Lending Club Case Study

Submitted by :
Suhas Naik

Subrata Das

Lending Club Case Study

Business Case

- Risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Current Approach / Roles and responsibility

Roles and Responsibilities :
In this project, my roles and responsibilities include
- Data Understanding
- Data Cleaning
- Data Manipulations and wrangling
- Univariate Analysis
- Bivariate Analysis
- Insights
- Conclusions

Aim

- The Lending club case study data contains information about past loan applicants and whether they 'defaulted' or not.
- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

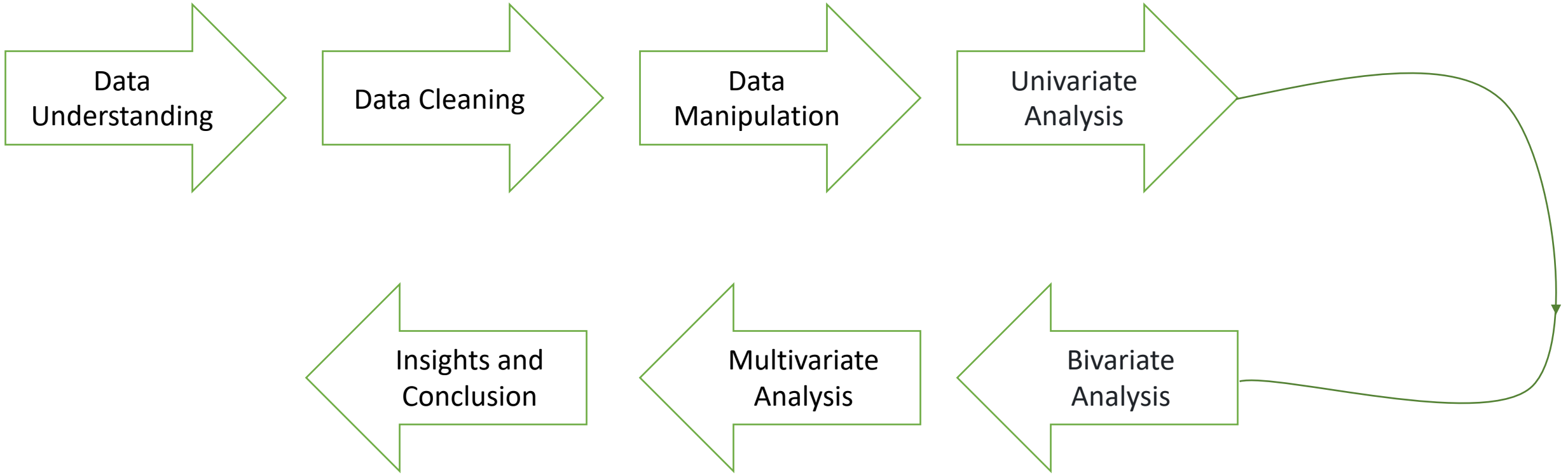
Project objectives

Identify patterns which indicate if a person is likely to default is the main objective.

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

Methodology



Analysis

Understanding the dataframe

- View Data: Use methods like head() and tail() to see the first and last few rows.
- Summary Statistics: Use describe() to get an overview of numerical columns.
- Check Data Types: Use info() to see the data types and non-null counts.

Data Cleaning

- Removing data based on the following factors
- 1. High number of Null/NAN Values
- 2. Single unique values
- 3. High number of unique values
- 4. Non numerical data (Member ID, Data Source etc)
- 5. Data that is only availble after the loan has been given
- 6. Handle Missing Values
- 7. Remove Duplicates
- 8. Correct Data Types¶

Analysis

Data Manipulations and wrangling

- Filtering: Select specific rows or columns based on conditions.
- Aggregating: Summarize data using group by operations (e.g., calculating averages).
- Feature Engineering: Create new variables or modify existing ones for analysis.

I see special characters in the below columns and will clean those :

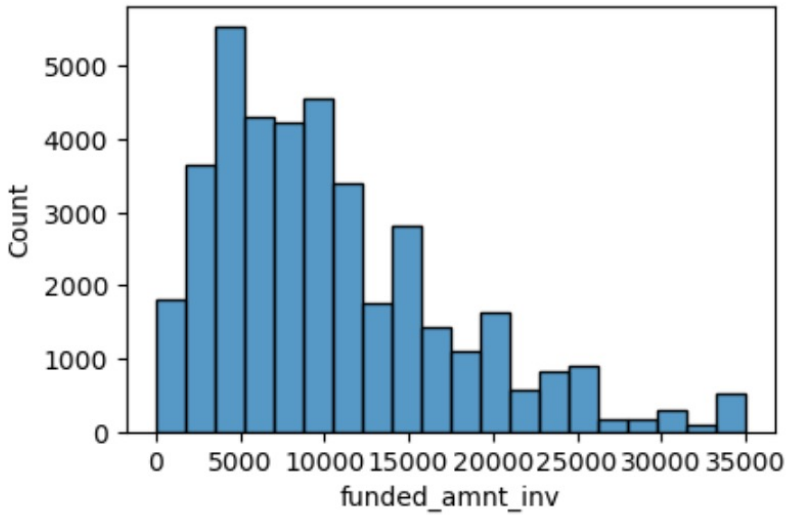
We can use the mode for emp_length (10+ years), revol_util (0%), and pub_rec_bankruptcies (0.0)

