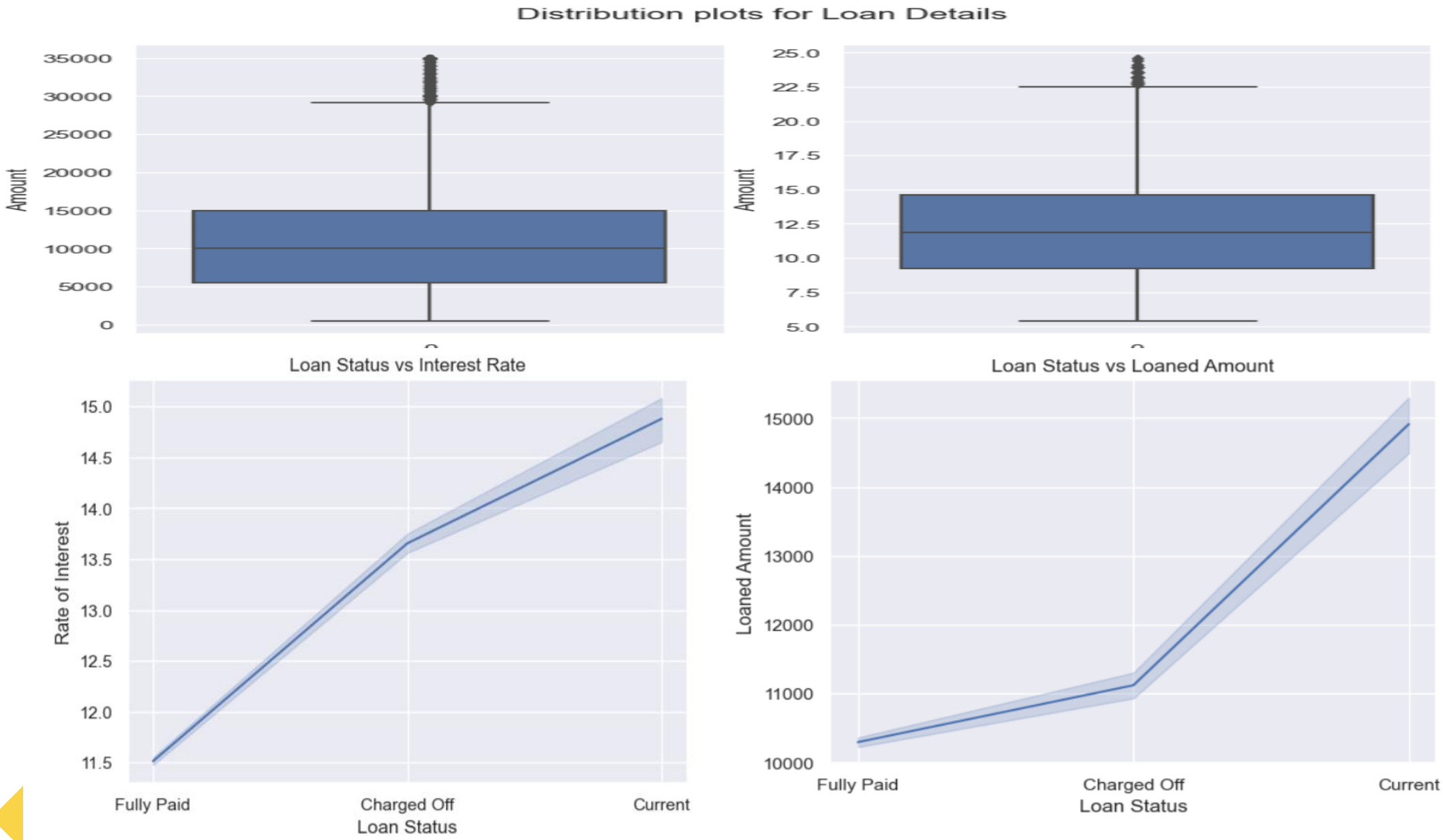
Bivariate and multivariate variable Analysis.

- We are analyzing two or more columns via Bivariate and multivariate in coming slides.

