

Leanding Club Case Study

ML63 Batch

Sujata and Shameer

Problem Statement

Lending Club is a marketplace for personal loans. That matches borrowers who are seeking a loan with investors looking to lend money and make a return.

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

The objective of the case study is determine how to reduce the risk in loan approval decision making.

- By identifying the markers of applicant who can repay the loan
- By identifying the applicants markers who can default the loan

Data understanding

The given Dataset had - 39717 Rows & 116 Columns.

- This dataset contains information about applicant profile(backgrounds/current situation) and Lending Club loan information.
- We have 5 years of data 2007 to 2011.
- Dataset contains mix of Categorical and Numerical - Loan Data.
- Dataset contains missing / NaN values records.
- Dataset contains additional columns after Loan approval.
- Dataset contains some single value columns.

Data Cleaning

Dataset is cleaned up to ensure the relevant columns are taken for analysis and also ensured the data completeness as well. Below processing was done for cleanup:

- Removed the 100% missing values 54 Missing columns dropped from Analysis
- Removed the single value 9 columns from Analysis
- Removed the 3 Columns missing at least 50% data
- Removed the 3 columns where all the values in column are unique.
- Removed 1 column where all column values are duplicate.
- Removed 4 columns with all the data is text where too much description given.
- Removed 18 columns data which is captured after loan approval.
- Total 92 columns removed from 111, we are analysis on 19 columns.

