Lending Club Case Study

Kiran Pittampally

Kafeel Ahmad

Problem Statement

Identify risky loan applicants for **Lending Club** using **EDA.**

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Overview – Lending Club Case Study

AVOID FINANCIAL AND BUSINESS LOSS FOR LENDING CLUB

Business Objectives:

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.

Business Understanding:

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.

Two **types of risks** are associated with loan approval decision:
- **Likely to repay** - **loss of business** on loan rejection
- **Likely to Default** - **financial loss** on loan approval

Analysis Approach

- Data Cleaning
- Data Manipulation
- Data Imputation
- Categorical and Quantitative classification
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Key Findings

Univariate Analysis

Loan Growth including default loans from years 2007-2011

- Lending club is grown significantly in their business of issuing loans to borrowers for the time period 2007-2011

- As years passing on from 2007-2011, Lending Club is failed to control the increase in Charged Off Cases.

- Lending Club should have had a strategy to control this problem as soon as possible it was identified.

Loan Issue Year Distribution
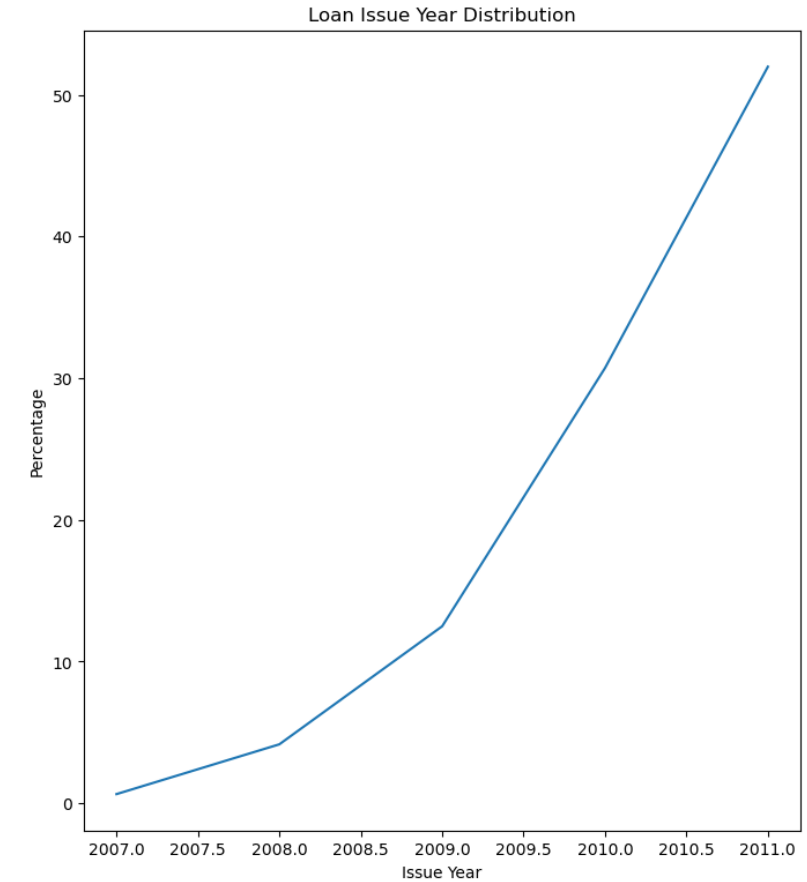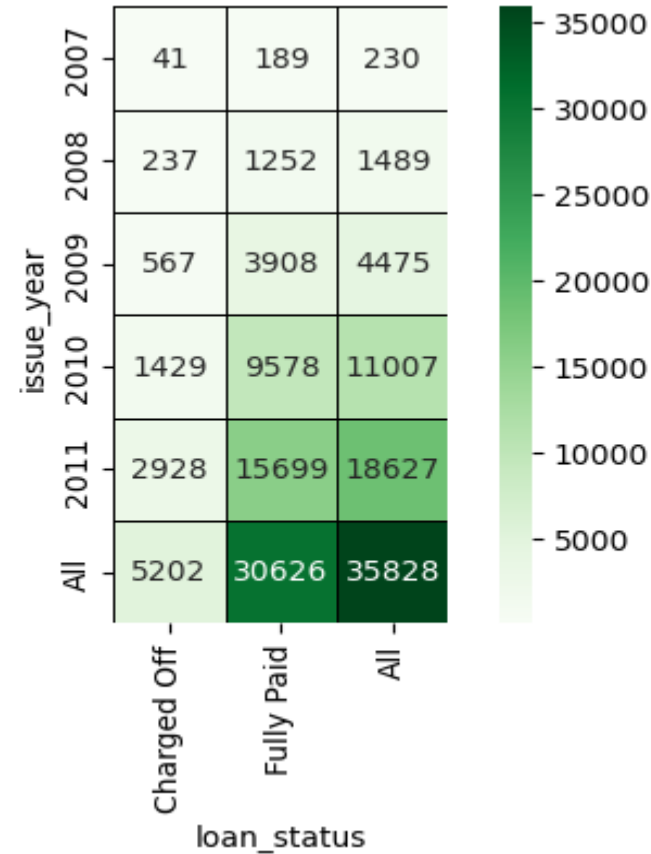
Loan Status Percentage Distribution

- **Observations:**
  - 14.2% loans are in Charged Off status which means 1 out of every 7 borrower is defaulter
  - 83% loans are fully paid
  - After removing current loan status records, the charged off loans will be at14.6%
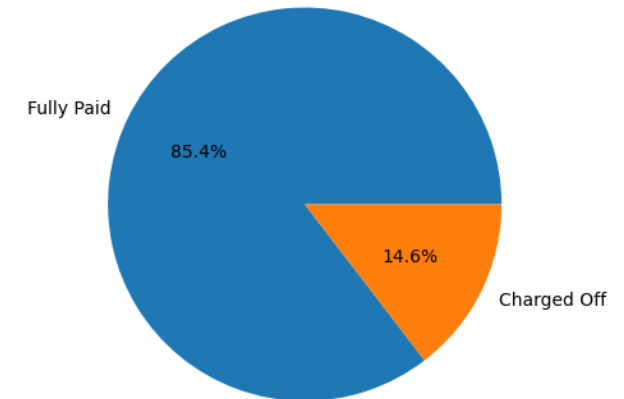
- **Potential Insights:**
  - Loans in defaulter state may risk financial health depending upon loan amounts
  - The high proportion of fully paid loans indicates a positive lending performance.

Distribution of Loan Status

Fully Paid 83.0%
Current 2.9%
Charged Off 14.2%
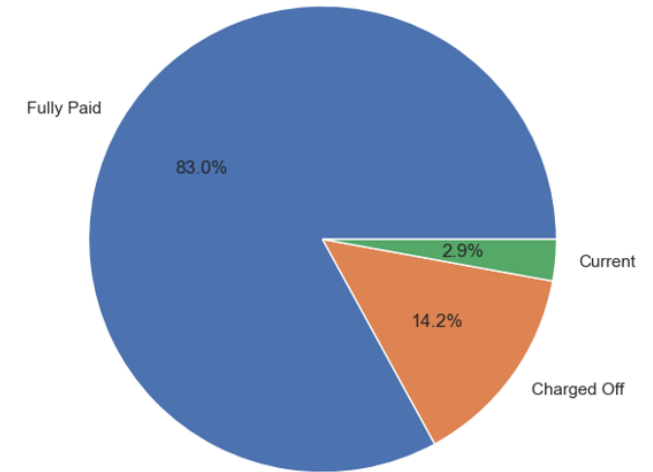
Loan Status Distribution after Removing Current loans

Fully Paid 85.4%
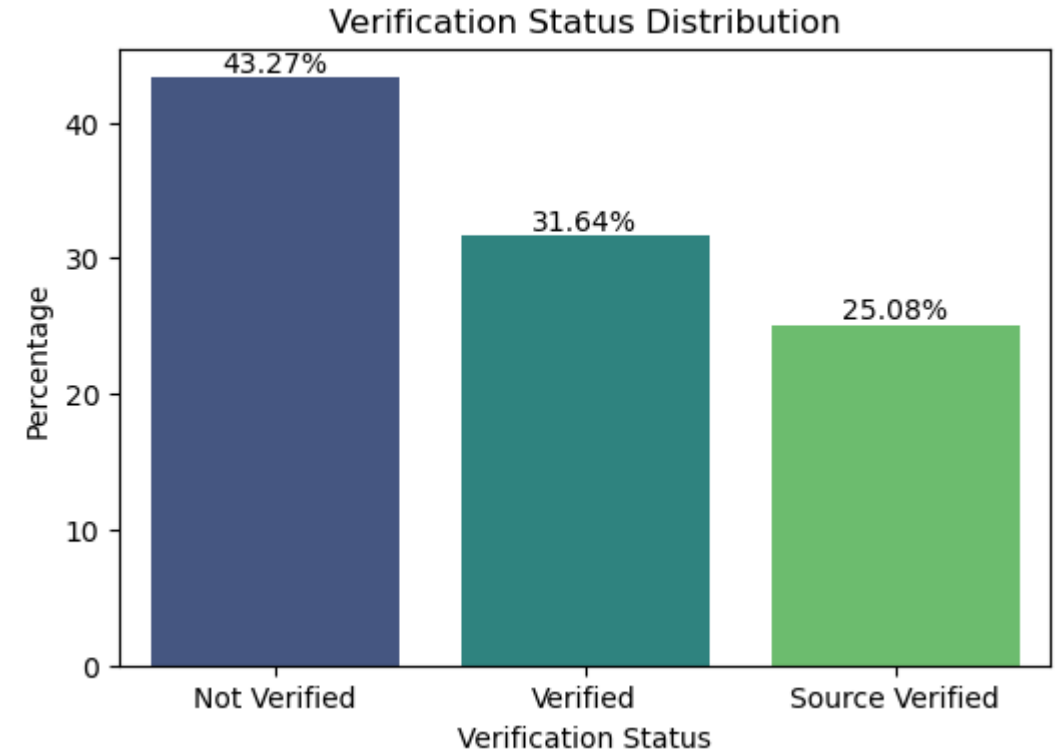Charged Off 14.6%

# Loan Verification status Distribution

- **Observations:**
  - The largest proportion of loans (43.27%) fall under the "Not Verified" category.
  - This suggests that a significant number of loans have not undergone a comprehensive verification process.

- **Potential Insights:**
  - The lack of verification for a significant portion of loans could potentially increase the risk of default or fraud.
  - The high proportion of "Not Verified" loans might suggest a need for streamlining the verification process


Verification Status Distribution
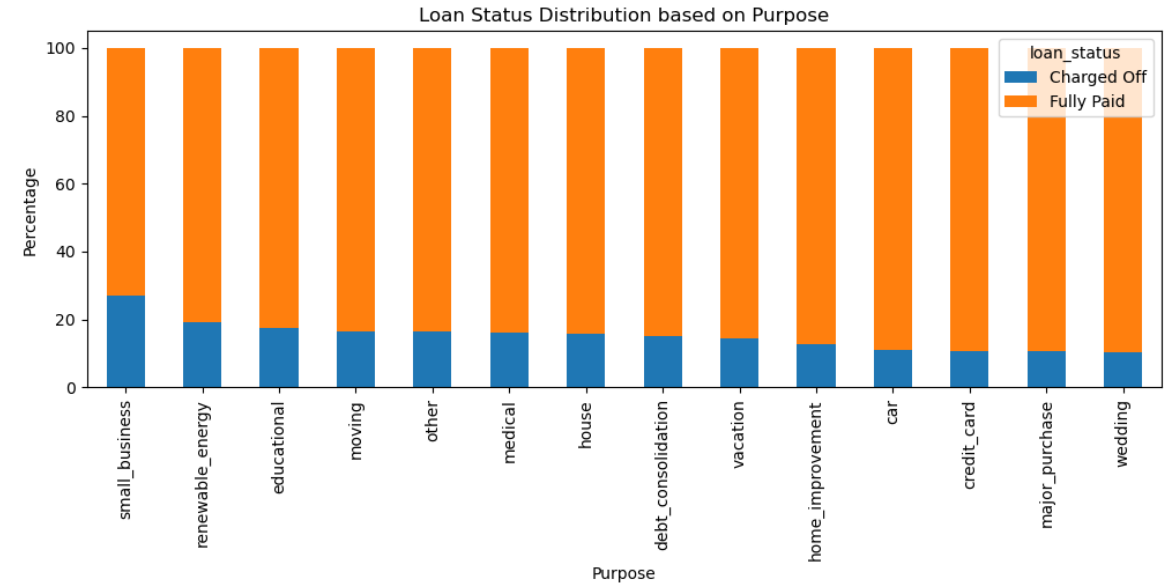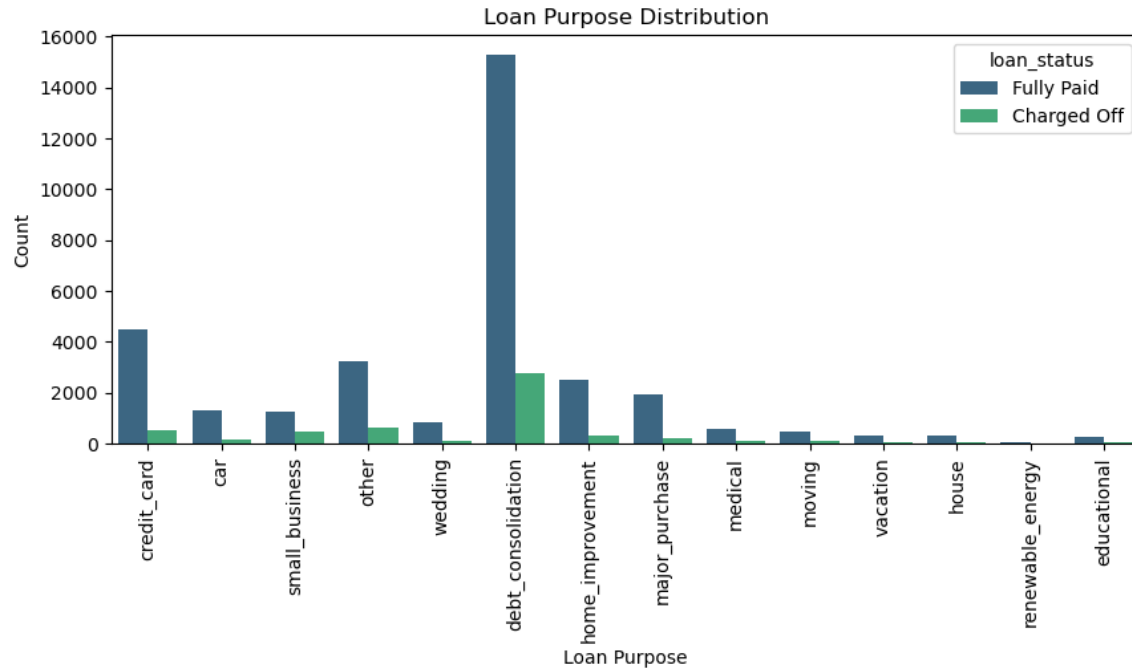
Loans distribution by Purpose segmented by loan status

- The highest count of loan purpose is 'Debt Consolidation' and 'Credit Card'

- Small Business Purpose is having highest percentage Charged Off Loans , followed by Renewable Energy Purpose.