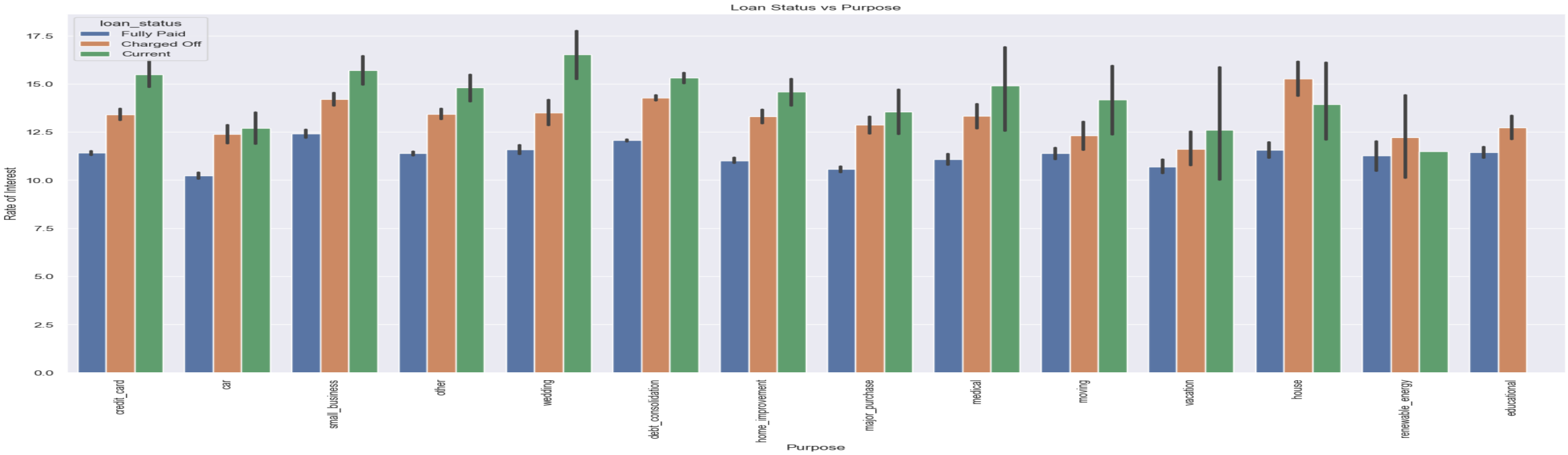
Bivariate and Multivariate Analysis

During our univariate analysis, we have observed the following co-relations between certain attributes. We will now investigate them in more details as a part of our analysis. We have considered the loans with loan status as **'Charged off'** as our category for defaulted loan.

- Interest Rate and Loan Amount varying with loan status.
- Co-relation between Annual income, Employment Length, Installments and DTI.
- Loan Grade and corresponding sub-grades changing with loan status.
- Impact of debt-consolidation(purpose) on the defaulted loans.
- Average number of open accounts for defaulting LCs - is it higher than 9?
- Co-relation between recovered amounts and the loan amount over time.
- Co-relation between annual income and the DTI.
- Time distribution of purpose of loans applied and loans defaulted