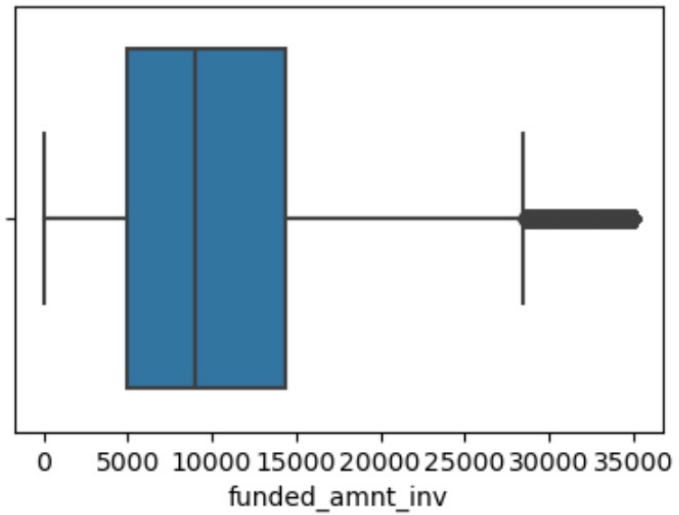
in each case, the mode appears significantly more frequently than the next most common value, making it the most representative and reliable estimate for imputation. Helps in retaining current distribution of data.

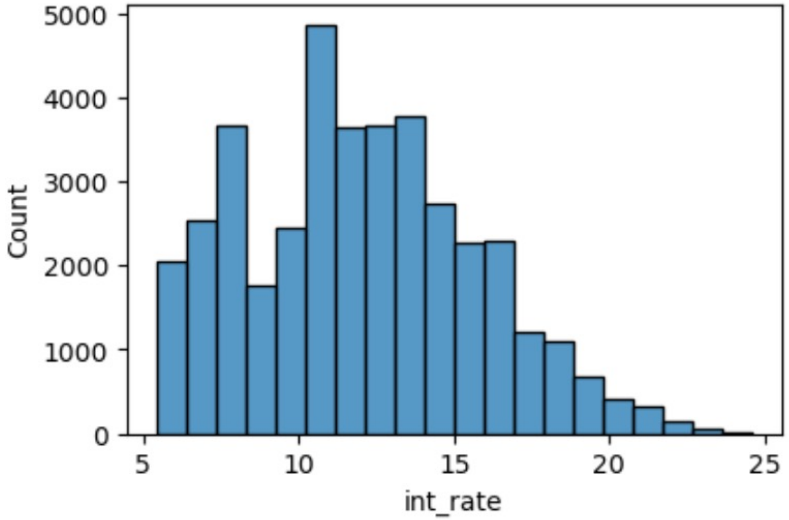
Analyze emp_length, revol_util and pub_rec_bankruptcies to determine the best value for imputing nulls