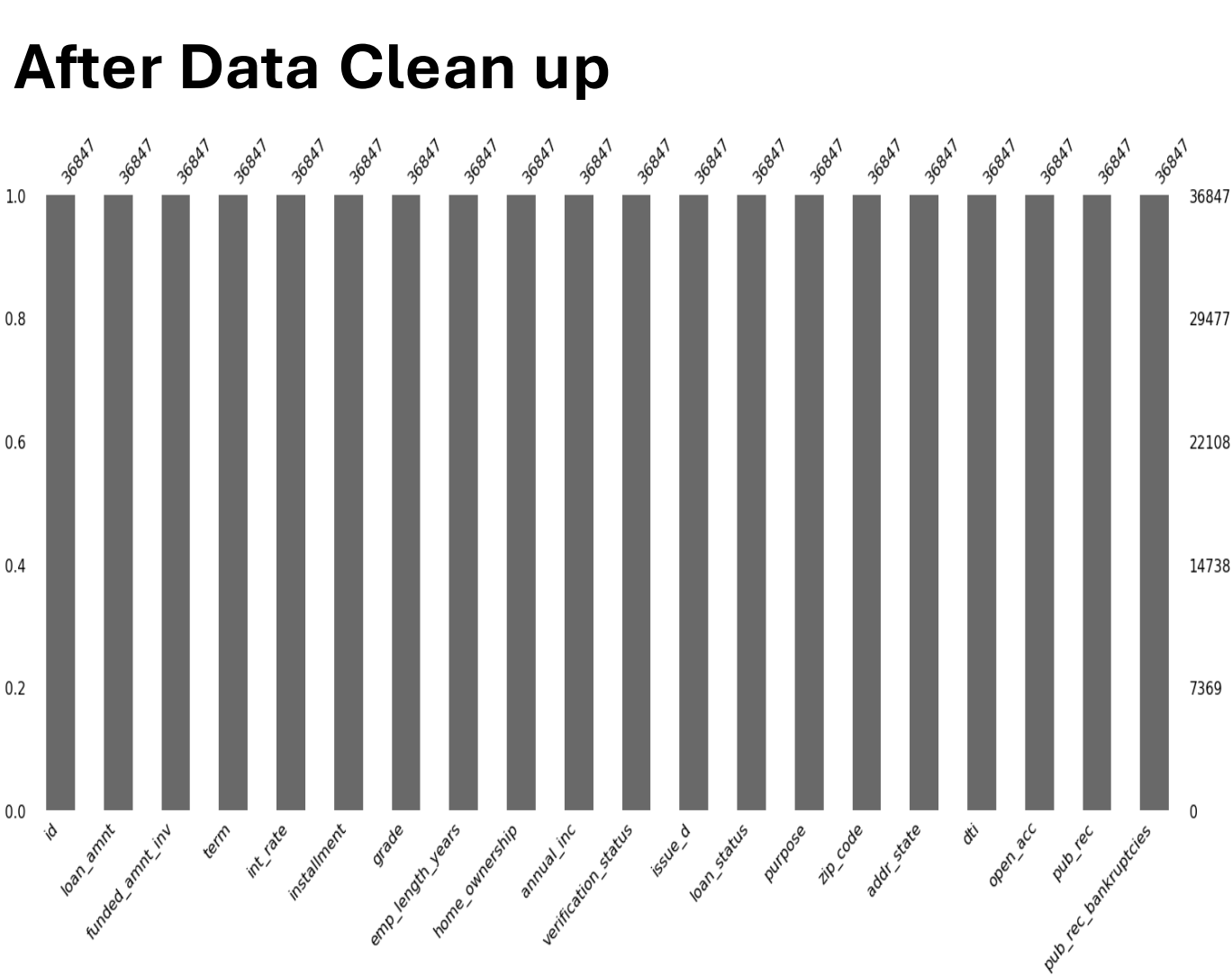
Data Manipulation

For Analysis aid below are few change performed.

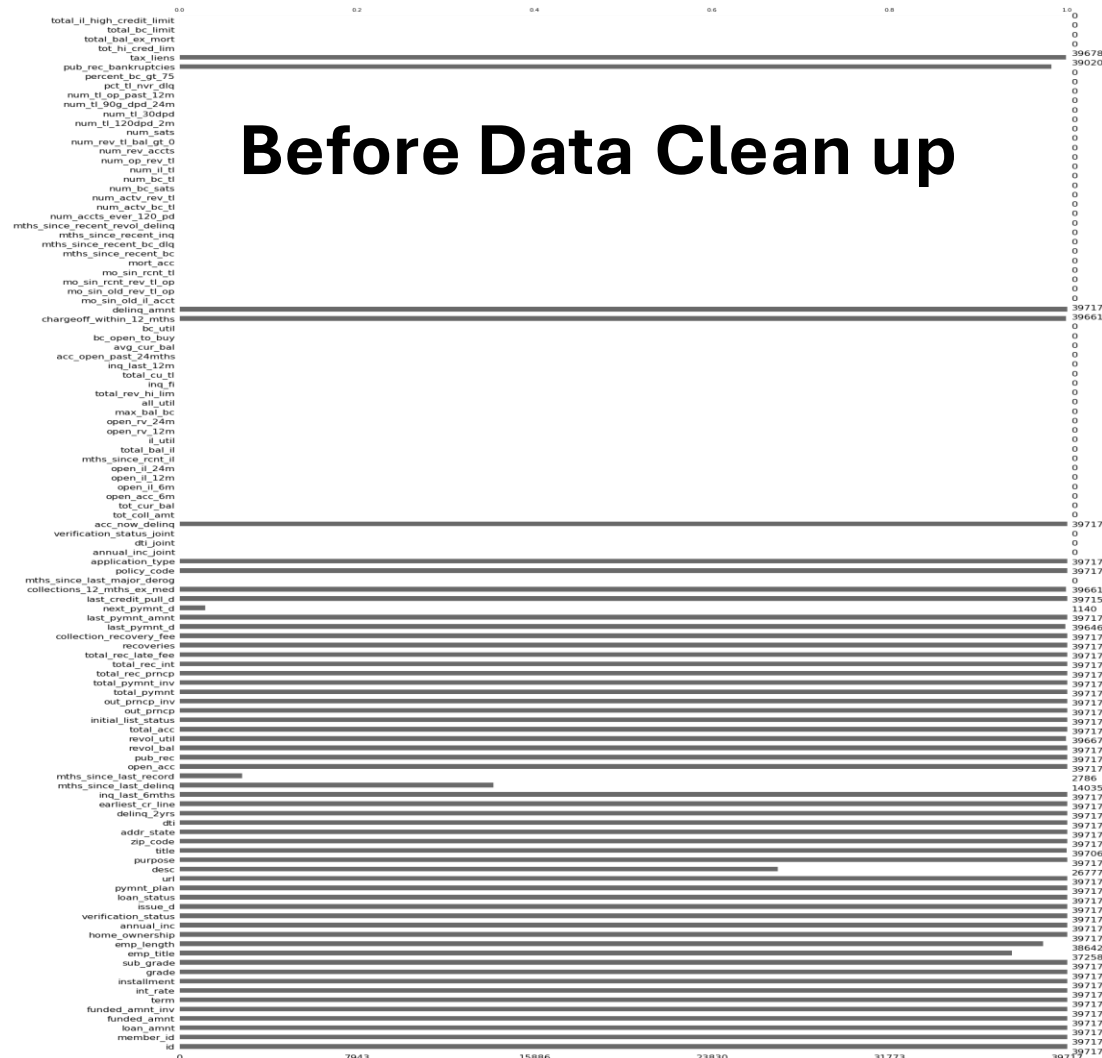
- Type conversion done for the relevant columns to perform aggregations and Math operations, few converted to Numeric data int/float.
 - Term, Interest rate.
- Created new numeric column based on the text value
 - New loan status – int, new column added
- Converting Date columns
 - Issue date column to Date time column type
- Updated Employee Length column to contain only Year, Removed additional text and total perc of null value rows in Employee length very low perc so deleted this rows.
- Dropped current status from load status as it is ongoing loan so not useful for current analysis