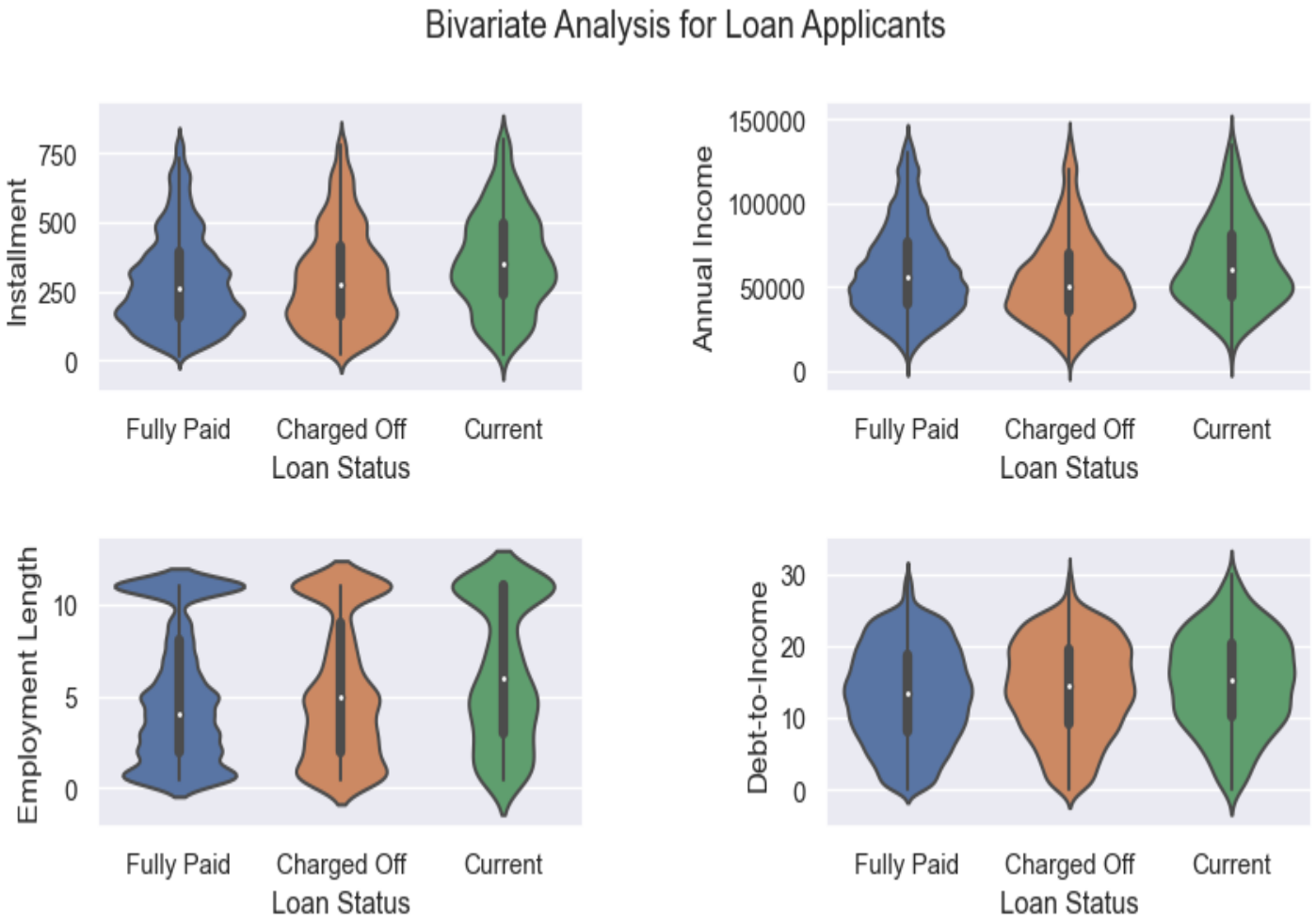
Bivariate and Multivariate Analysis



Observations:

- Loans with lower interest rates are paid off but loans with interest rate higher than 13.5% are more charged off.
- Loan amount has similar pattern for both fully paid loans as well as charged off loans.
- Small BUsiness, House and Debt consolidation requires attention as the outliers can be seen extended till 22.5%.
- Renewable energy has good repayment rate but the number of current loans approved is very less.