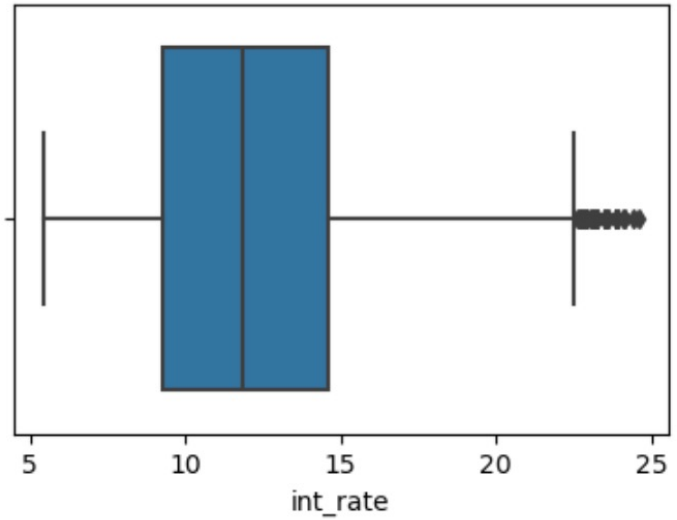
Univariate Analysis

Representing and interpreting a single variable at a time





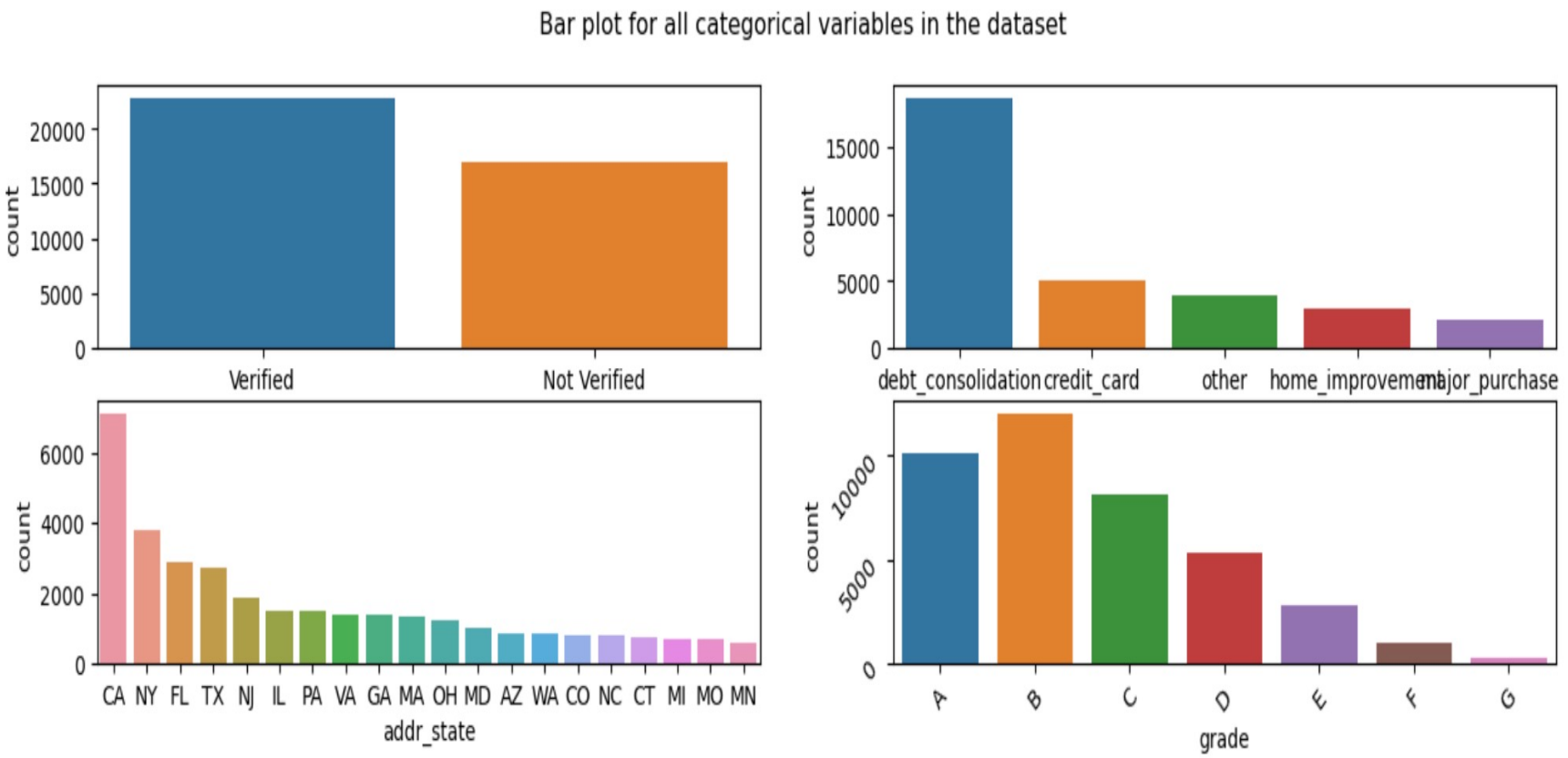




- Funded amount committed by investor is ranges **from 5000 – 15,000 (25% - 75% percentile)** which covers most of the data points
- Investors have less funded amount when the price is high which tends to say that **risk** is also high
- Avg. Interest rate **10-12%** is high when compared to others.
- The average interest rate on loans is **12.02%**, with a minimum of 5.42% and a maximum of 24.59%.
- Interest rate with higher slab is also **directly proportional** to the funded amount and where there is a risk of high default

Univariate Analysis

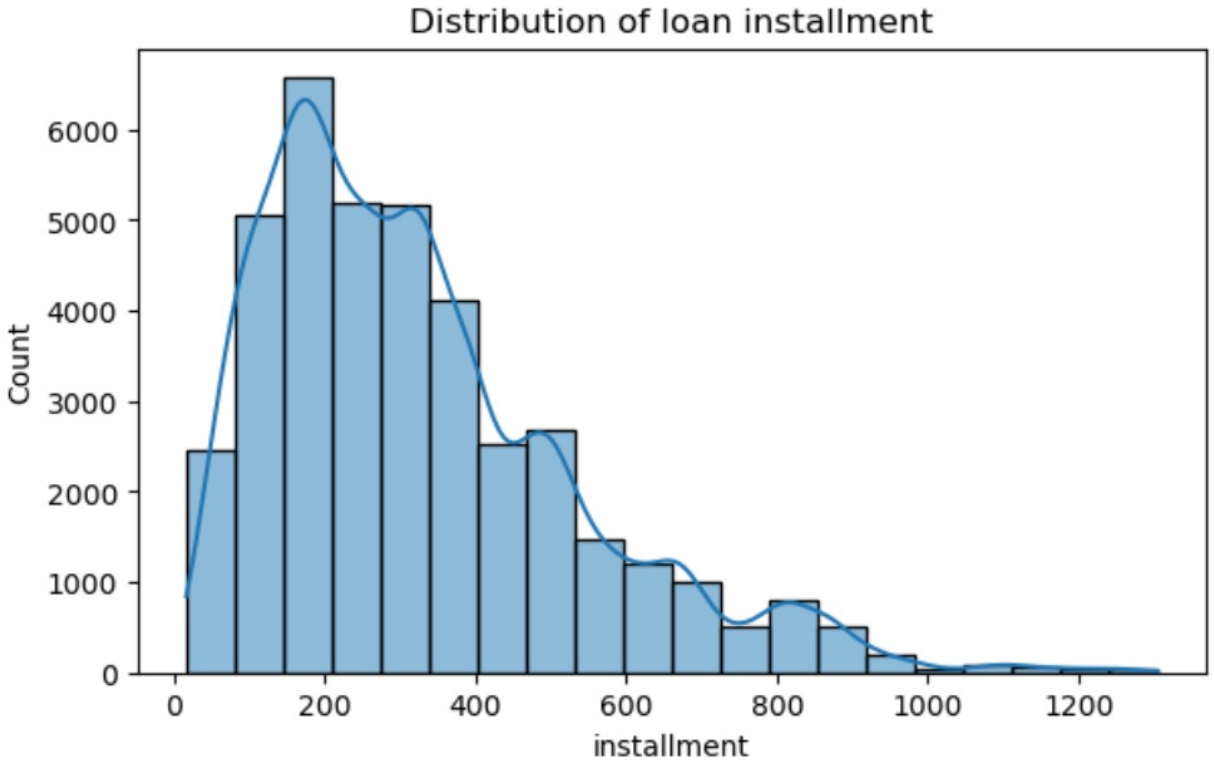
Representing and interpreting a single variable at a time