- LC should add more credit profile checks while issuing loans to Small business borrowers

Borrowers based on Address State
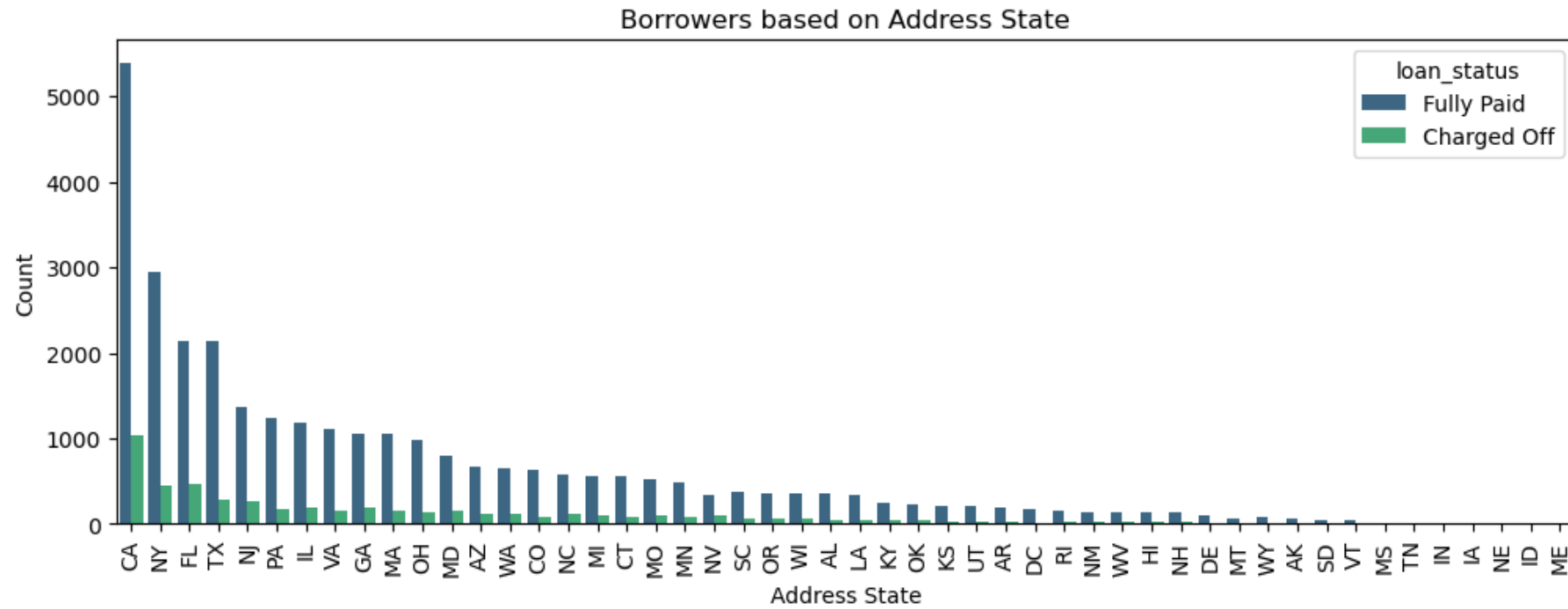
Count of Loans in States and segmented by Loan status

- Lending Club has highest count of borrowers with loan status (Fully Paid, Charged Off) from California, followed by New York and Florida.

Loan Verification Status segmented by Loan Status

- The verification status 'Verified' has the highest number of loans that are charged off.

- Lending Club should restructure verification process to avoid any verified loans to become defaulted.

- There are significant borrowers who paid fully loan amount though they were not verified. **So, the insight here, some other factors like Interest rate, Term are dominating the borrowers to default their loans.**


Verification Status Distribution

Home Ownership versus Loan Status
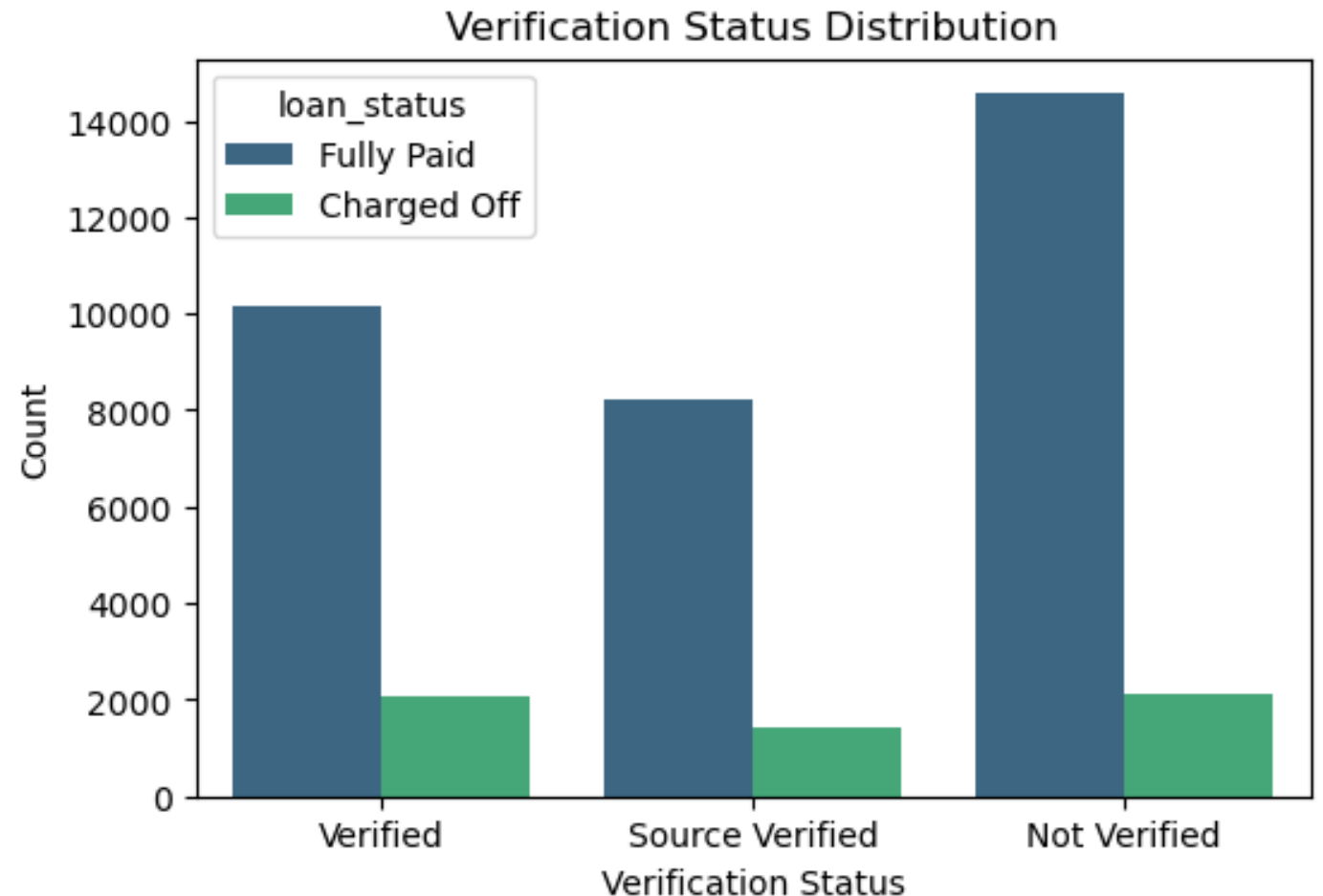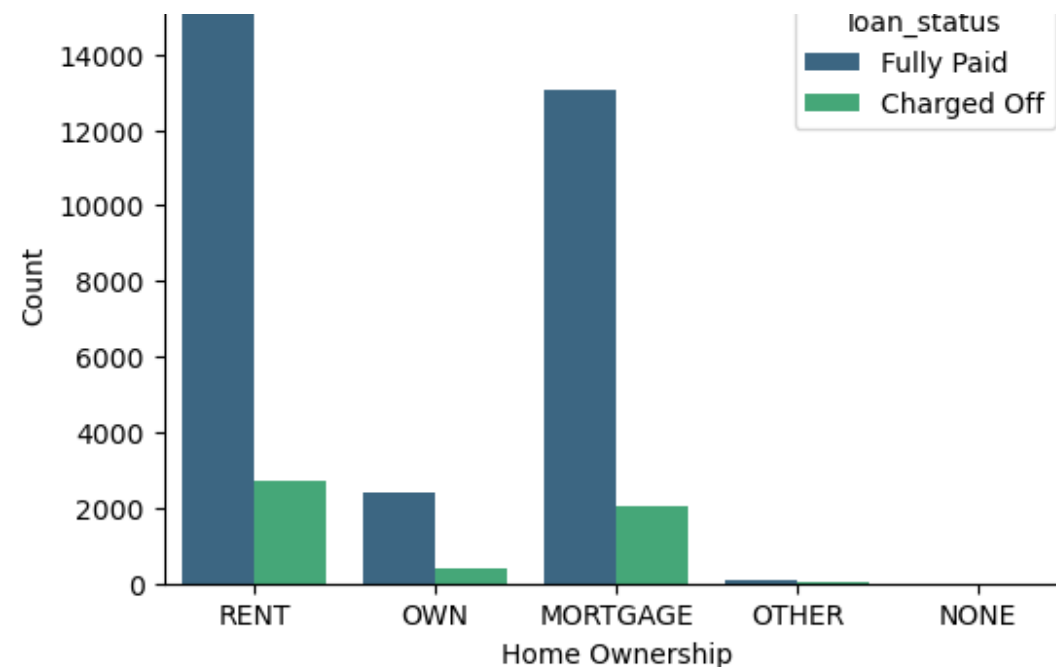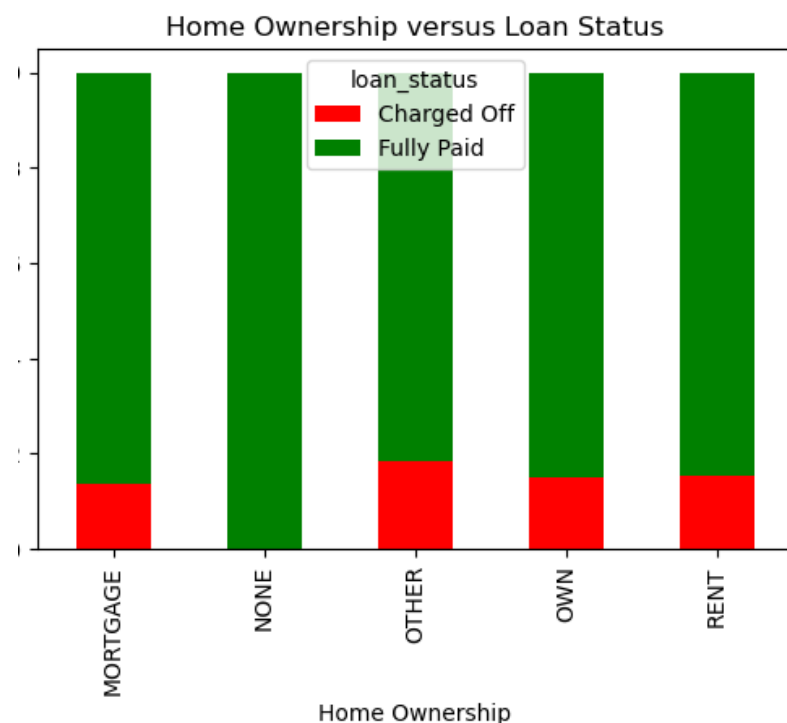
Borrower's home ownership's impact on Loan status

- Borrowers with home ownership like OWN, RENT, Mortgage are having 20 % defaulting Loans.
- Lending Club should focus on recovery process as the defaulter can be tracked through home location.
- **Here the data for home ownership like NONE, OTHER is almost negligible.**

LC Borrowers public records and bankruptcy impact on loan status

- Borrowers with lowest derogatory records and low bankruptcy filings are also defaulting their loans.

- Though checking derogatory and bankruptcy history records is mandatory to check before issuing the loans.