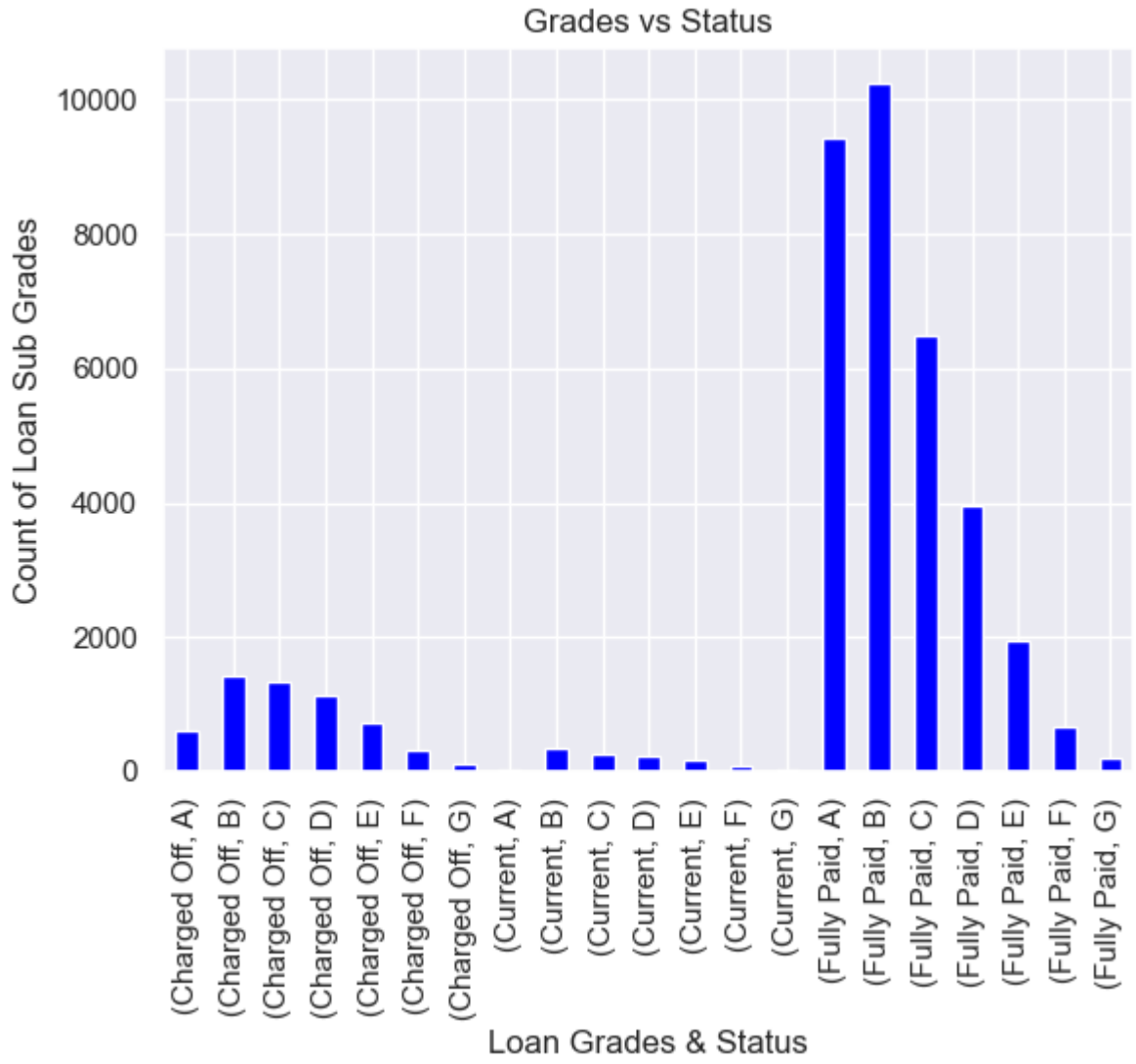
Bivariate and Multivariate Analysis



Observations:

- Frequency distribution does not show significant variation with the loan status for these parameters.