Data Analysis – Missing Values Check

After Data Clean up



Before Data Clean up



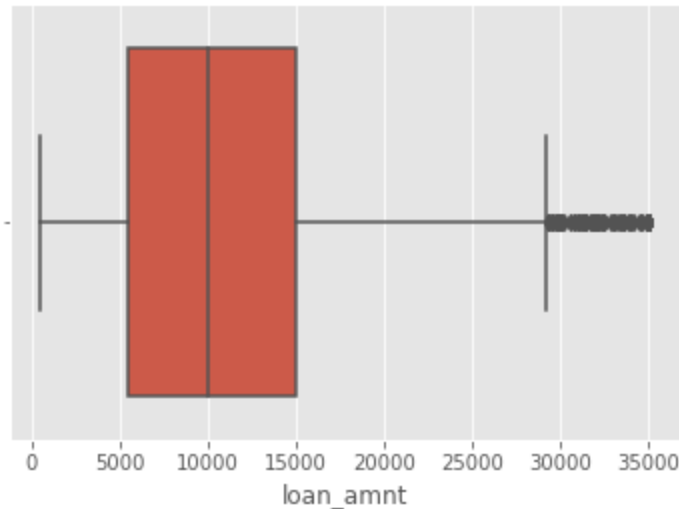
Data Analysis – Pre check

Below set of columns were considered for Data Analysis

- **11 Numeric**
- **6 Text**
- **1 Date**
- **1 Additional**

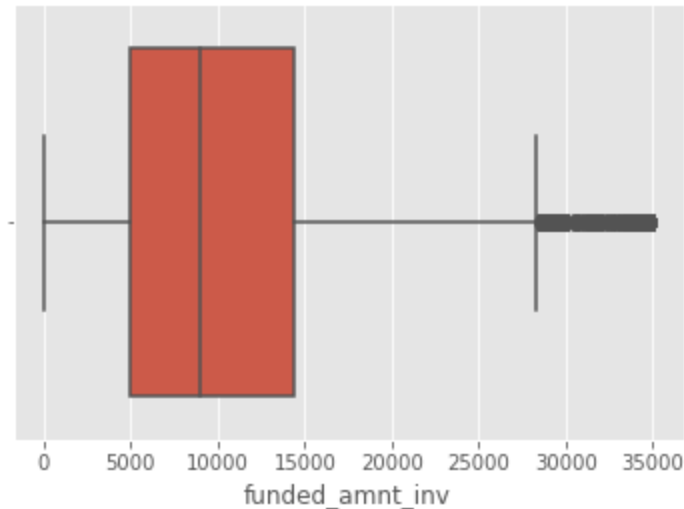
Leanding Club – Data Analysis

Loan Amount Distribution



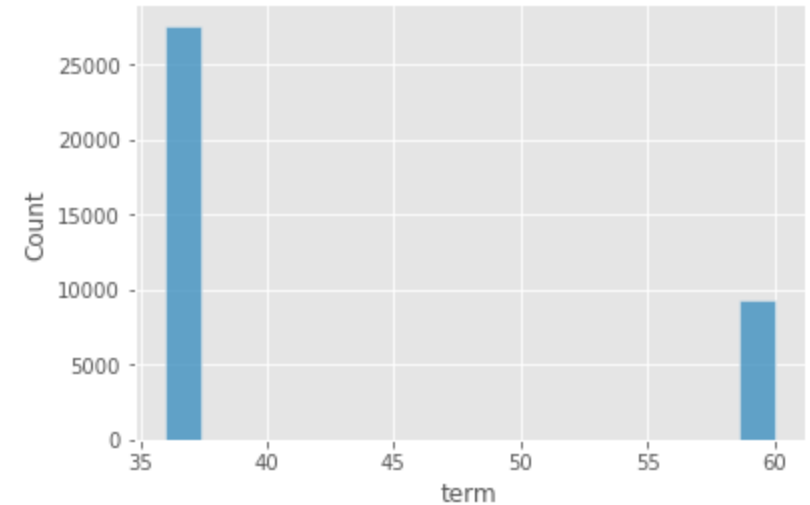
Max loan amount is 35000, however the 75% is only at 15000.

total amount committed by investors



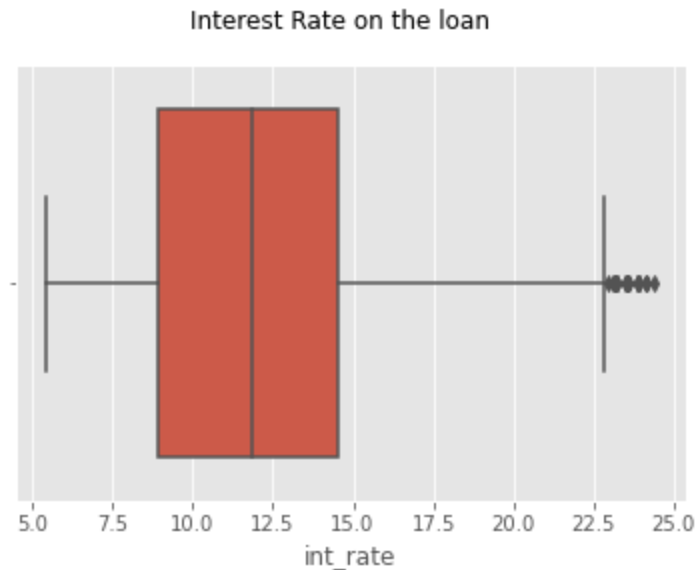
The total amount committed by investors – 93.67%.