- **LC should focus other factors like DTI in verification methods though borrower looks green on Public derogatory and bankruptcy records.**

Public Records Bankruptcies Distribution

Public Records Distribution
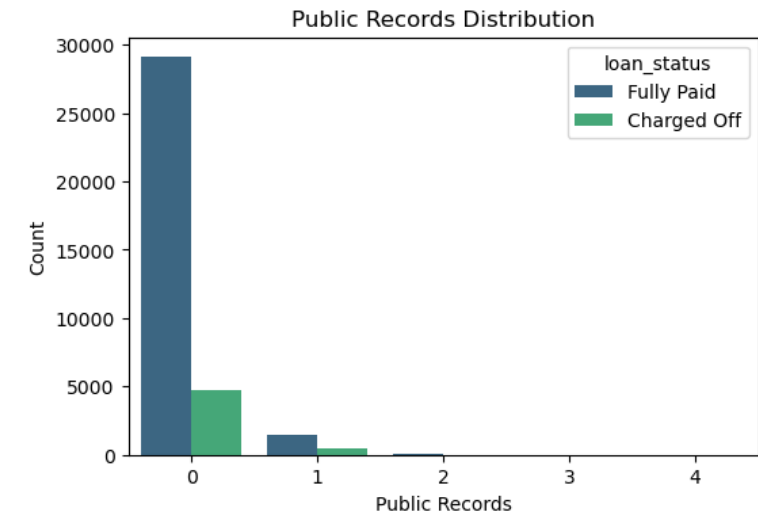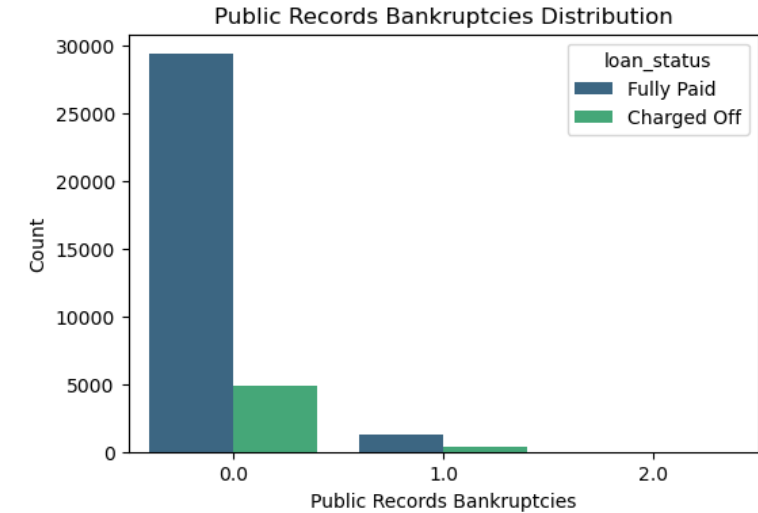
LC – Borrower Annual Income and Loan Amount distribution

These both variables contain outliers which needs to be handled before constructing any plots

Histogram for Annual Income (After Outlier Handling)

Histogram for Loan Amount (After Outlier Handling)
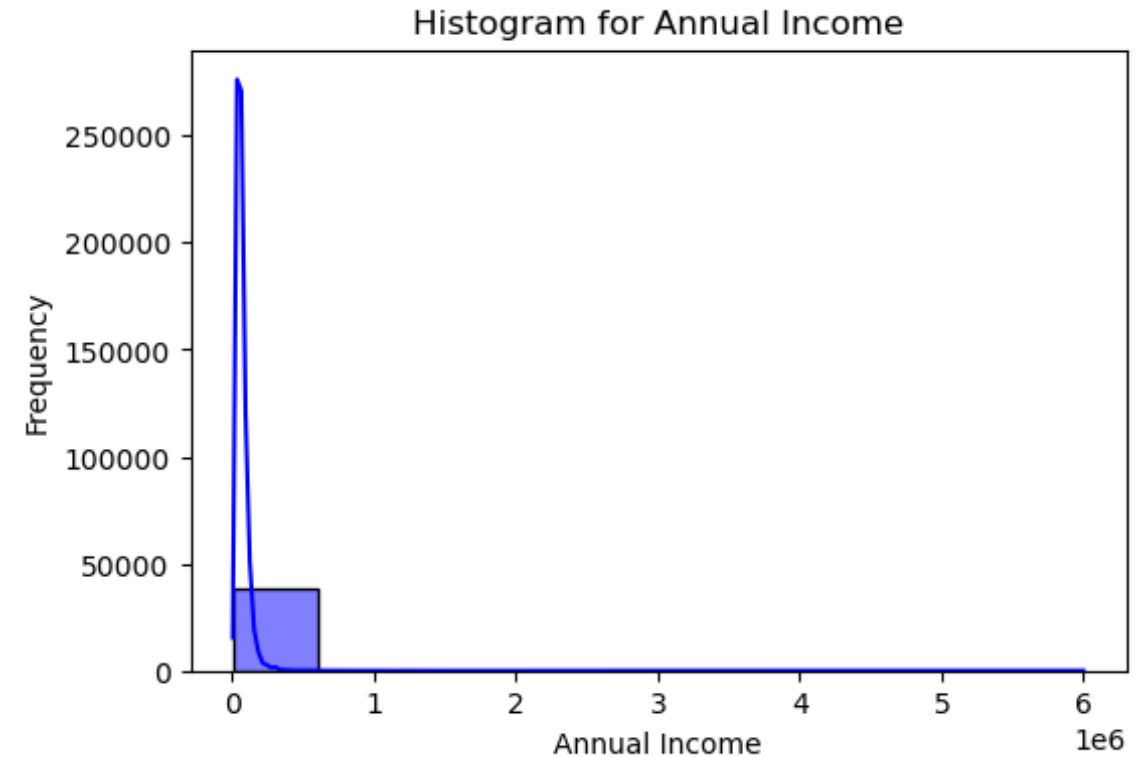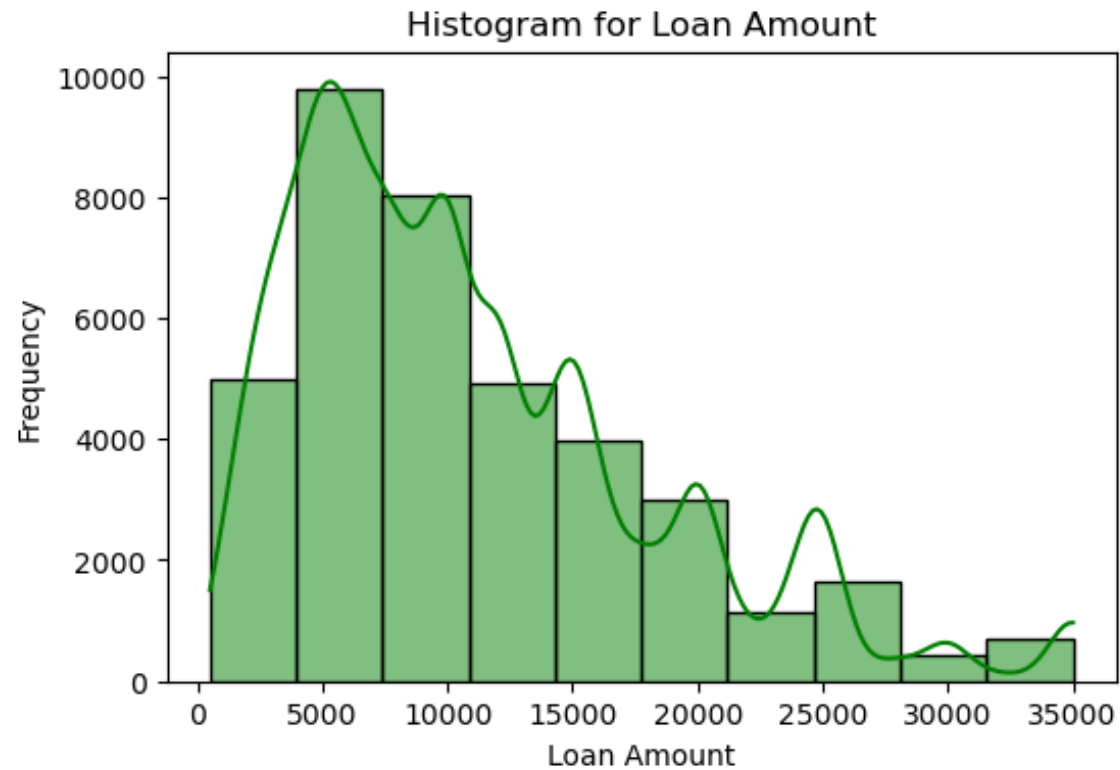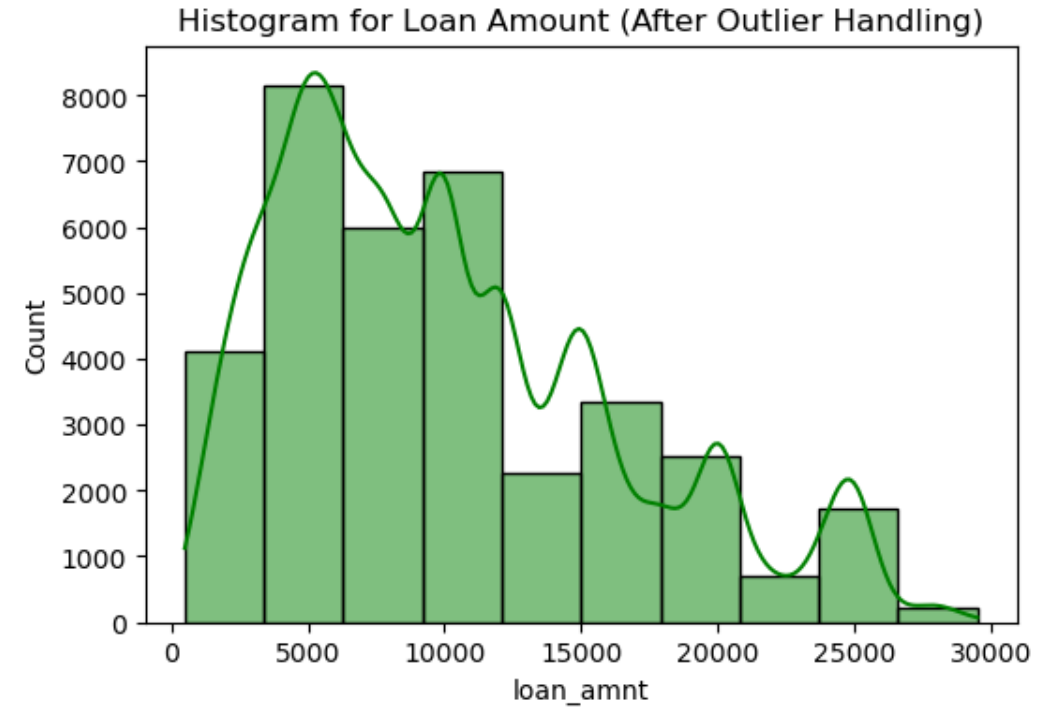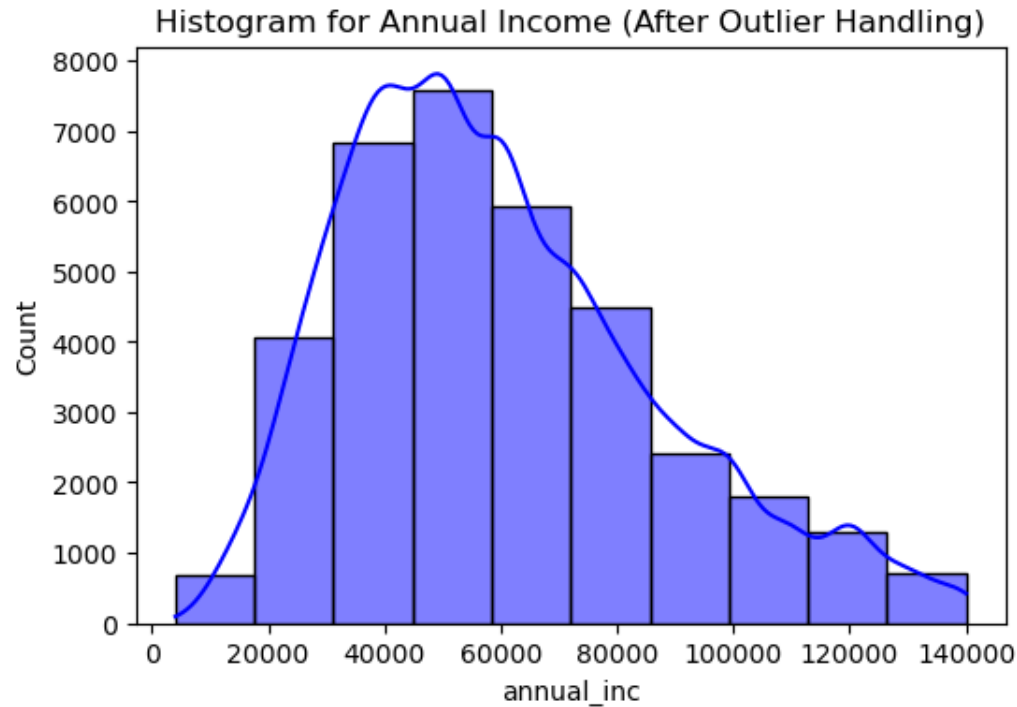
LC – Borrower Annual Income and Loan Amounts frequency

After outlier treatment:

- Average annual income is between 40k-60K
- Average loan amount is between 5K-10K