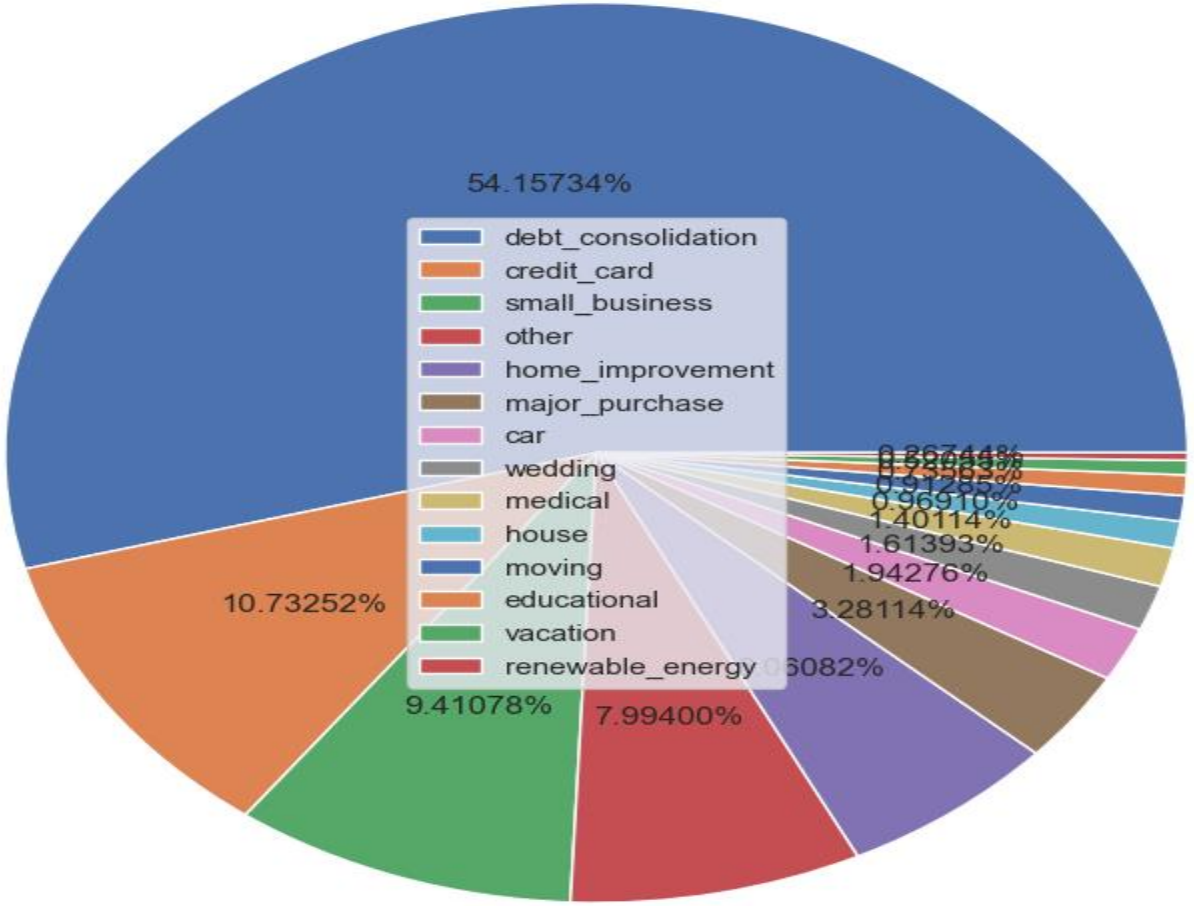
Bivariate and Multivariate Analysis



Observations:

- Grade A & B has highest number of defaulters for a good number of applicants
- Chances of loan default is more for Subgrades of C and D although the number of applications are not very high
- Subgrades of A have very good repayment rate.