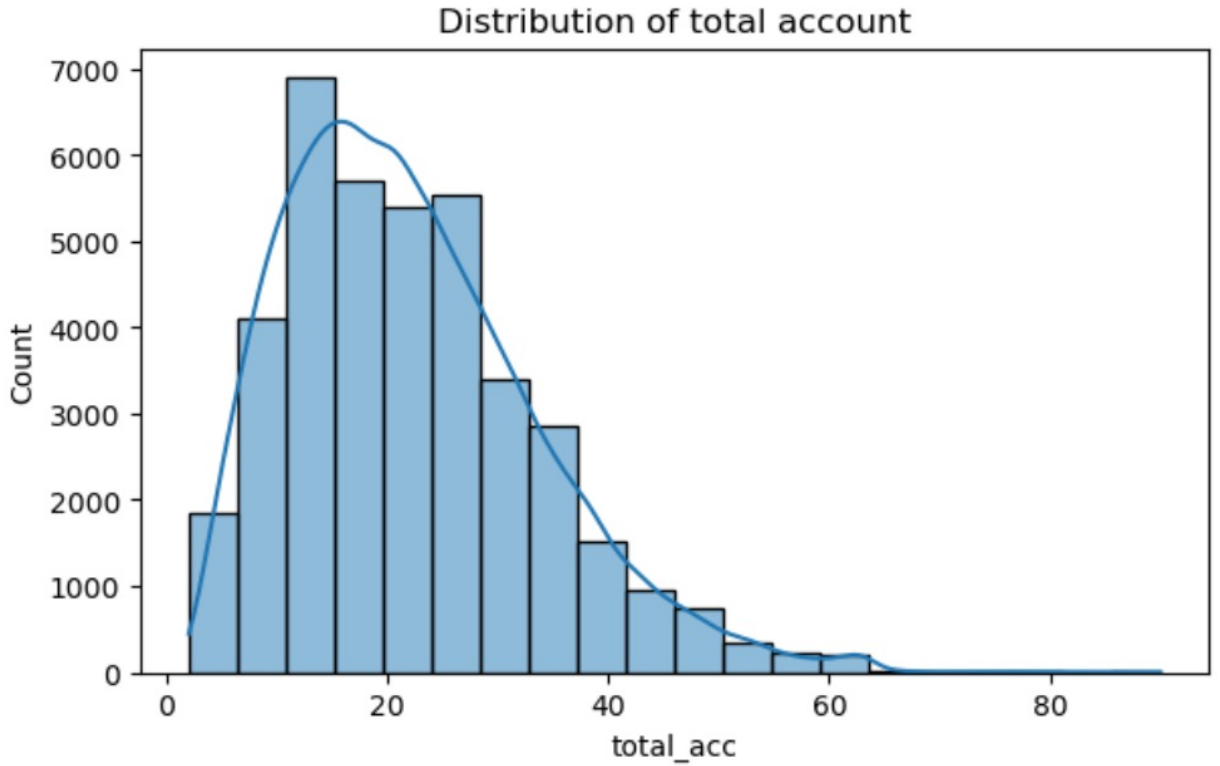


- **58%** of the accounts are verified by LC
- 42% is yet **not verified** and there is a risk of not verified account
- Major purpose of the loan is **"debt consolidation "**
- Major % of the accounts held in **CA (California)** which means that the marketing for CA which holds good in all sense which need to be implemented in other states as well in order to maximize the business
- ~12500 accounts, LC assigned loan **B grade.**