LC – Histograms for DTI, Interest rates

The peak DTI is noted around 15, and the most frequented Interest rates around 12.5 %
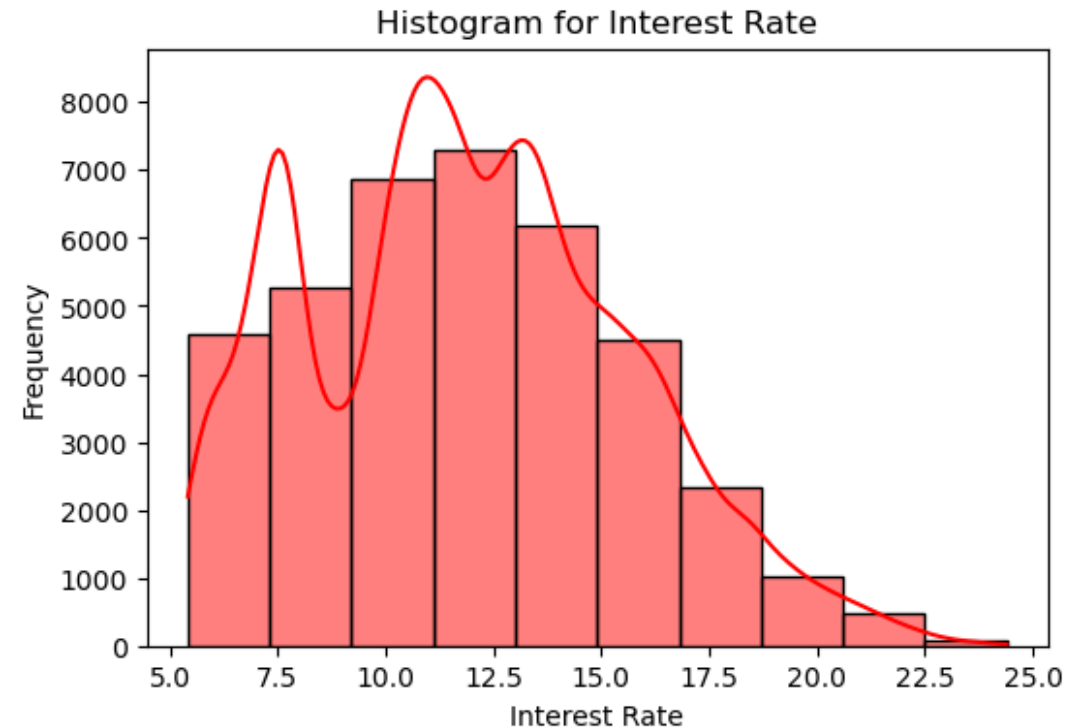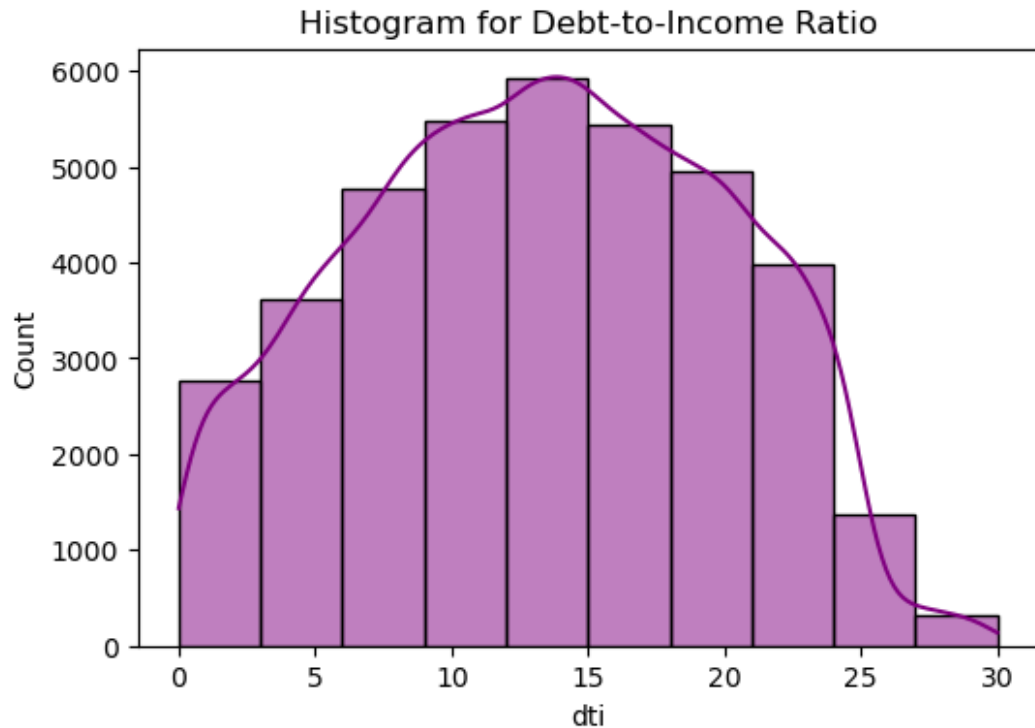
LC Funded loan amount by Investors

Investors lost 15% of their funded amount in Lending Club due to Charged off loans.
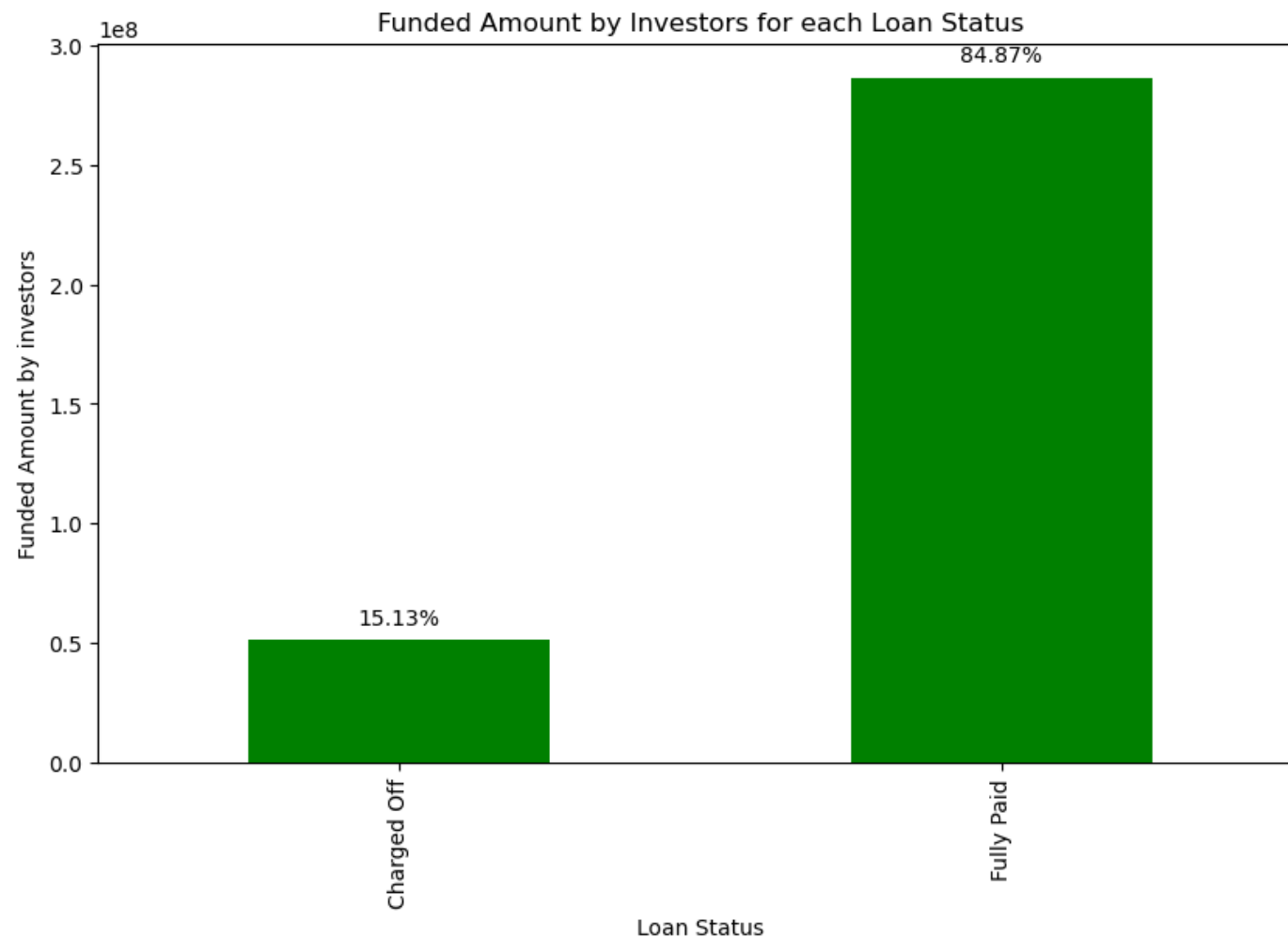
LC Sub Grade Loans frequency in "Charged – Off" loans

B5 sub grade is the most frequent defaulted loan type followed by C1.

Sub Grade Distribution for Charged Off Loans
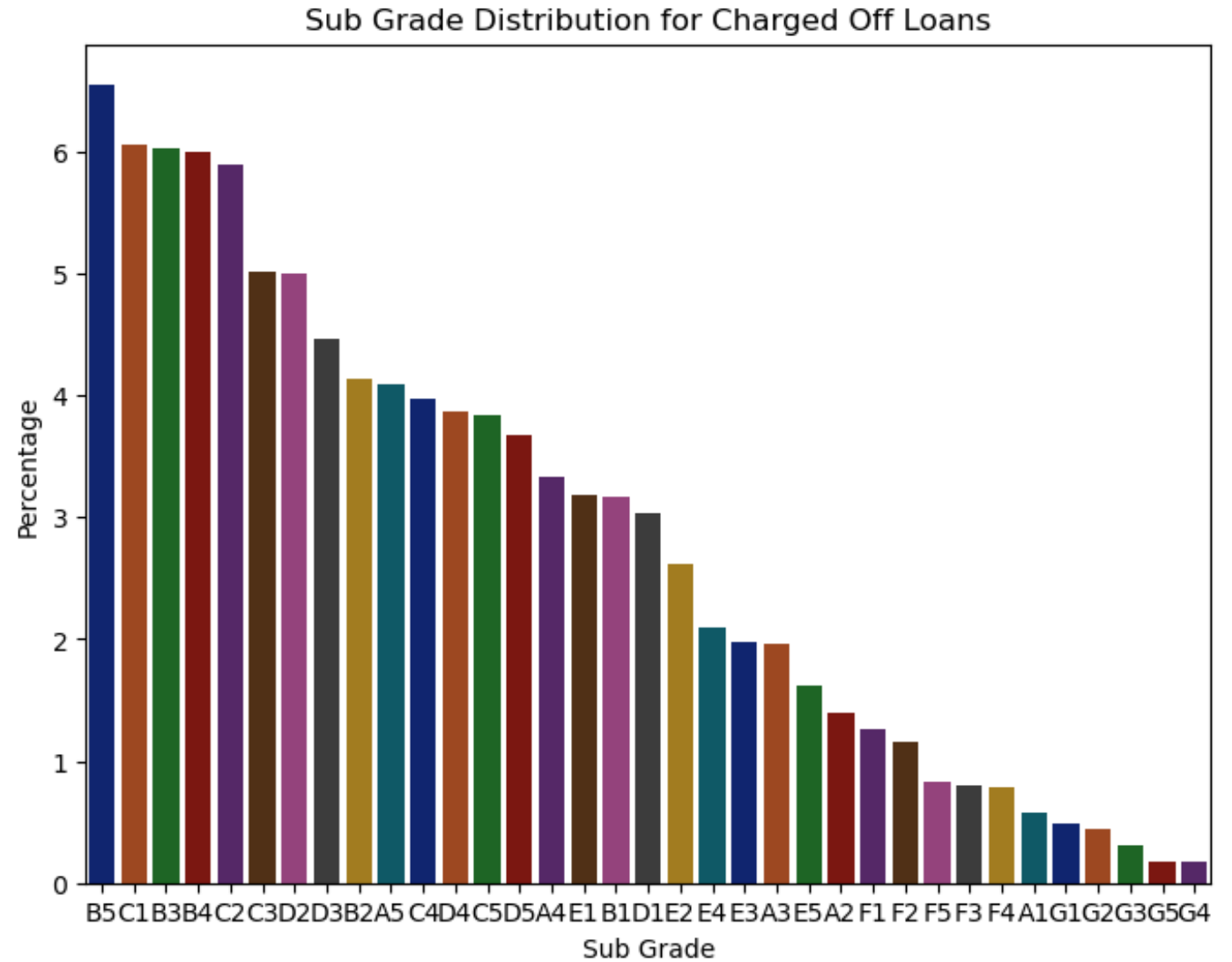
Bivariate Analysis
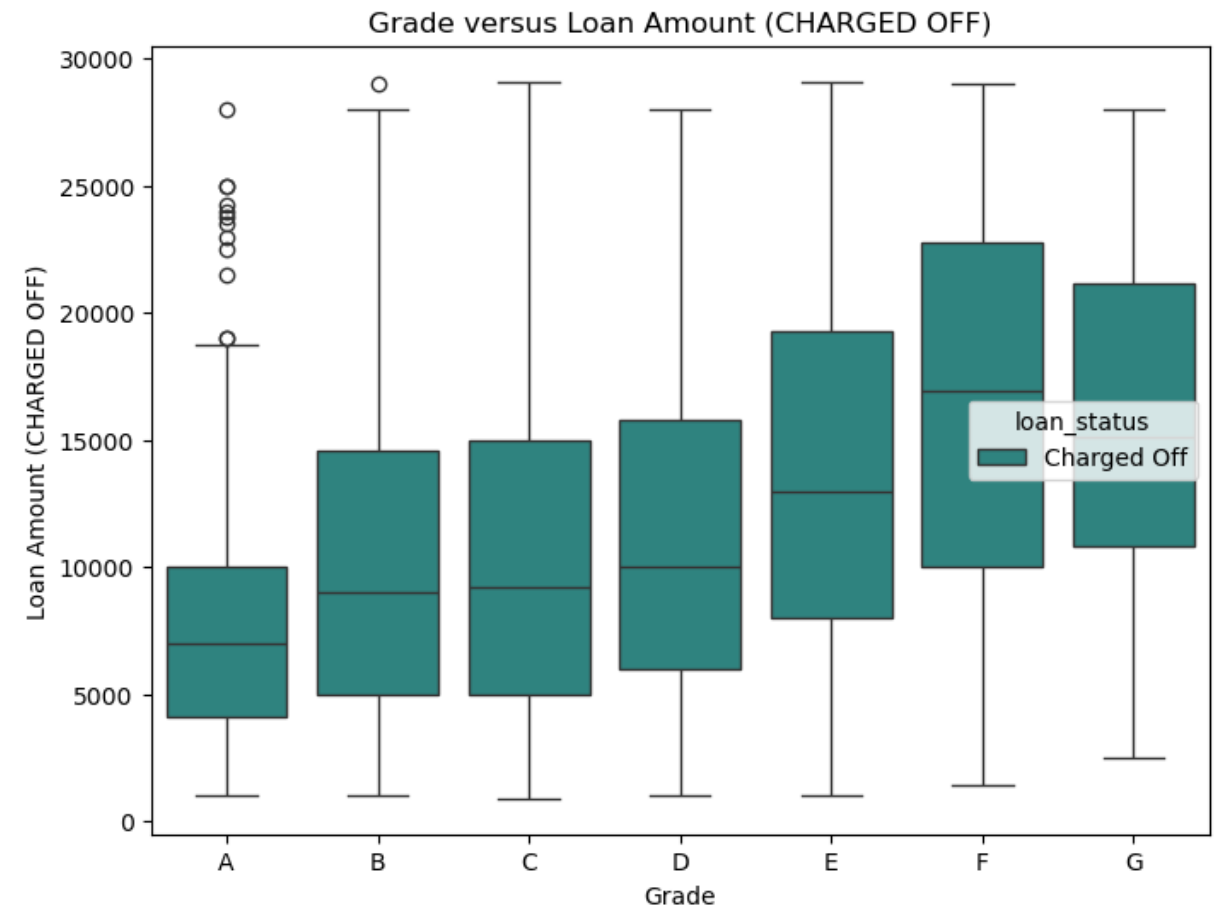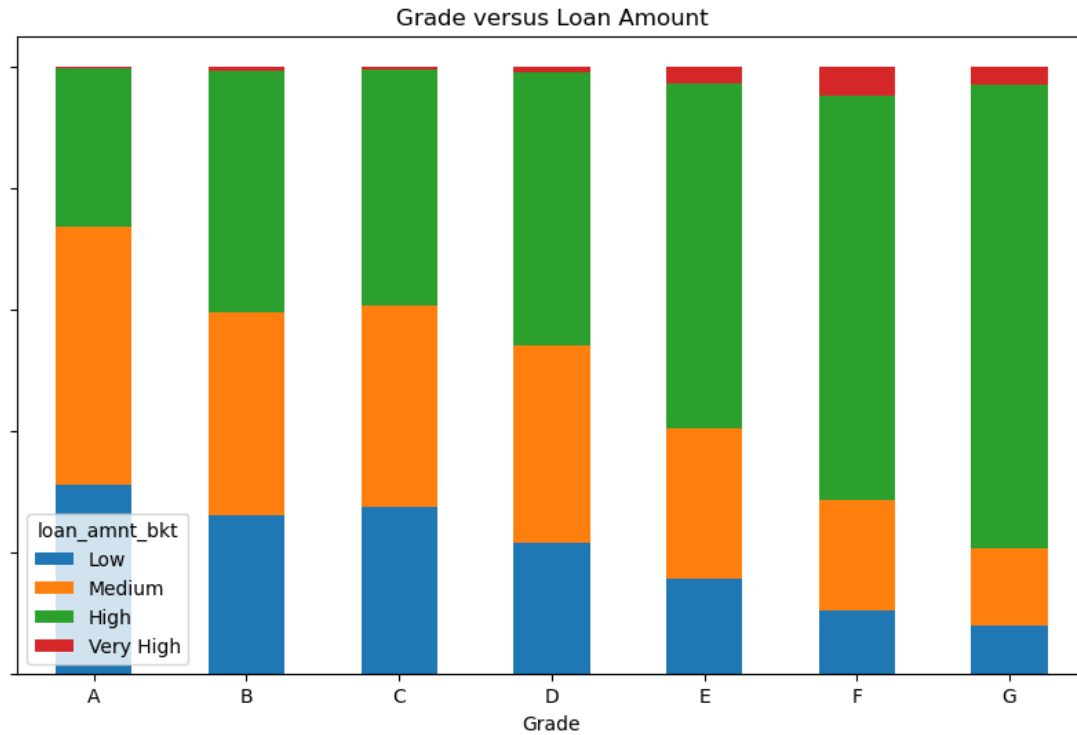
LC – Loan Amounts with their Grades