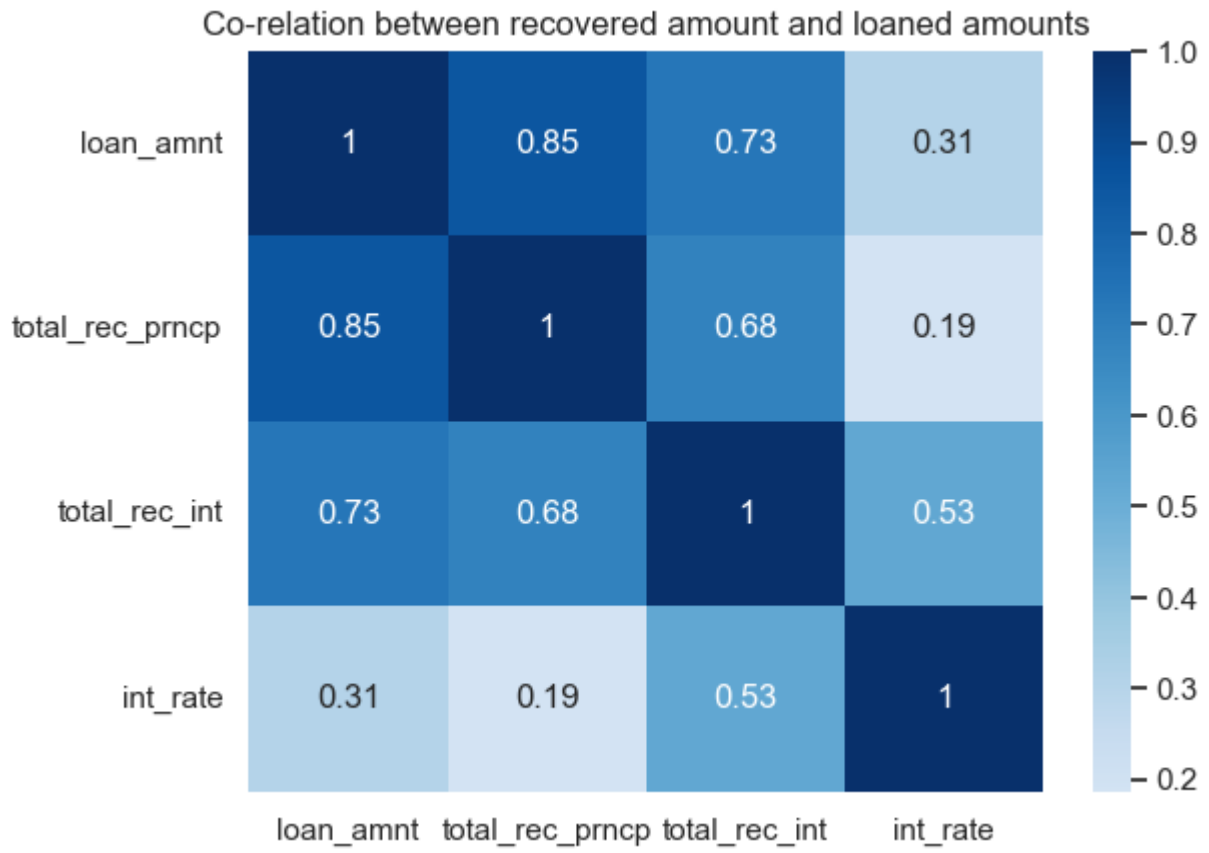
Bivariate and Multivariate Analysis



Observations:

- 54% of the total charged off loans are for debt consolidation. This should be a concerning factor for the organization.
- For renewable energy the total charged off loans is only 0.27%. The organization should invest more on renewable energy.