Univariate Analysis

Representing and interpreting a single variable at a time

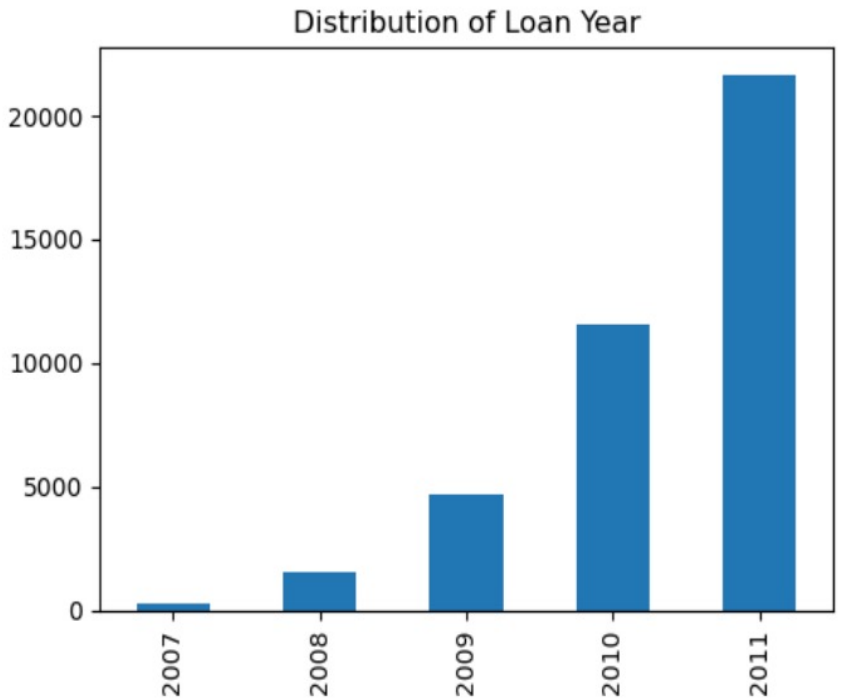


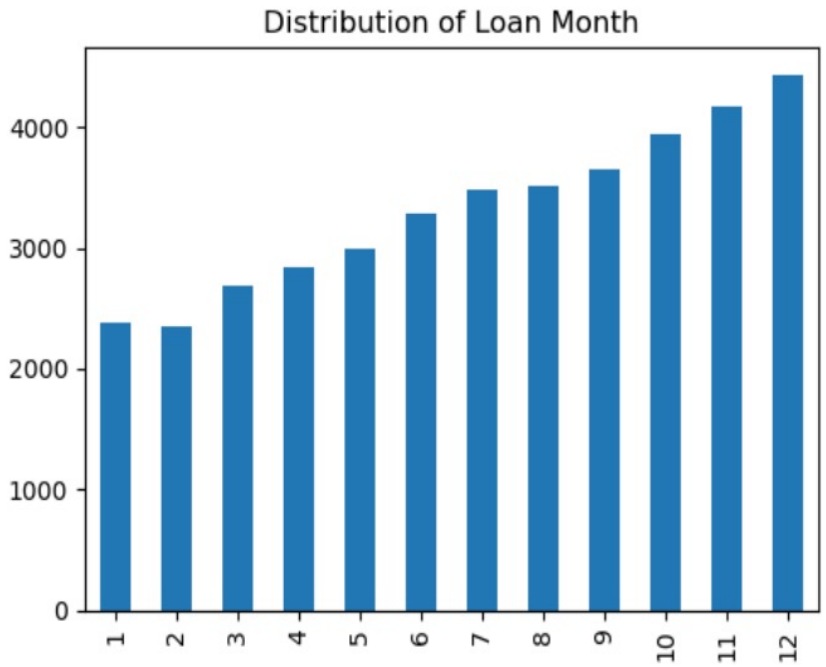


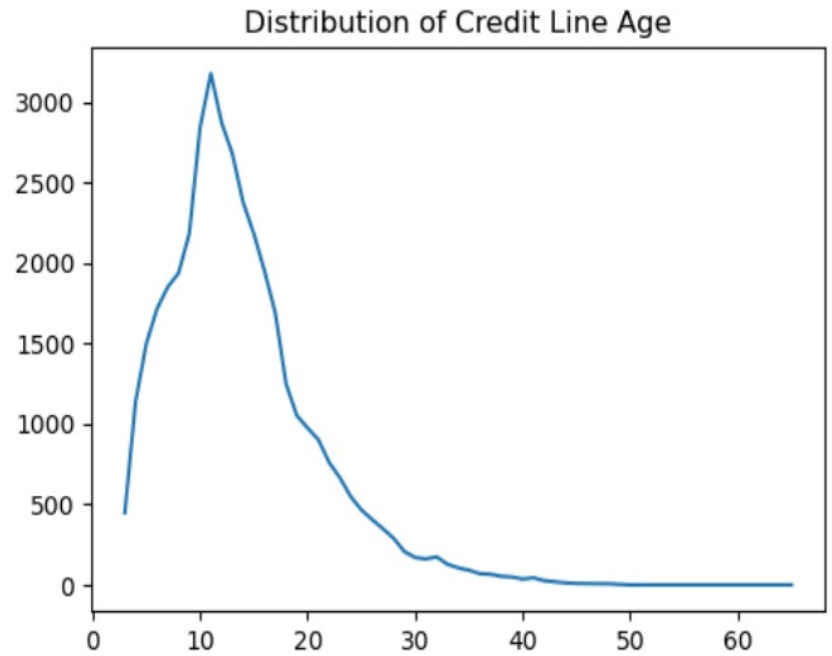
- The average interest rate on loans is **12.02%**, with a minimum of 5.42% and a maximum of 24.59%.
- The average **annual income(annual_inc)** of borrowers is 68,967, with a minimum of 4,000 and a maximum of **60,00,000**.
- The average debt-to-income ratio (DTI) of borrowers is **13.3**, with a minimum of 0 and a maximum of 29.99.
- Most loans are for a term of 36 months, with a total count of **39716** loans and **24,000** of them being source verified.