- High Grade loan types like G and F are mostly used in HIGH LOAN AMOUNT deals whereas Low Grade loan types like A and B are mostly used in LOW LOAN AMOUNT deals.

- Highest Charged off loans are found in Grade F followed by G.

- LC should tighten issuing policies for the Grade F and G

LC Charged Off - Loans pattern against borrower's annual income

- Medium (5k) ad High (10K) range loans are most issued to the borrowers and most defaulted by the borrowers.
- These range loans are issued most to the borrowers having annual income > 50000.

Annual Income Group versus Loan Status
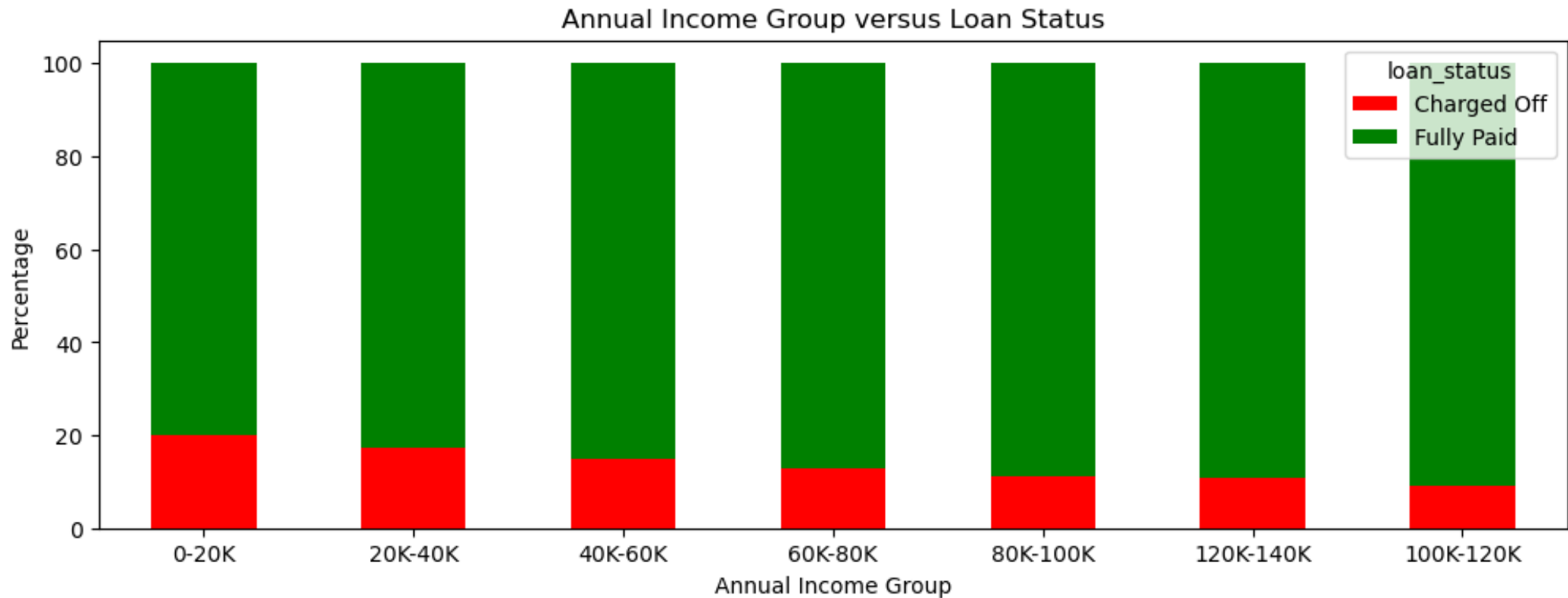
Borrower Annual Income vs Loan Status

- Charged Off loan's percentage is found at high in **Low Income groups** and decreased consistently while the annual income grows of the borrower.

LC – Loan segments vs Interest rates

- Highest interest rate loans are tended to charge-off. Charged Off loan rate is growing while loan interest rate is increasing.
- In Very-high amount loan group, when the installment amount is relatively less to "Fully Paid loans", Still borrower are tending to default their loans.
- **Insight :Borrowers are taking high range loans with low-interest rate and defaulting over the period.**

Interest Rate versus Loan Status

Loan amount bucket versus Installment

Loan Amount Bucket versus DTI

Annual income Bucket versus DTI
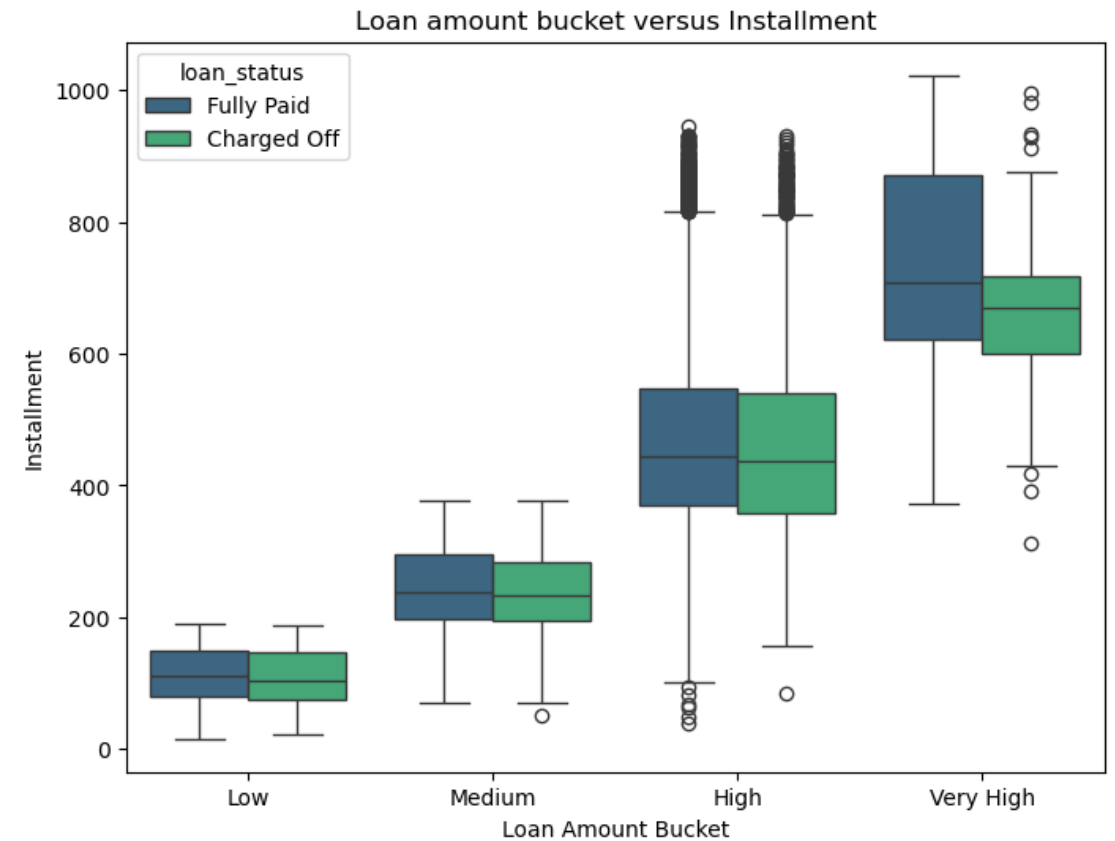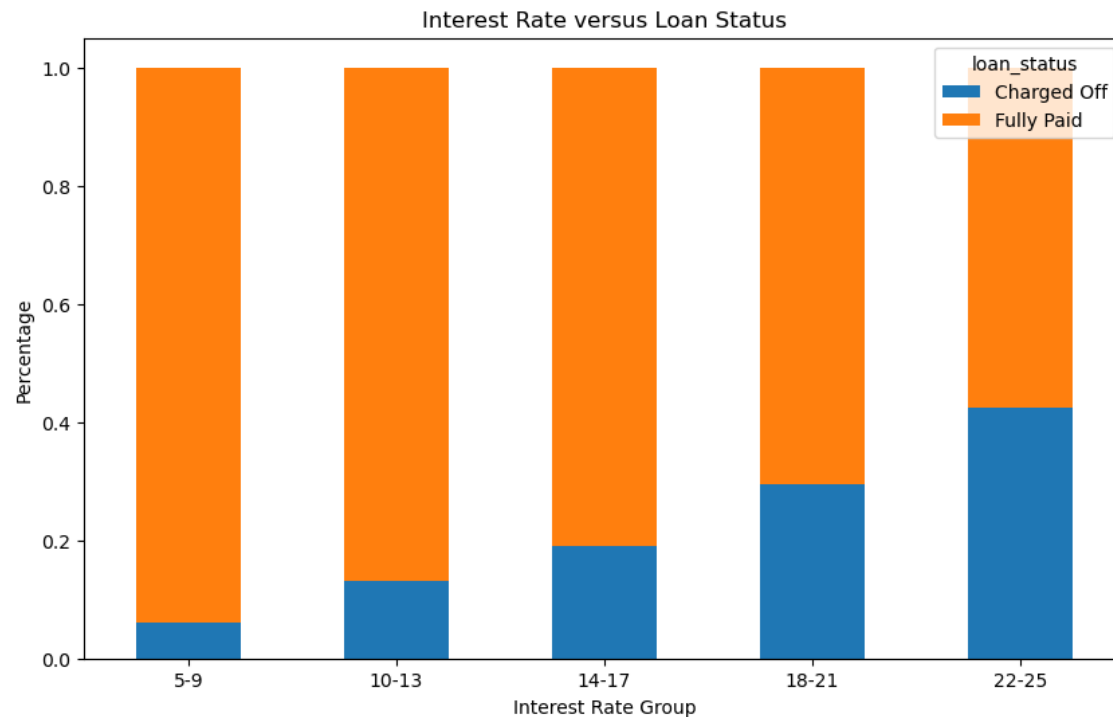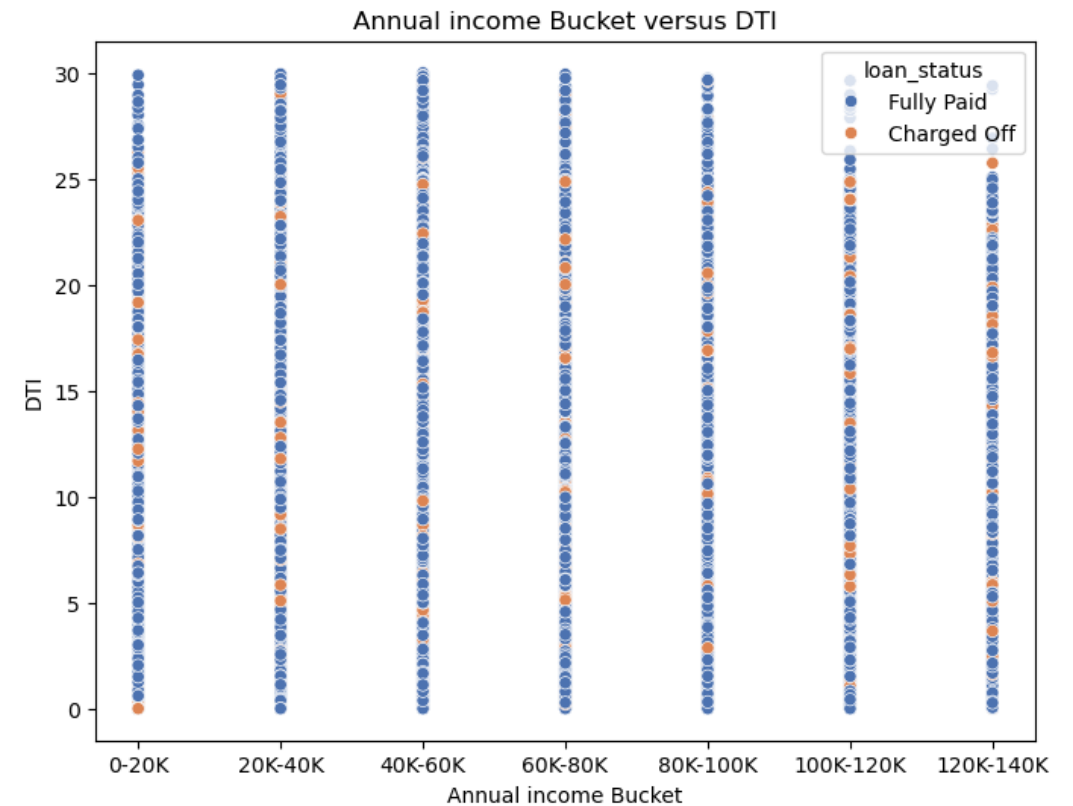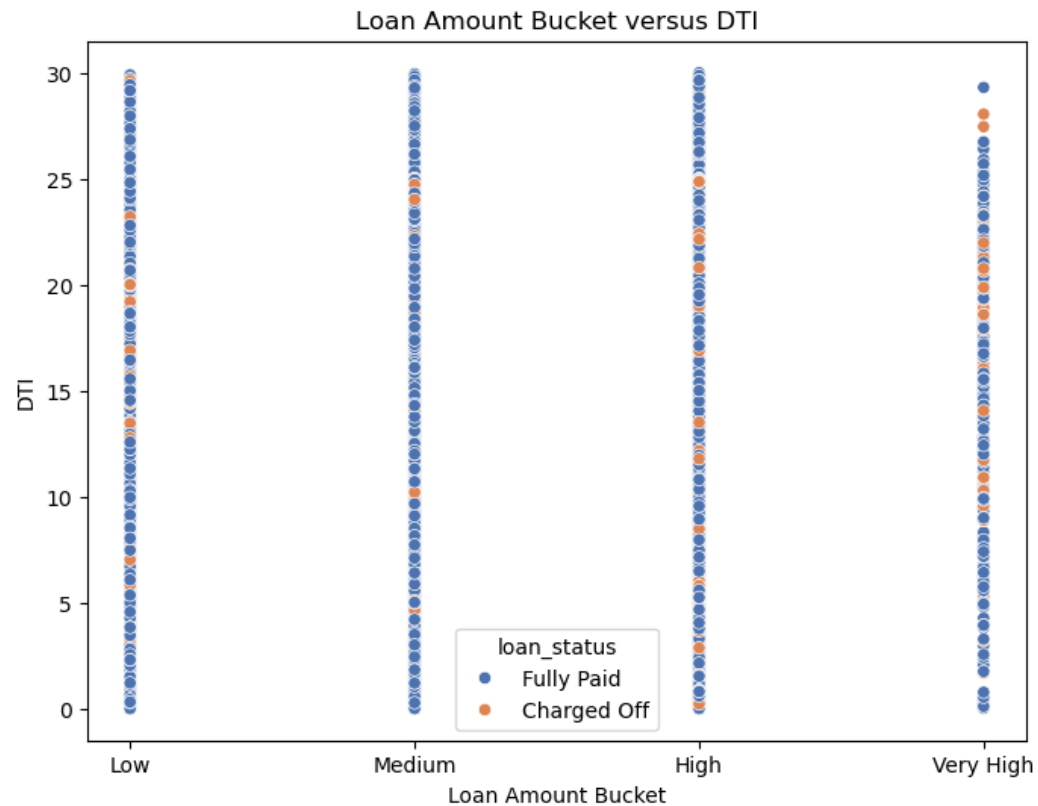
Borrower's DTI relation with Annual Income and Loan Amount