number of payments on the loan

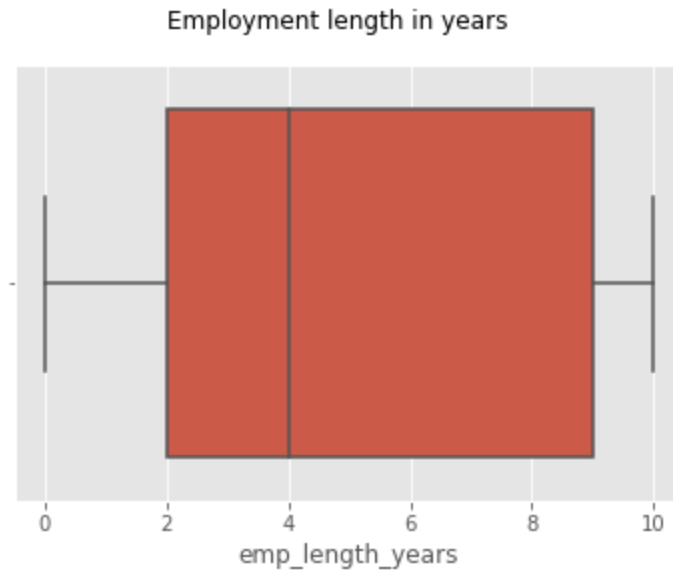


Majority of the loans are 36 Month term

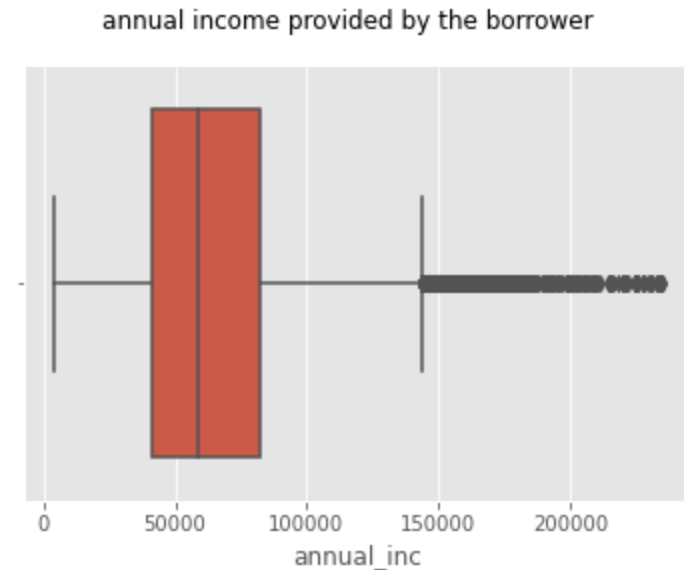
Leanding Club – Data Analysis



Max interest rate is 24%,
however the rate between – 9%
to 15%



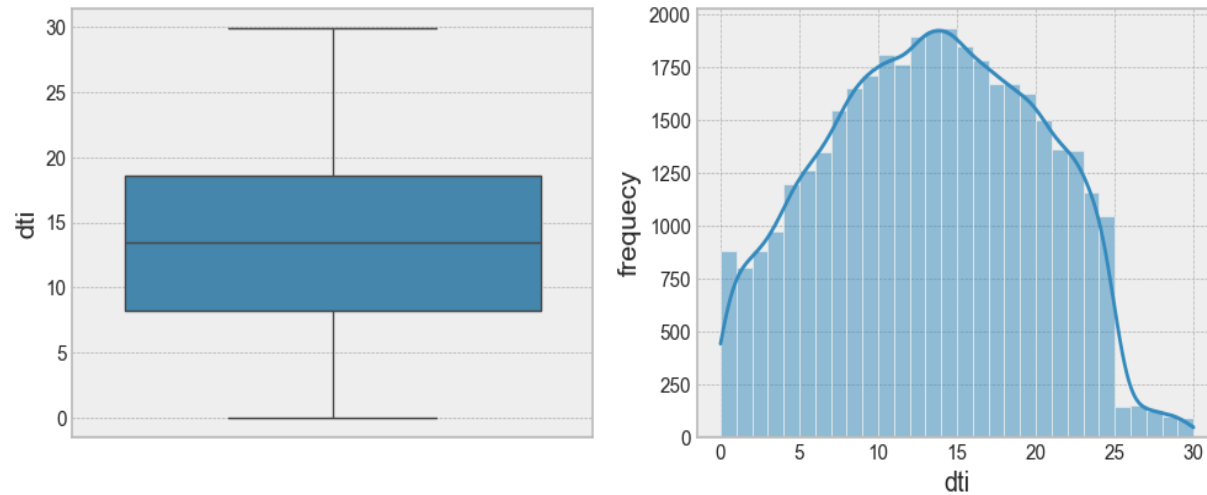
50% has employment length
has 4 years. Max is 10 years.



Most of the Annual incomes are
in range of 40000 to 81600

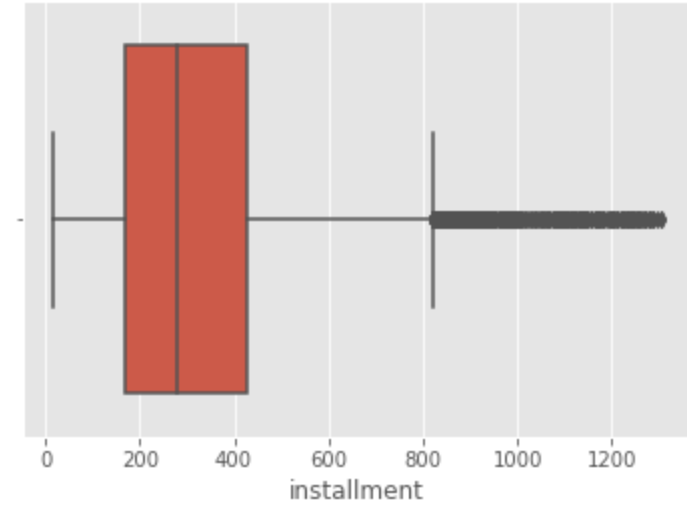
Leanding Club – Data Analysis

Ratio calculated using the borrower's total monthly debt payments on the total debt obligations