Univariate Analysis

Representing and interpreting a single variable at a time



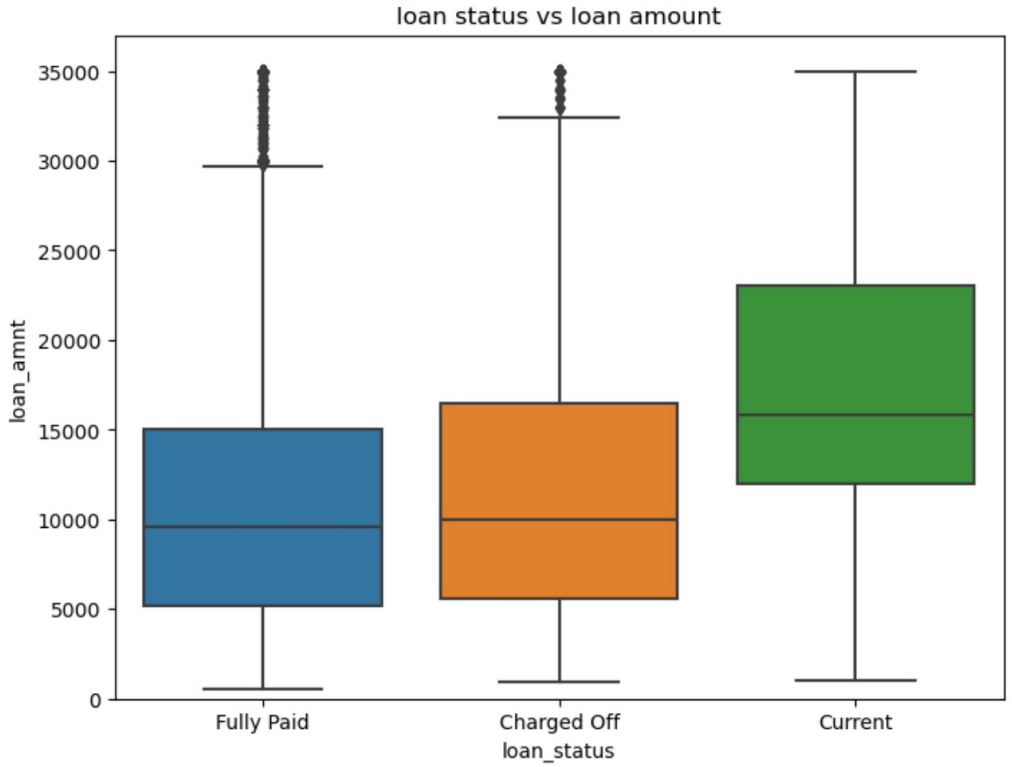


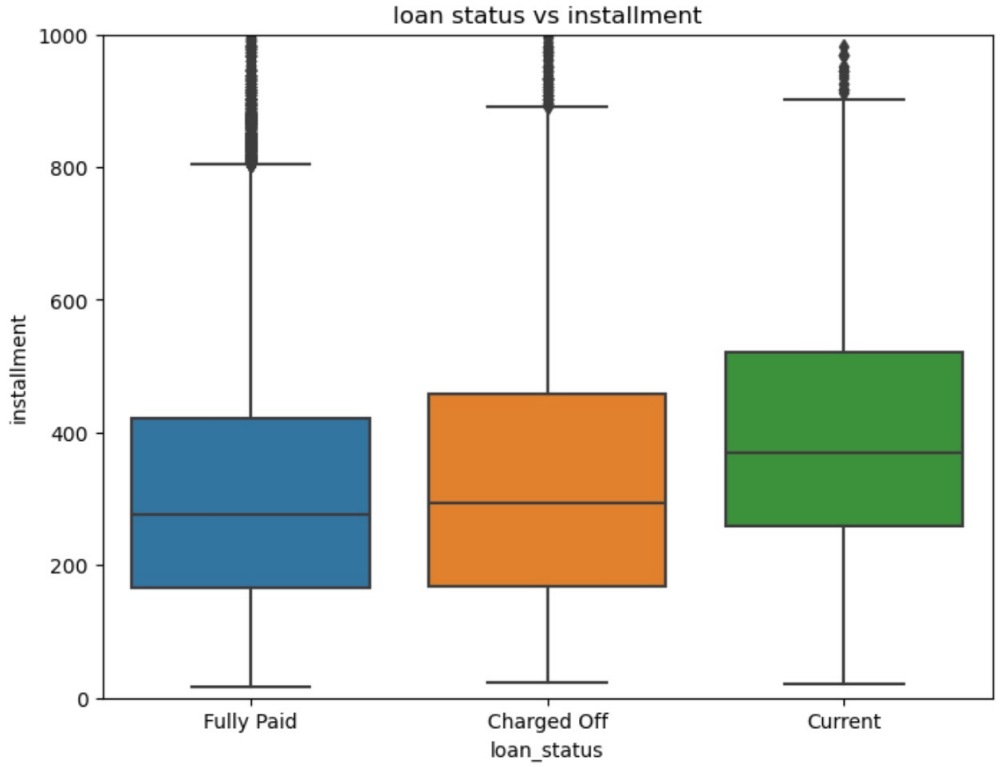


- **Loan Year:** The bar chart shows a significant increase in the number of loans issued from 2007 to 2011, with **2011** having the highest number of loans by far.
- **Loan Month:** The distribution of loans shows a clear seasonal pattern throughout the year. Loan numbers are lowest in January and February, then gradually increase with the highest number of loans occurring **in December (month 12).**
- **Credit Line Age:** The distribution of credit line age displays a pronounced **right-skewed pattern** with a peak occurring sharply between 10 and 15 years.