- When DTI is either low or high , the charged off loans are found more in the HIGH and VERY HIGH range loan amounts.

- When DTI is low, the charged off loans are found less in borrowers with low-income range.

- **DTI is not giving any strong indication whether loan will be charged off or fully paid. Along with DTI, LC should consider other factors like loan amount and interest rate also to analyze the risk.**

Borrower Employee Length vs Purpose in charging off loans

- In Overall, Borrowers with employee length 10+years are more defaulters than any other employee length.
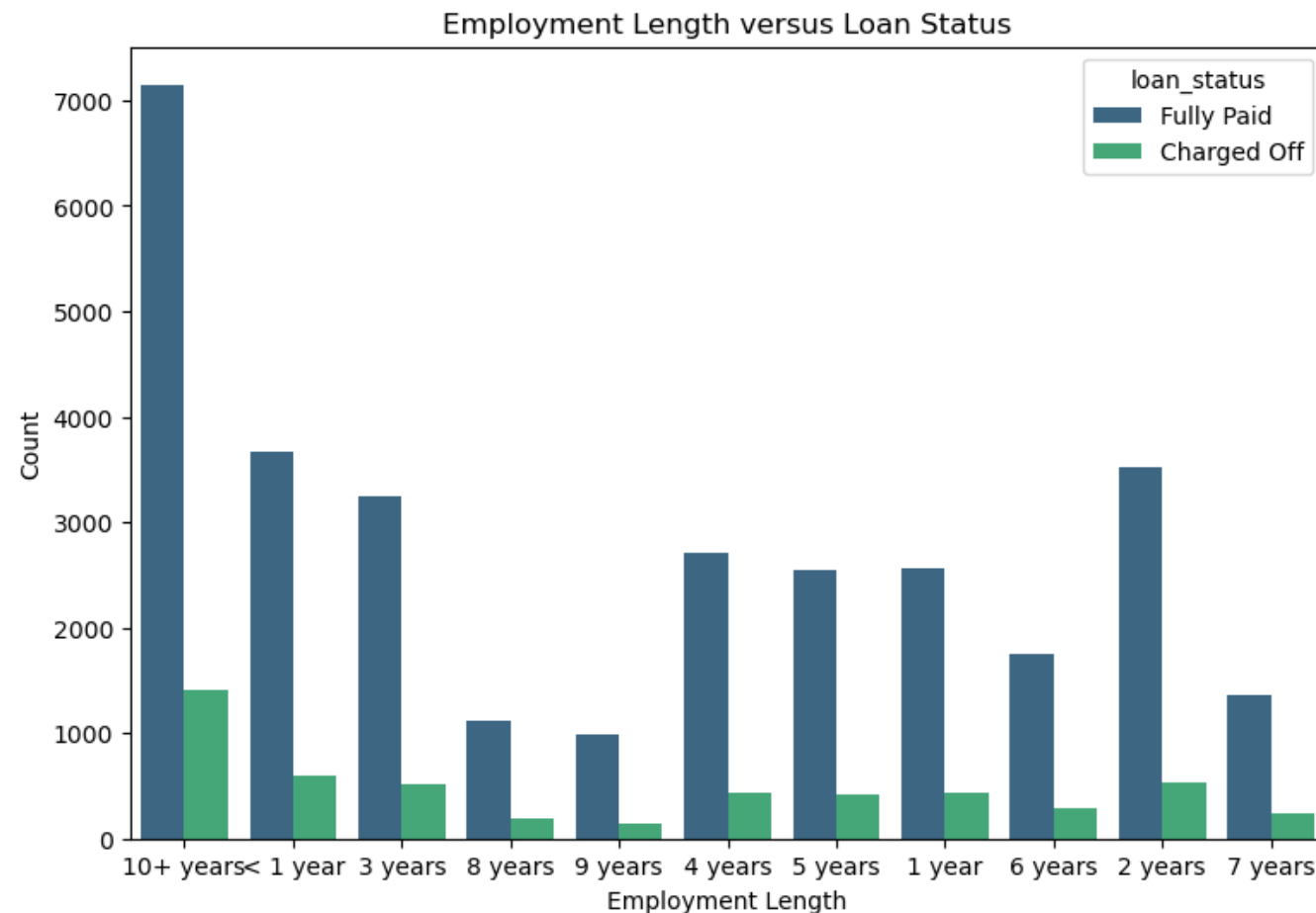
Loan Term Impact on Loan Status
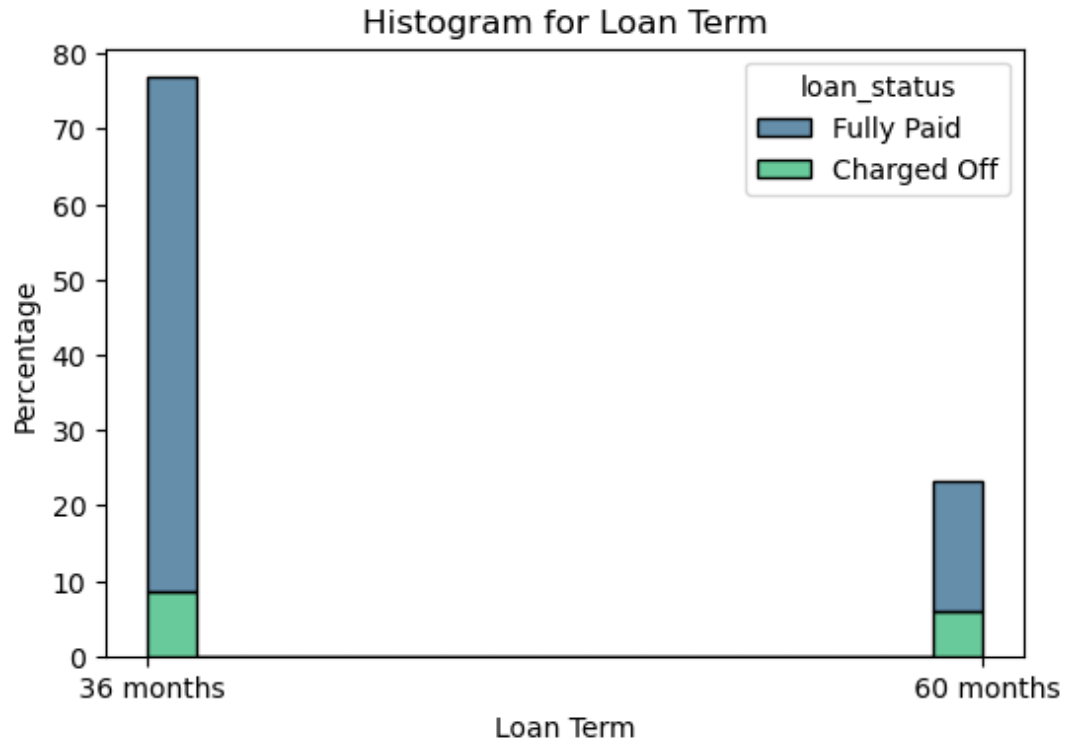
- Loan term of 36 months has the highest percentage of fully paid loans. Loan term of 60 months has the highest percentage of Charged off Loans.

- In Average, the loan amount lost in charged off loans is higher in 60 months term when compared to 30 months term.

Annual Income Group versus Revolving Balance

Revolving Balance versus Installment (CHARGED OFF loans)
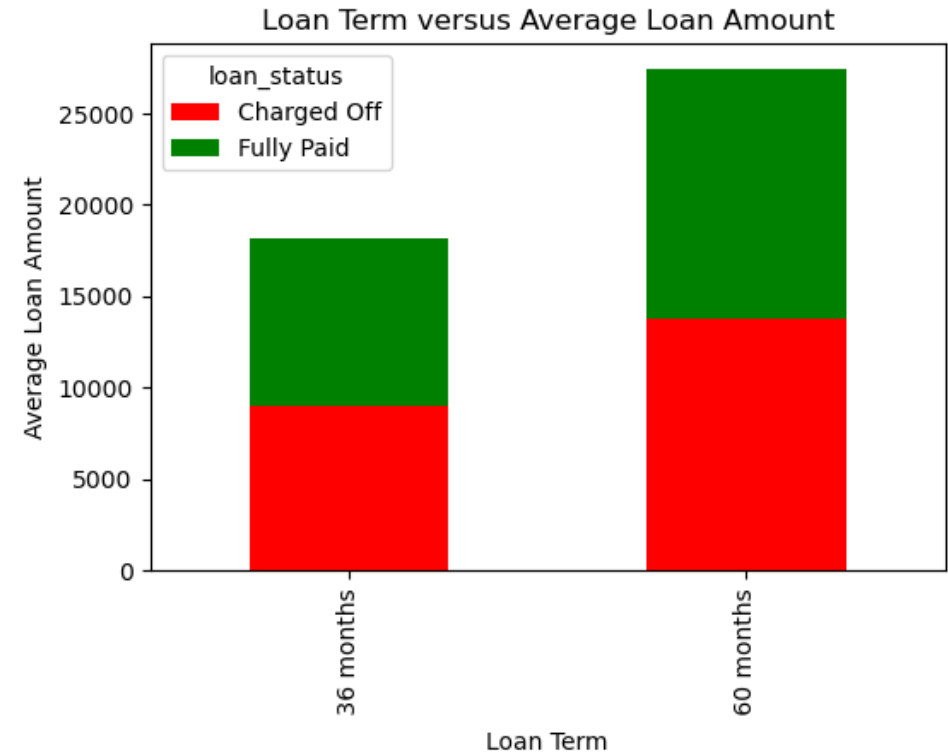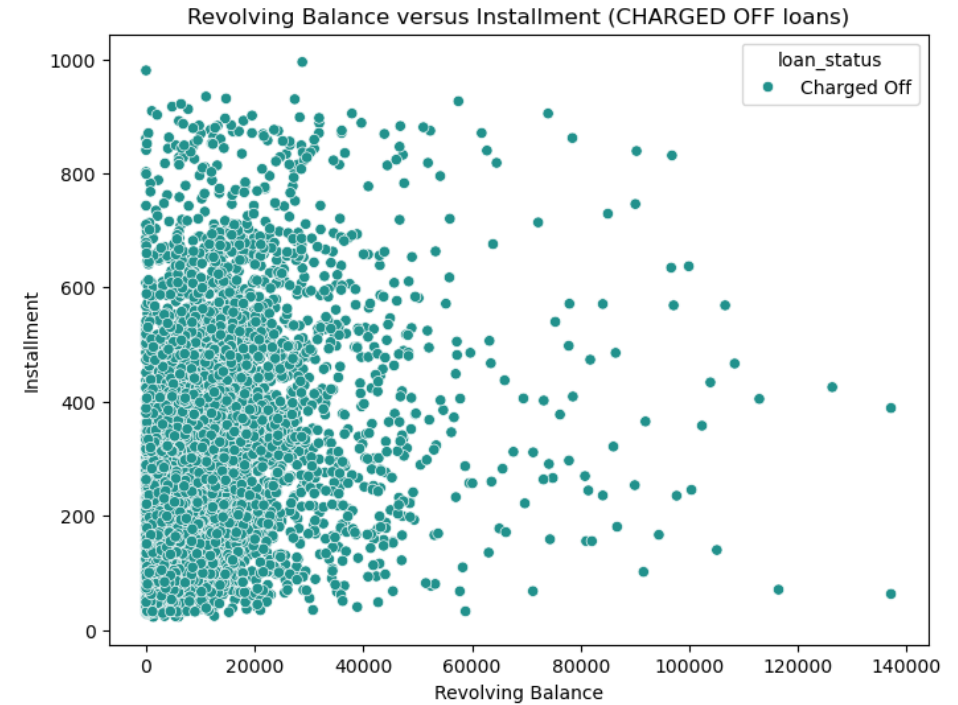
Borrowers Revolving Credit Vs Annual Income Vs Loan Installment

- In all Income groups, Average revolving utilization rate is above 50% for all Charged Off loans.