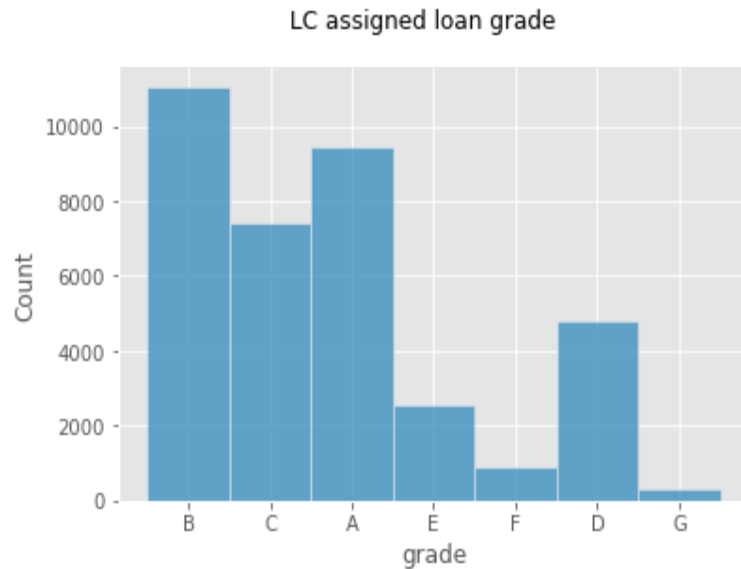
Maximum dti of debt obligations is 30%. However the mean is 13%.

Installment distribution

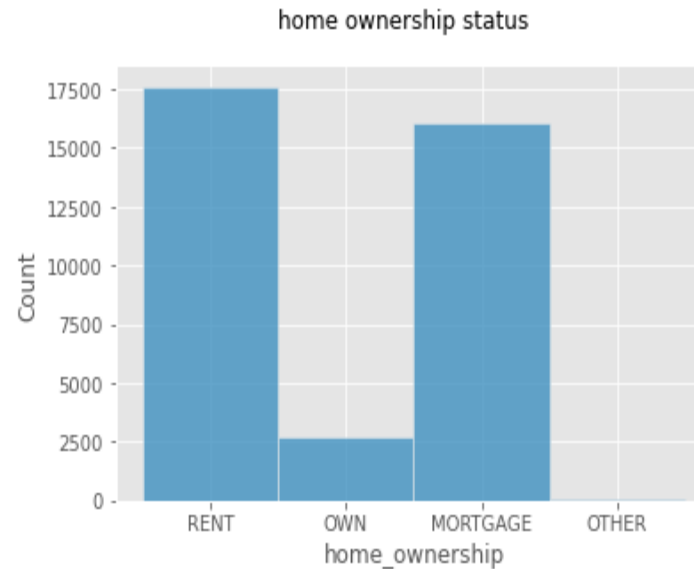


Most of the monthly payments is between 150 to 400.