Bivariate Analysis

Representing and interpreting a single variable at a time



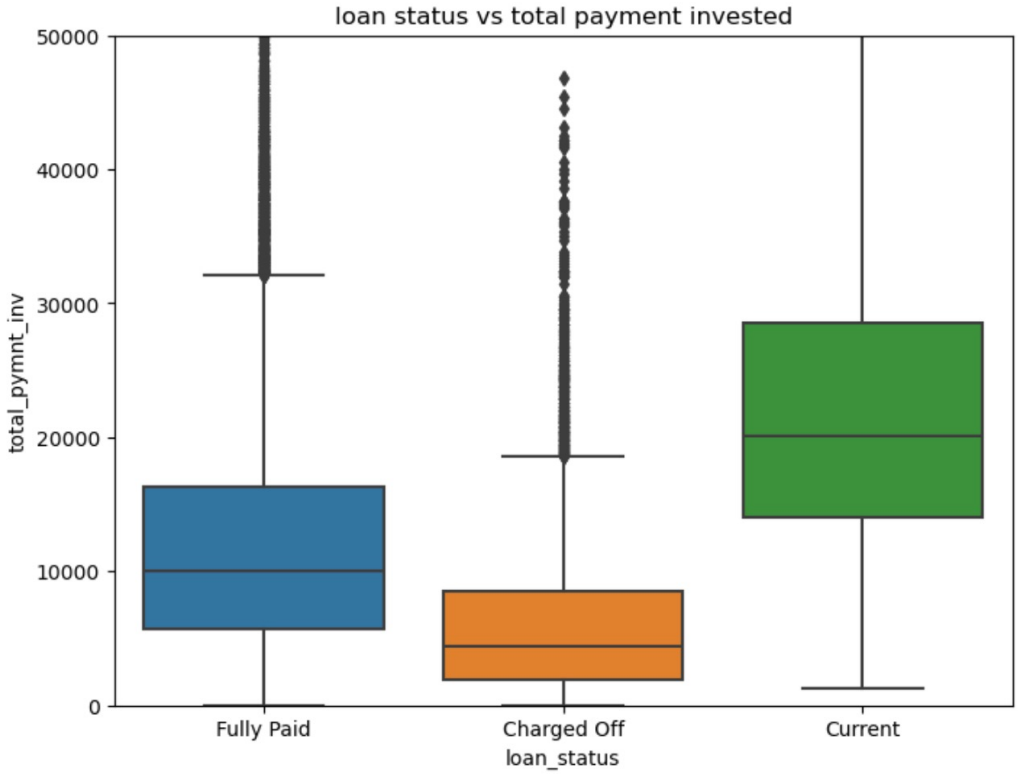


- Loan amount on the current status tends to have high % when compared to fully paid, charged off.
- 12,500 loan amount is the 25% percentile
- Installment(The monthly payment owed by the borrower if the loan originates) on the current status tends to have high % when compared to fully paid, charged off.

Bivariate Analysis

Representing and interpreting a single variable at a time

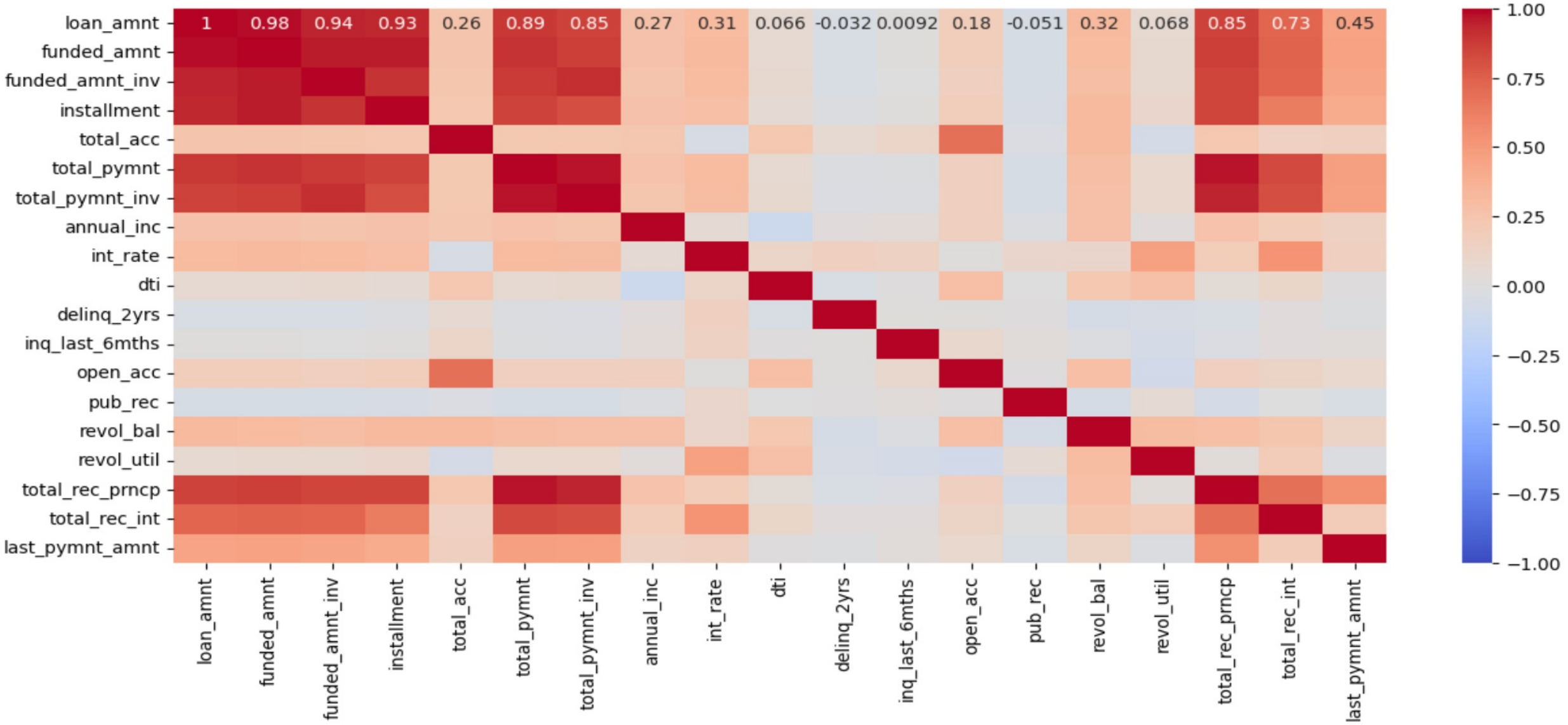
loan_status	home_ownership	count
Charged Off	MORTGAGE	2327
Charged Off	OTHER	18
Charged Off	OWN	443
Charged Off	RENT	2839
Current	MORTGAGE	638
Current	OWN	83
Current	RENT	419
Fully Paid	MORTGAGE	14693
Fully Paid	NONE	3
Fully Paid	OTHER	80
Fully Paid	OWN	2532
Fully Paid	RENT	15641