- There are many charged off loans with low installments even though borrowers have low revolving balance. In this case, some other reasons are influencing them to default their loans.

- **LC should not issue loans to borrowers with average revolving utilization > 50%**

Open Accounts versus Loan Status

Open Accounts versus Revolving Balance
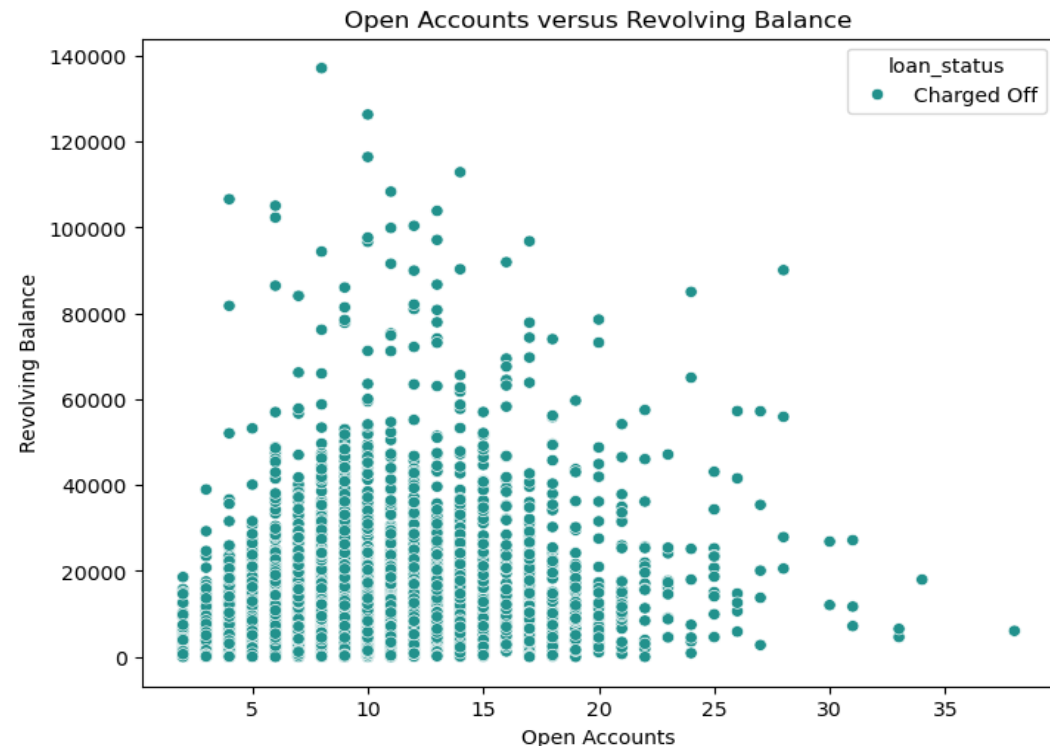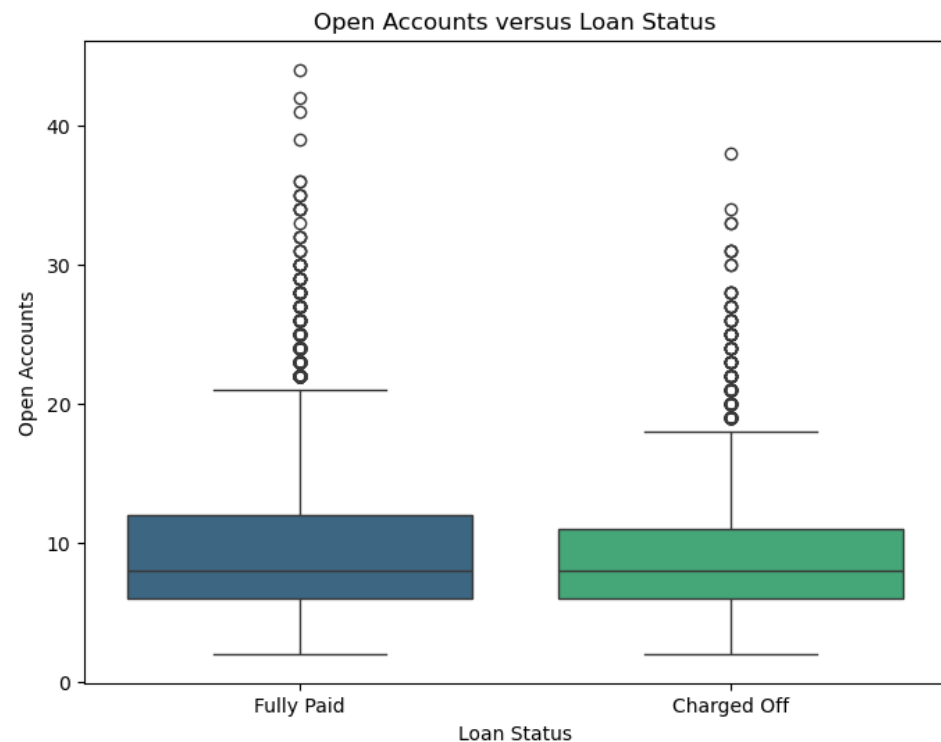
Borrower open credit lines, Revolving balance Vs Loan Status

- For Charged Off loans, the open credit accounts are 7 in median.

- There are many charged off loans which has higher number of open credits with lower revolving balance. This indicates that borrowers with higher number of open credit lines and spending huge and unable to maintain low revolving balance.

- LC should not issue loans to borrowers with higher revolving balance with high number of open credit lines.

Multivariate Analysis

LC Pair plot on Annual Income, Loan Amount, Interest rate (Charged Off Only)

- High loan amounts with low installment are leading to default their loans. Here, the term could be more, and borrower is unable to pay back due unforeseen conditions.

- Loan Amounts are equally distributed on the borrower annual income. It means - Borrower with either 50k, 100k and 150k can avail the maximum loan amount i.e., 30K.

- **Insight: Lending Club should restrict high loan amounts to the borrowers of higher income.**
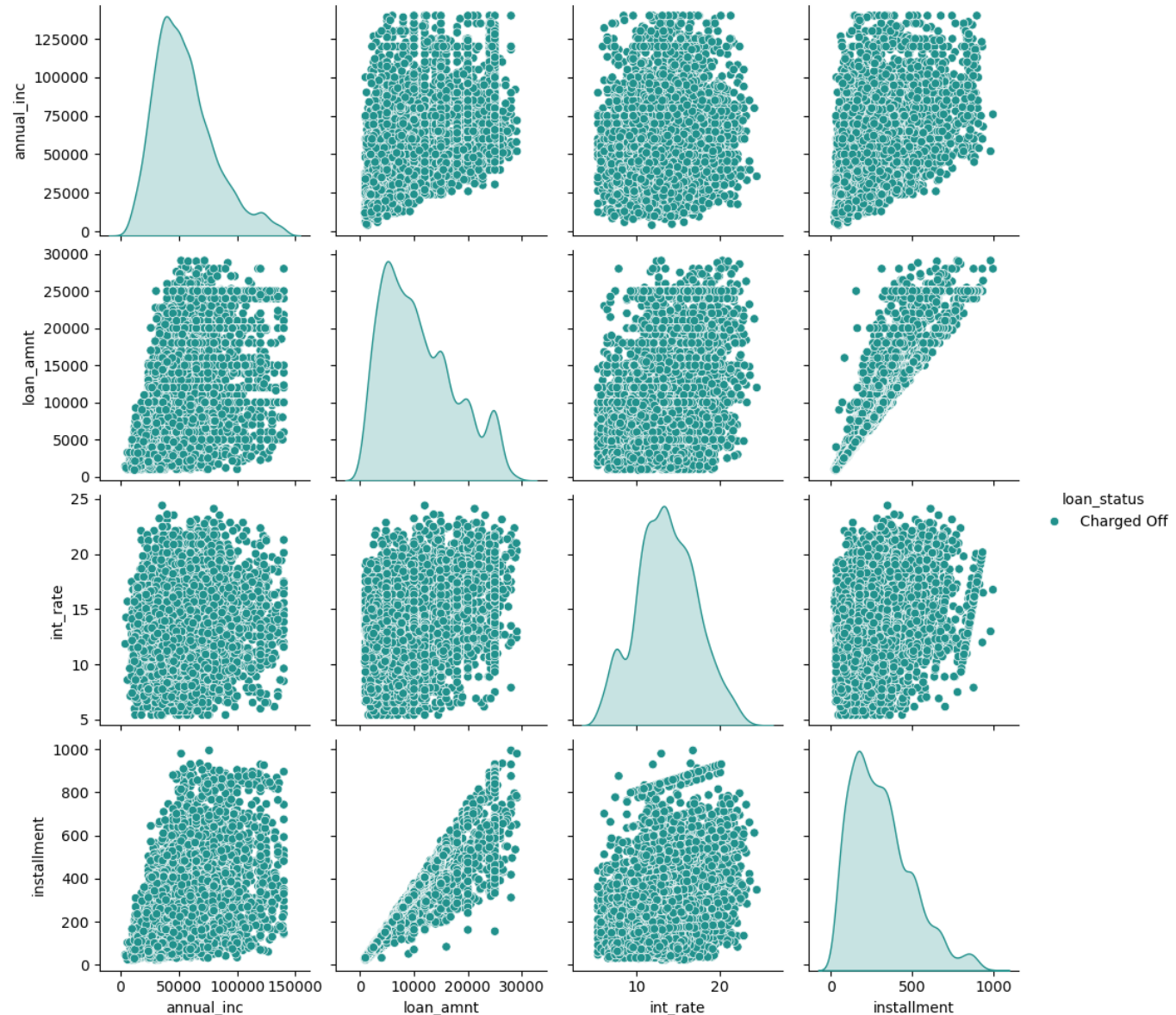
Correlation Matrix on Key Numeric values

- Loan Amount, Installment, Funded Amount, Funded Amount by Investors are having strong positive correlation

- DTI is having <mark>very weak correlation</mark> with all other elements - Loan Amount, Installment, Funded Amount, Funded Amount by Investors, Annual Income

- Annual Income has <mark>weak correlation</mark> with all other elements - Loan Amount, Installment, Funded Amount, Funded Amount by Investors, DTI