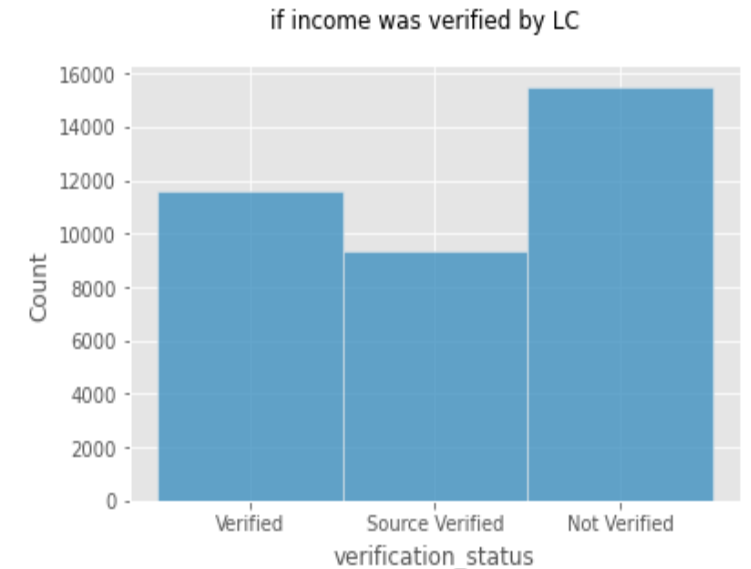
Leanding Club – Data Analysis



Majority of the loan from Grade A, B & C

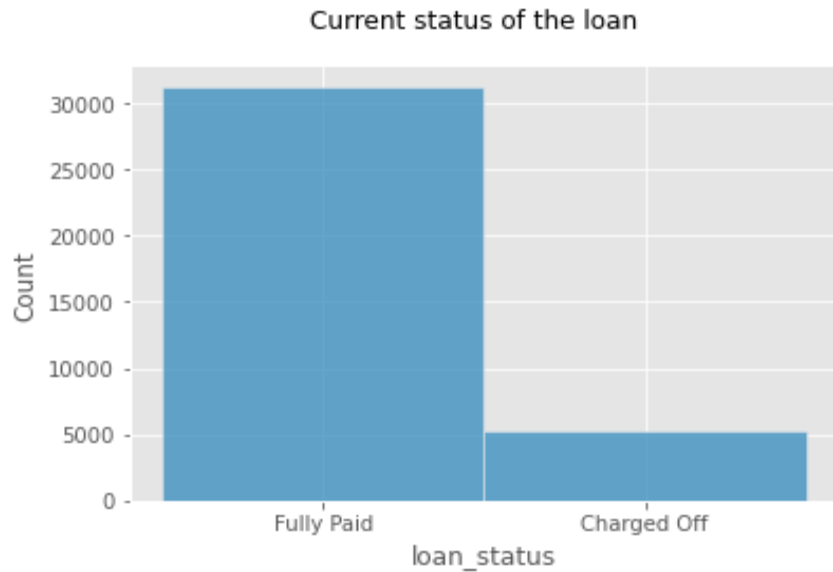


Most people took the loan have rented Home or Mortgage