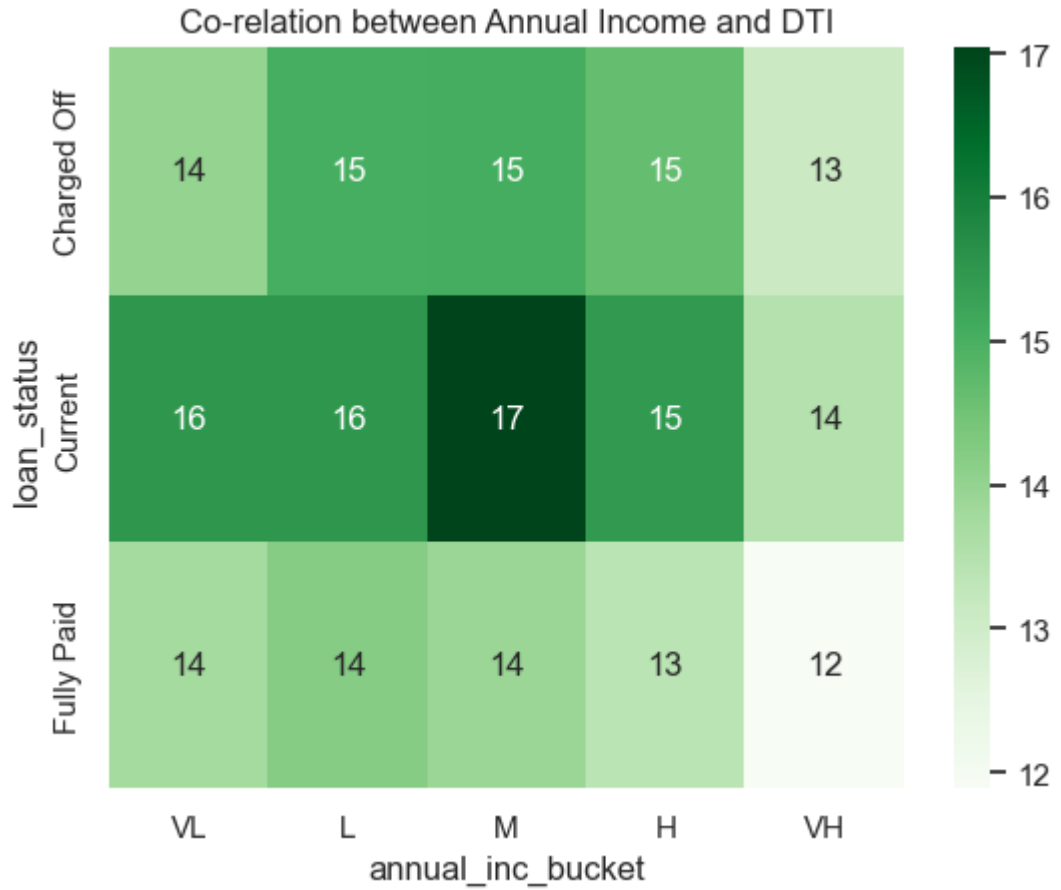
Bivariate and Multivariate Analysis



Observations:

- Higher the interest rate, lower is the amount loaned which seems to be an obvious fact.
- With increasing interest rate the rate of recovery is lowering down supporting our previous findings.