Correlation Matrix

	annual_inc	loan_amnt	funded_amnt	funded_amnt_inv	installment	int_rate	dti
annual_inc							
loan_amnt	0.37						
funded_amnt	0.37	0.98					
funded_amnt_inv	0.35	0.93	0.95				
installment	0.37	0.93	0.96	0.89			
int_rate	-0.026	0.25	0.26	0.25	0.23		
dti	-0.077	0.091	0.09	0.099	0.078	0.11	
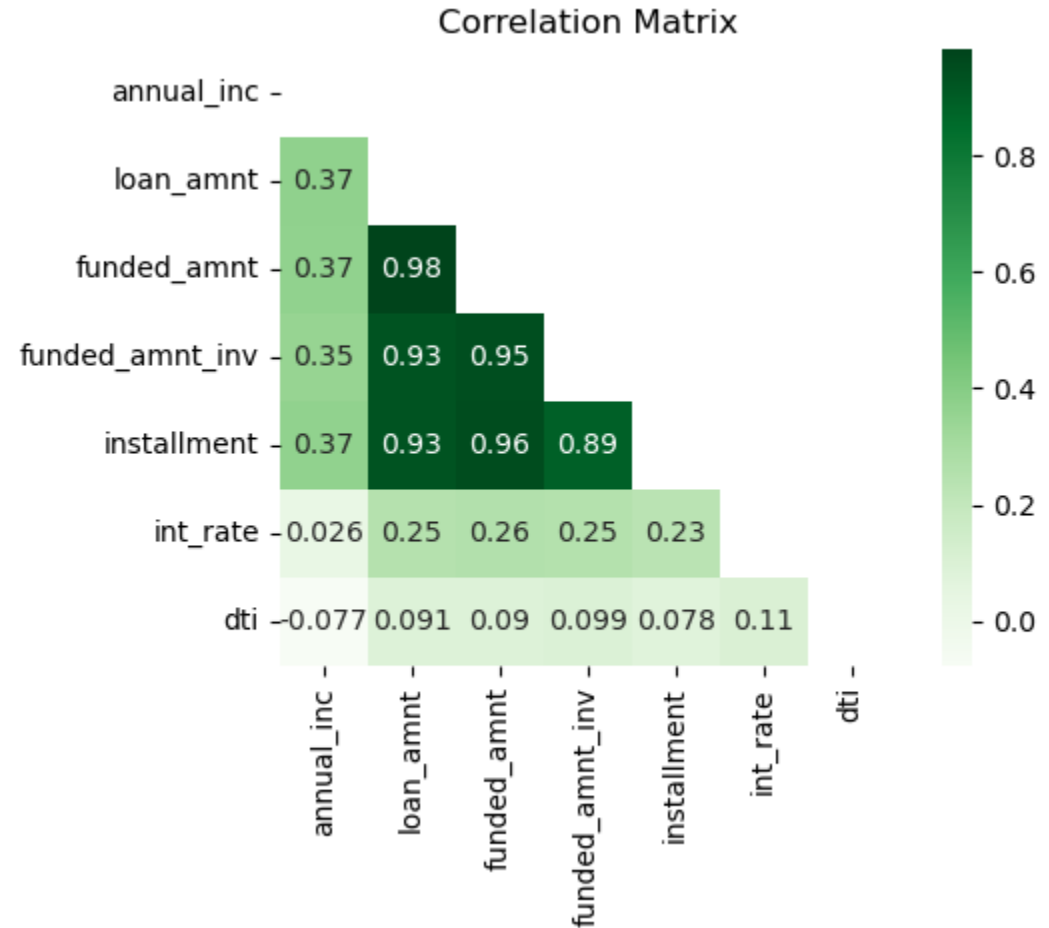
Employment Length versus Purpose for Charged Off Loans
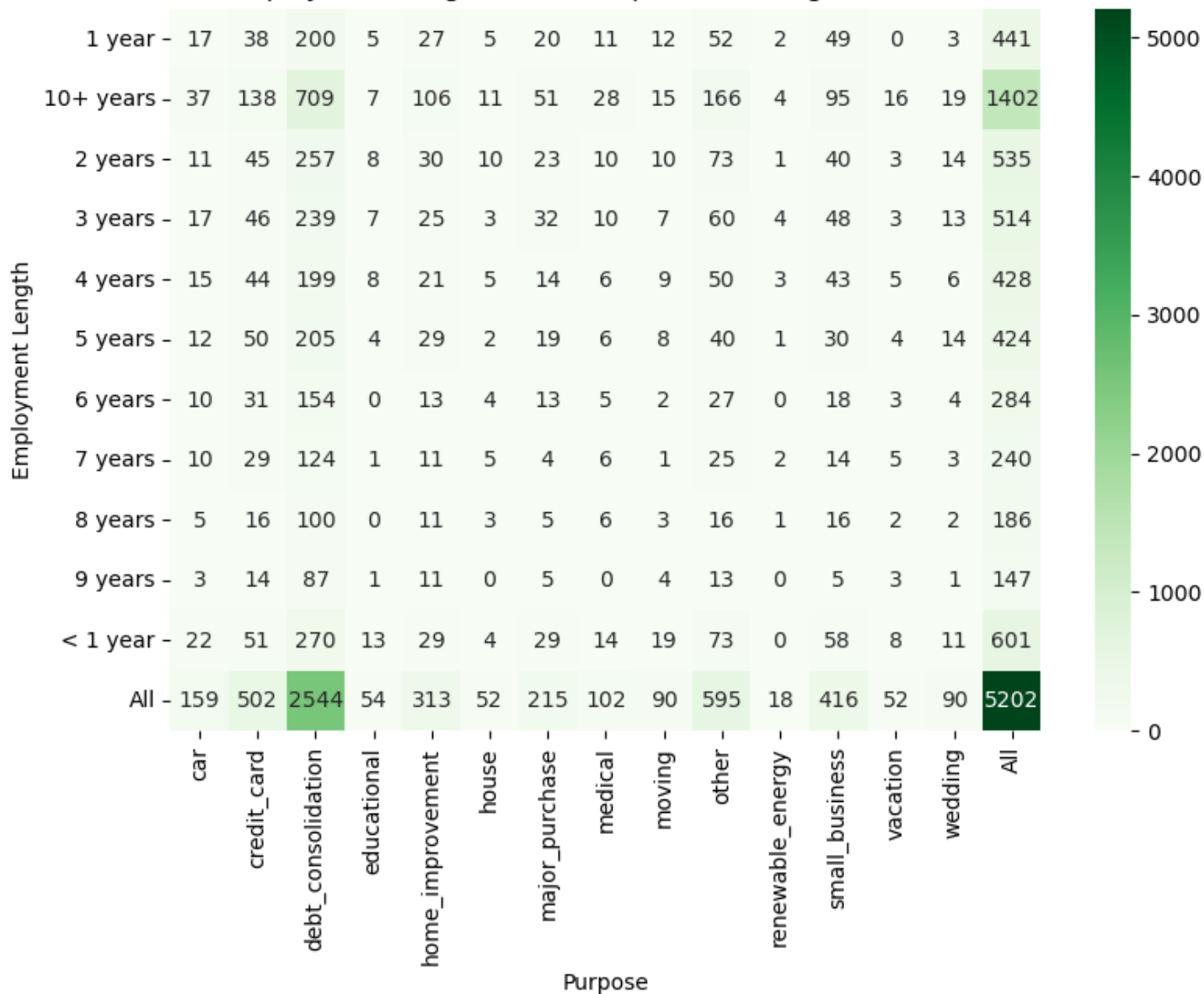
Defaulters by Employee Length and their purpose

- In Overall, Borrowers with employee length 10+years are more defaulters than any other employee length. That too, they charged off loans are with the purpose "debt consolidation".
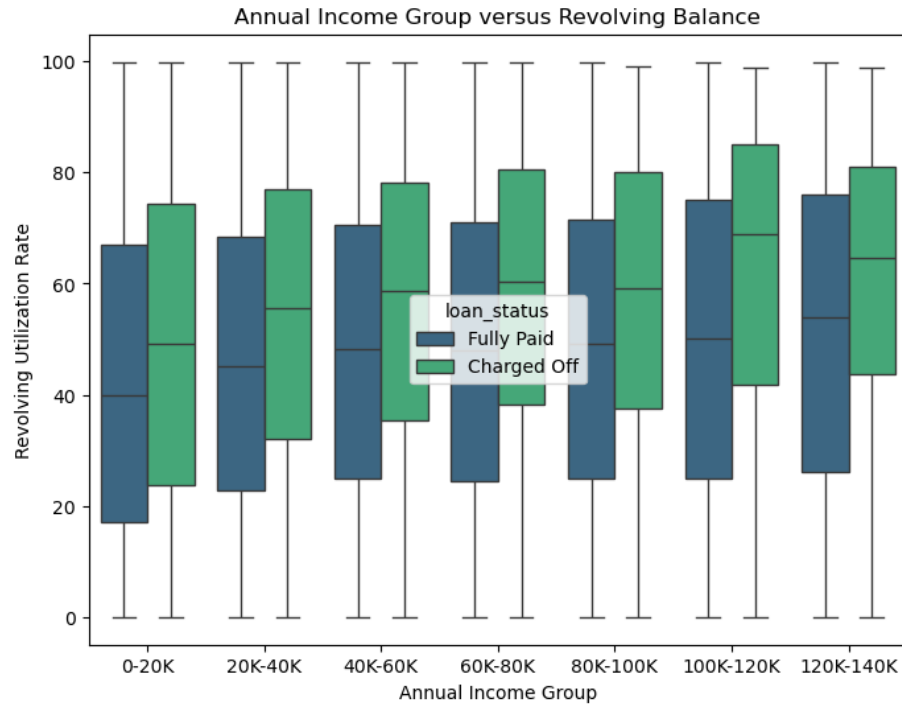
- **So, LC should add more verification checks for this combination while issuing loans .**

LC verification on Home ownership Vs Loan Status.

- There are significant charged off Loans from Borrowers who are residing in RENT, MORTGAGE , OWN houses but they didn't undergo verification process.
- **This gives insight of weak Verification process from Lending Club on borrowers' homeowner ship. It needs to corrected.**

- At the same time, there are significant borrowers who got verified On home ownership but still Charged off their loans.
- **It gives insight of some other factors dominating the borrowers profile to lead defaulting condition.**

Home Ownership versus (Loan Status and Verification Status)

home_ownership	Charged Off-Not Verified	Charged Off-Source Verified	Charged Off-Verified	Fully Paid-Not Verified	Fully Paid-Source Verified	Fully Paid-Verified	All
MORTGAGE	836	489	1002	6080	3341	5273	17021
NONE	0	0	0	3	0	0	3
OTHER	10	1	7	42	9	29	98
OWN	187	117	139	1272	625	635	2975
RENT	1109	827	903	7155	4268	4218	18480
All	2142	1434	2051	14552	8243	10155	38577

loan_status-verification_status

LC – Loan Term, Interest Rates, Loan Grades relation in Charged off loans

- - Grades G and F are the type of highest defaulted loans in both Loan Term 36 months and 60 months

- **However, 60 months term is having highest defaulting loans as their loan interest is high.**

Term versus Interest rate with grade for Charged Off Loans
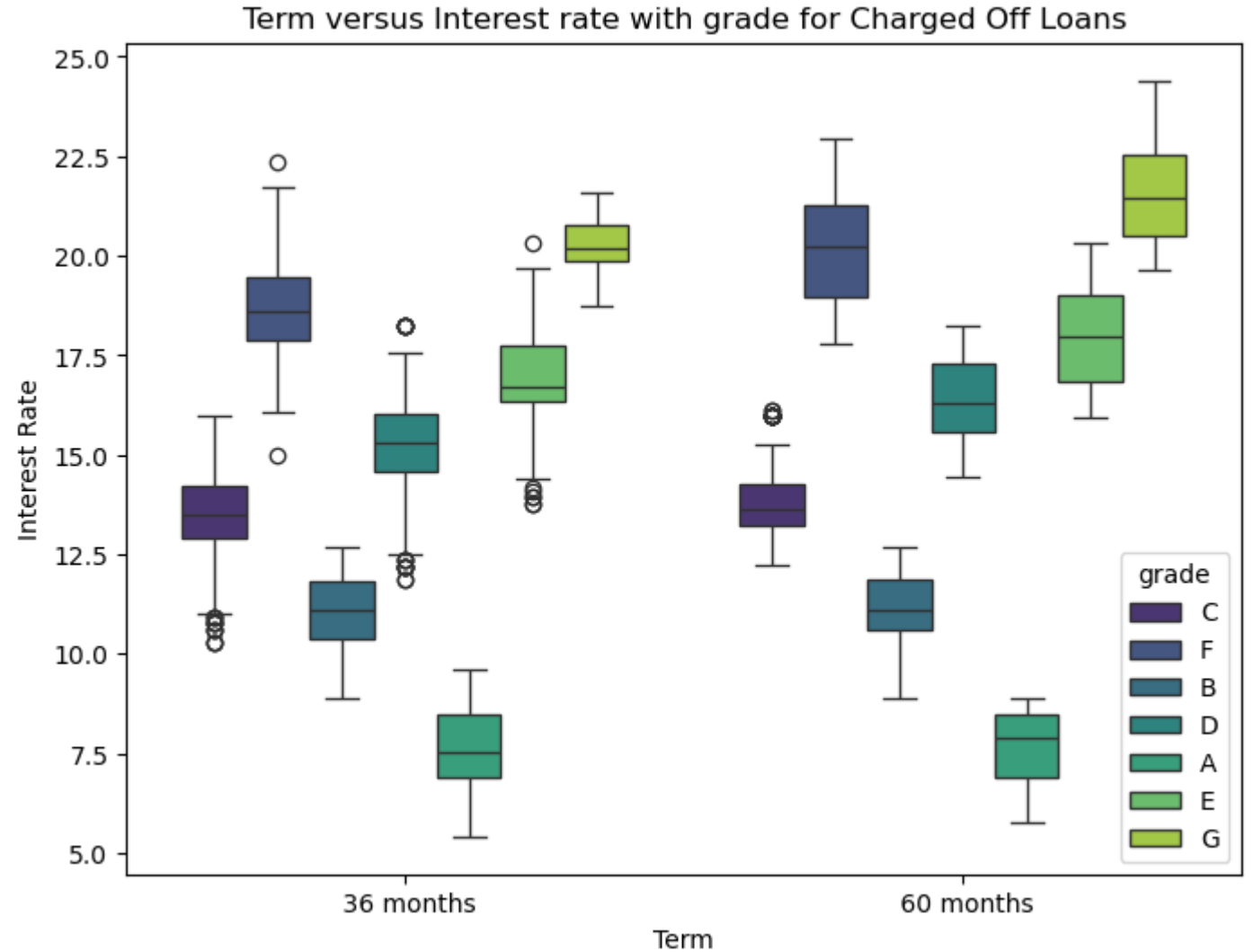
Summary

Lending Club (LC) experienced significant growth from 2007 to 2011. However, during this period, they struggled to control the increase in charged-off loans, with the latest charged-off percentage reaching approximately 14.5% by 2011.

Around 45% of LC's loans were not verified by their team, indicating that the verification process was insufficient. To address this issue, LC should enhance their verification process by adding additional checks to reduce the number of charged-off loans. Interestingly, it was observed that "Not-Verified" loans were also fully paid off in higher numbers compared to verified loans.

The highest number of borrowers took out loans from LC for the purpose of "Debt Consolidation." However, the highest percentage of charged-off loans was found in the "Small Business" category.

LC had the highest count of borrowers with loan status from California, but the highest percentage of charged-off loans was found in Nevada (NV).

Borrowers with home ownership statuses of Rent and Mortgage were availing the highest number of loans and had a higher percentage of charged-off loans.

Summary (Cntd.)

- Lending Club (LC) borrowers have an average income of $60K and typically take out loans of around $10K.

- Borrowers often have a high Debt-to-Income (DTI) ratio of 15 and an interest rate of 12.5%.

- Loans in the $5K to $10K range are the most issued and the most frequently defaulted, particularly among borrowers with annual incomes over $50,000.

- LC investors have lost 15% of their investment due to charged-off loans.

- LC should focus on strengthening verification for Loan types F and G, as these grades have a higher incidence of charged-off loans.

- The percentage of charged-off loans is significantly higher in "Low Income" groups and decreases consistently as the borrower's annual income increases.

- Borrowers are taking out high-range loans with low-interest rates and defaulting over time.

- Borrowers with employment lengths greater than 10 years are the highest defaulters, especially for loans taken out for "Debt Consolidation."

Summary (Cntd.)

Loans with a 36-month term have the highest percentage of fully paid loans, while those with a 60-month term have the highest percentage of charged-off loans.

The average revolving utilization rate for all charged-off loans is above 50%. Therefore, LC should avoid issuing loans to borrowers with an average revolving utilization rate greater than 50%.

Additionally, LC should refrain from lending to borrowers with high revolving balances and many open credit lines

High loan amounts with low installments are leading to defaults, so Lending Club should restrict high loan amounts to borrowers with higher incomes.

Recommendations

Lending Club (LC) should be vigilant with following measures to prevent loan defaults:

- Borrower above10 years experience for the purpose- debt consolidation
- Loans with 60 months term and high interest rates
- High Grade loans with high interest rates
- Any high amount loans with higher interest rates
- Limit loan amount based on borrowers revolving balance, bankruptcy history
- Small business loans with high loan amount
- Borrower with revolving utilization above 50%
- Every single loan application should be verified before issuing the loan.
- Tighten loan verification process in Nevada which has highest % of defaulters.