- Rent home ownership customers tends to have a loan than any others
- Targeting the people who wants to upgrade in their living status tends to avail loan
- Current loan status on mortgage home is higher when compared to rent and own house
- Total payment invested in current loan status seems positive
- The strategy on charged off loan status on spending total payment less is working good

Multivariate Analysis

Representing and interpreting a single variable at a time



Insights from the correlation matrix

- Loan amount(loan_amnt) has very strong positive correlation with funded_amnt, funded_amnt_inv which is an Indication that the profit in the operations to meet its commitments.
- Loan amount(loan_amnt) has very strong positive correlation with installment which is obvious since higher loan_amnt will lead to higher installment controlled for term.
- Loan amount(loan_amnt) has strong positive correlation with total_pymnt, total_pymnt_inv which is a strong indicator that the CFC is operationally efficient and prioritizes investors.
- Loan amount(loan_amnt) has weak correlation with annual_inc which needs to be considered while giving loan.
- Loan amount(loan_amnt) has very strong positive correlation with total_rec_prncp , total_rec_int , last_pymnt_amnt which is a great indicator that the CFC is great at collection management reducing structural risk.
- Total_pymnt , total_pymnt_inv , total_rec_prncp , total_rec_int has strong positive correlation as they are leading and lagging indicator of efficient collection system.
- Int_rate has strong positive correlation with revol_util which may be an indicator that the CFC is catering to a segment which is loan starved from other sources or their interest rate is still substantially higher.
- total_acc has strong positive correlation with open_acc which is an indicator that most of the credit files are open, which is an indicator of strong loan book.
- Int_rate has strong positive correlation with total_rec_int which is an obvious conclusion since higher interest rate leads to higher interest recovery considering its not a bad loan.

Conclusions

- **Loan Purpose:** Debt consolidation and credit card payoff are the most common purposes, indicating a prevalent use of these loans for managing existing debt.

- **Seasonal Trends:** There is a clear seasonal pattern in loan issuance, peaking in December, which may reflect borrowers financial planning trends at year-end.

- **Income Verification and Default Rates:** The status of income verification does not consistently correlate with lower default rates, suggesting that it should not be overly relied upon, especially for larger loans.

- **Account Number Risks:** The number of open accounts presents a U-shaped risk profile, where very low and very high counts are associated with increased default risks, highlighting the complexity of financial management among borrowers with extreme numbers of credit lines.

Loans with below criteria are highly contributing to charged off:

- Loans with higher interest rate (>12%)
- Loans with grade 'F' and loan amount > 20K
- Higher amount loans (>13K) for small business, debt consolidation or credit card
- Borrower Debit to income ratio > 25% and loan amount > 15K
- Borrower annual income <50K and loan amount > 5K

Recommendations

- **Implement Risk-Based Pricing:** Develop a nuanced interest rate model that reflects diverse risk factors including DTI, credit utilization, and loan purpose, ensuring rates are commensurate with potential risks.
- **Geographic Risk Management:** Establish tailored lending criteria for different regions, especially in states with historically higher default rates, to mitigate location-based risks.
- **Specialized Debt Consolidation Programs:** Offer tailored financial products for debt consolidation with accompanying advisory services to help borrowers manage their debts more effectively.
- **Enhance Employment Length Evaluation:** Balance the emphasis on employment length with comprehensive assessments of financial health to avoid over-prioritizing tenure over actual ability to repay.
- **Credit Utilization Strategy:** Tighten approval criteria for borrowers with high credit utilization rates (>60%), recognizing this as a significant predictor of potential default.
- **Seasonal Lending Adjustments:** Plan for seasonal fluctuations in loan demand, optimizing capital allocation and marketing strategies to match the observed end-of-year surge in borrowing.
- **Stricter DTI Thresholds:** Enforce more rigorous reviews for loans where the DTI exceeds 20-25%, identifying this range as a critical risk threshold.
- **Robust Income Verification:** Strengthen the income verification process, especially for larger loan amounts, to ensure that reported incomes are accurate and reliable.
- **Adjust Loan Terms Based on Amount:** Tailor the terms of loans, particularly the maximum amount and duration, to better manage the risk profile of longer-term, larger loans.

Acknowledgement

Thanks to the instructors from UpGrad and IIITB for guidance and feedback. Acknowledgment

Group Facilitator :
Name: Subrata Das
Email ID: subrata.pucsd@gmail.com
Phone No:918861506628

Team Member Detail:
Name: Suhas Naik
Email ID: suhasnaikk@gmail.com
Phone no: 918722476171