Bivariate and Multivariate Analysis

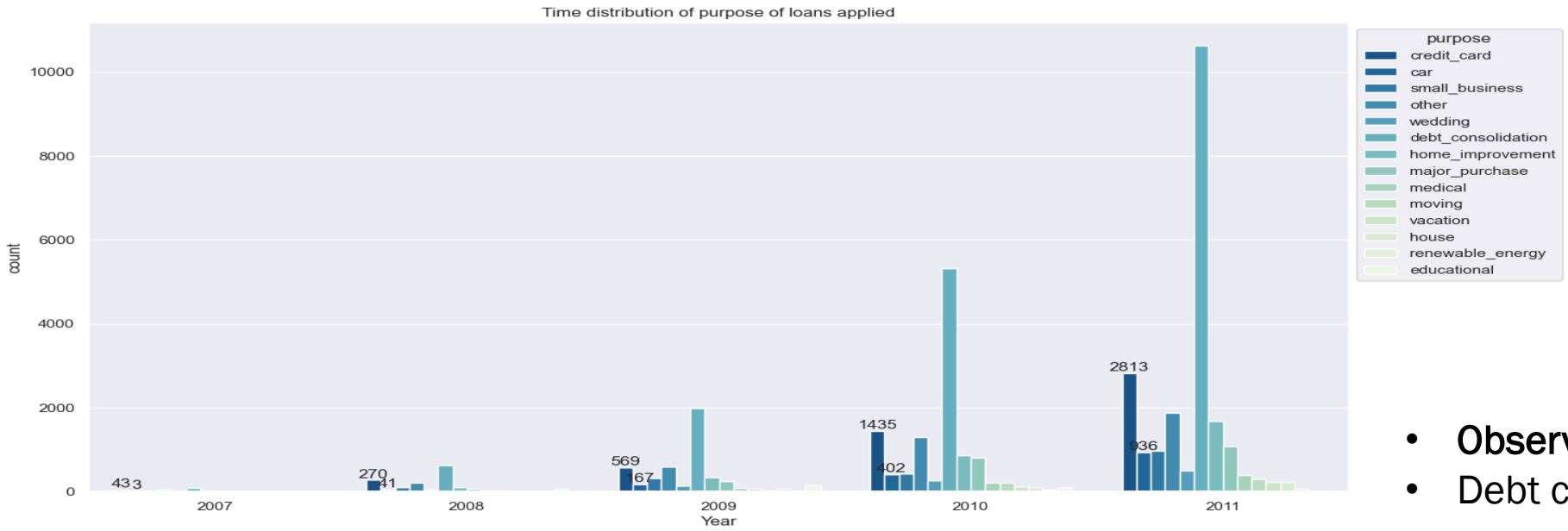


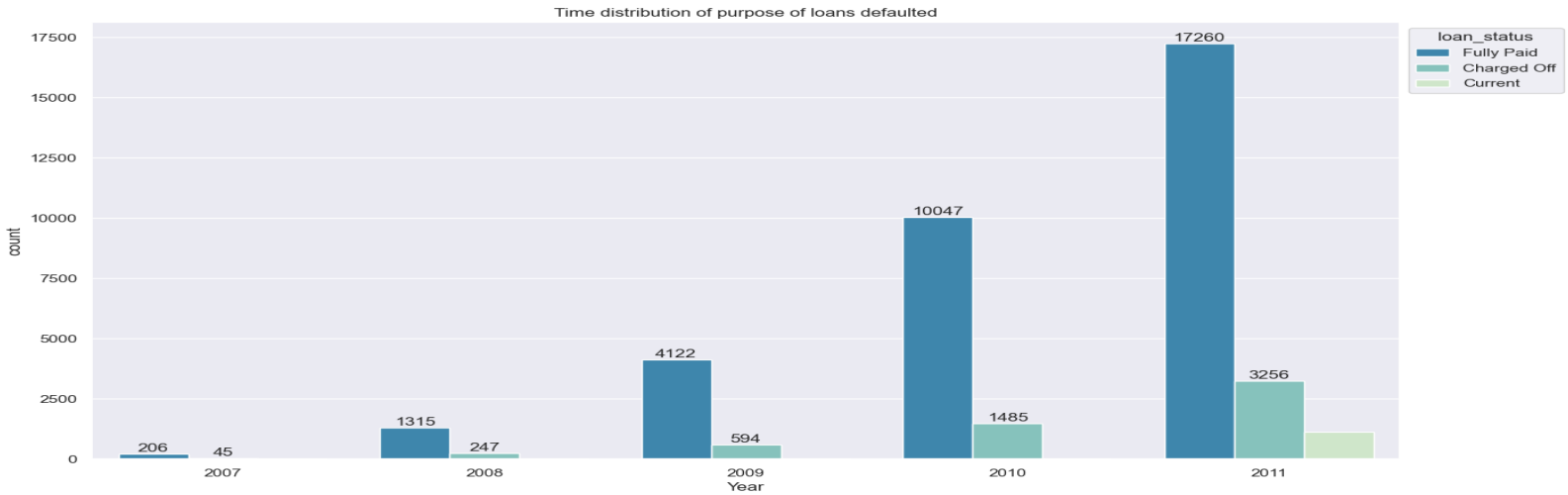
Observations:

- Current loans need immediate attention to remediate risk of loan defaults.
- Users under Block 16, i.e users with Low Income(<40k) and High Debt ratio(>19) is second highest risk are more prone to default loans as well.
- Users with High Income have medium to low DTI.
- Block 15 with users having low to high income but Medium DTI shows some risk of defaulting loans.

DTI plays major role in defaulting.

Bivariate and Multivariate Analysis





- **Observations:**
- Debt consolidation and credit card purpose seems to increase gradually as months passby where rest remains almost same..
- Loans issues on debt consolidation almost doubles by end of year, where as credit card loans increases by 3x.

Bivariate and Multivariate Analysis

Observations:

- Current loans need immediate attention to remediate risk of loan defaults.
- Users under Block 16, i.e users with Low Income(<40k) and High Debt ratio(>19) is second highest risk are more prone to default loans as well.
- Users with High Income have medium to low DTI.
- Block 15 with users having low to high income but Medium DTI shows some risk of defaulting loans.

DTI plays major role in defaulting.

Bivariate and Multivariate Analysis

Conclusion

- Loans approved should have lower interest Rate. Amounts loaned at higher interest rates(typically more than 13.5%) are mostly charged-off.
- For higher rate of interest, the annual income of the applicant should be high (approx. more than 83K) for a safer decision.
- Amounts for loan subgrades of A & B should be favored over the remaining to reduce charged-offs rates.
- Number of loans defaulted is lesser in count for housing loan customers who owns the house while it is more for the ones with rented or mortgaged properties.
- Short term(36 months) loans are defaulted more often than long term loans(60 months).
- Investments on Renewable Energy should be increased and Debt Consolidation should be reduced.
- Applicants from Canada and NY have more chances to be defaulters.
- Credit pull for LCs should continue at the same rate.
- Applicants with Low/Moderate Annual Income and more than average to high Debt ratio should be validated carefully before approving the loan application.

Thank you

Mallika Bera & Mamatha.K
mamathak2647@gmail.com
Mallika.bera4@gmail.com