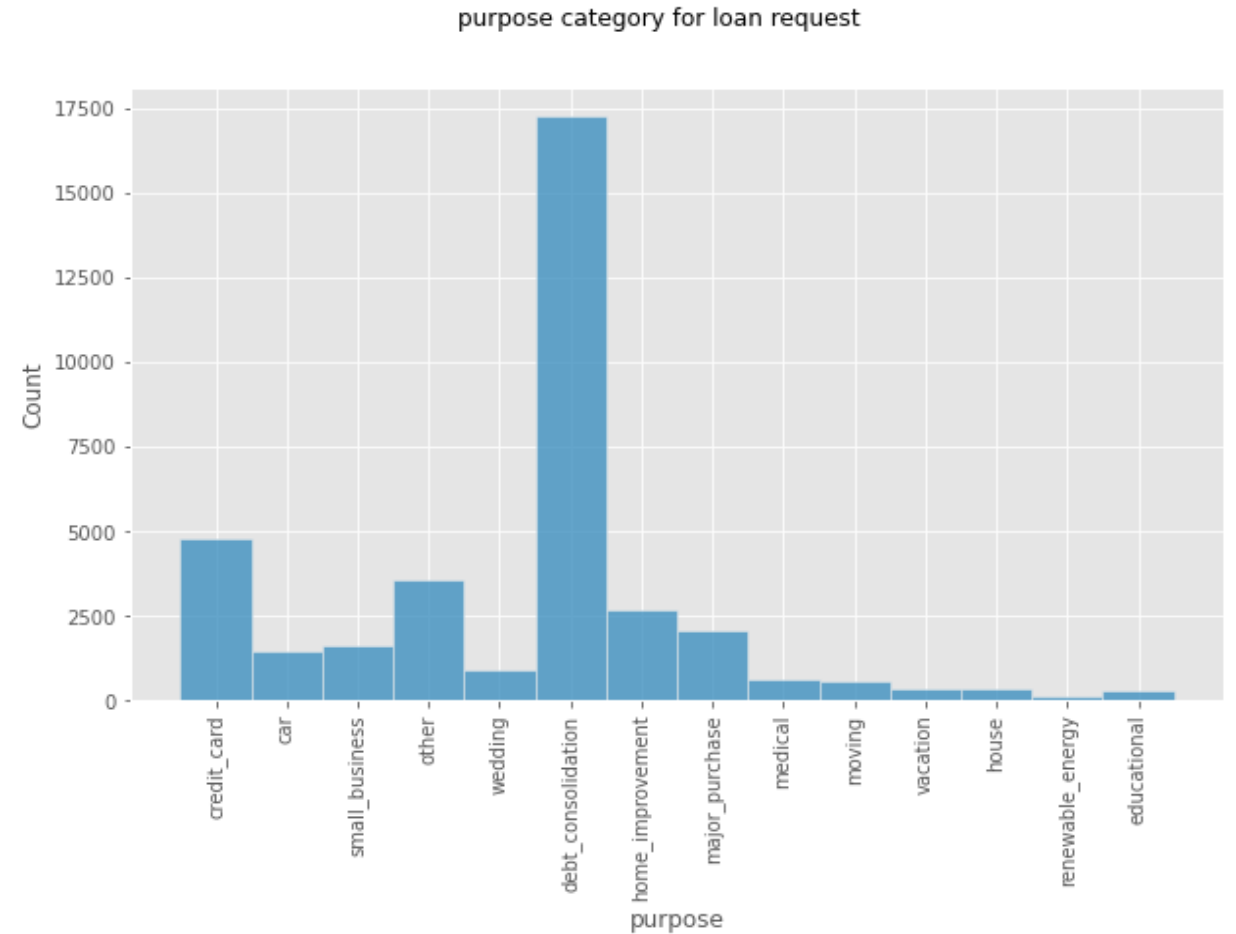


Large amount of people verification status is not verified

Leanding Club – Data Analysis



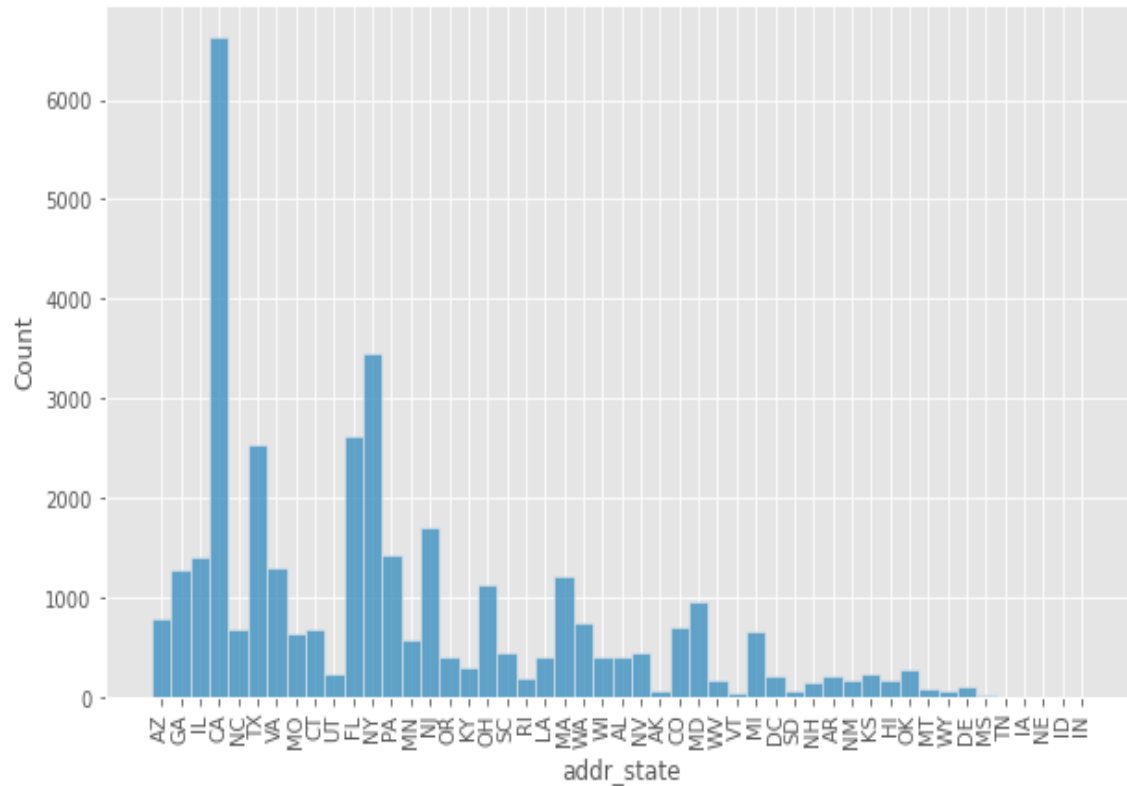
85.38% people paid the loan and
14.61% people are defaulters.



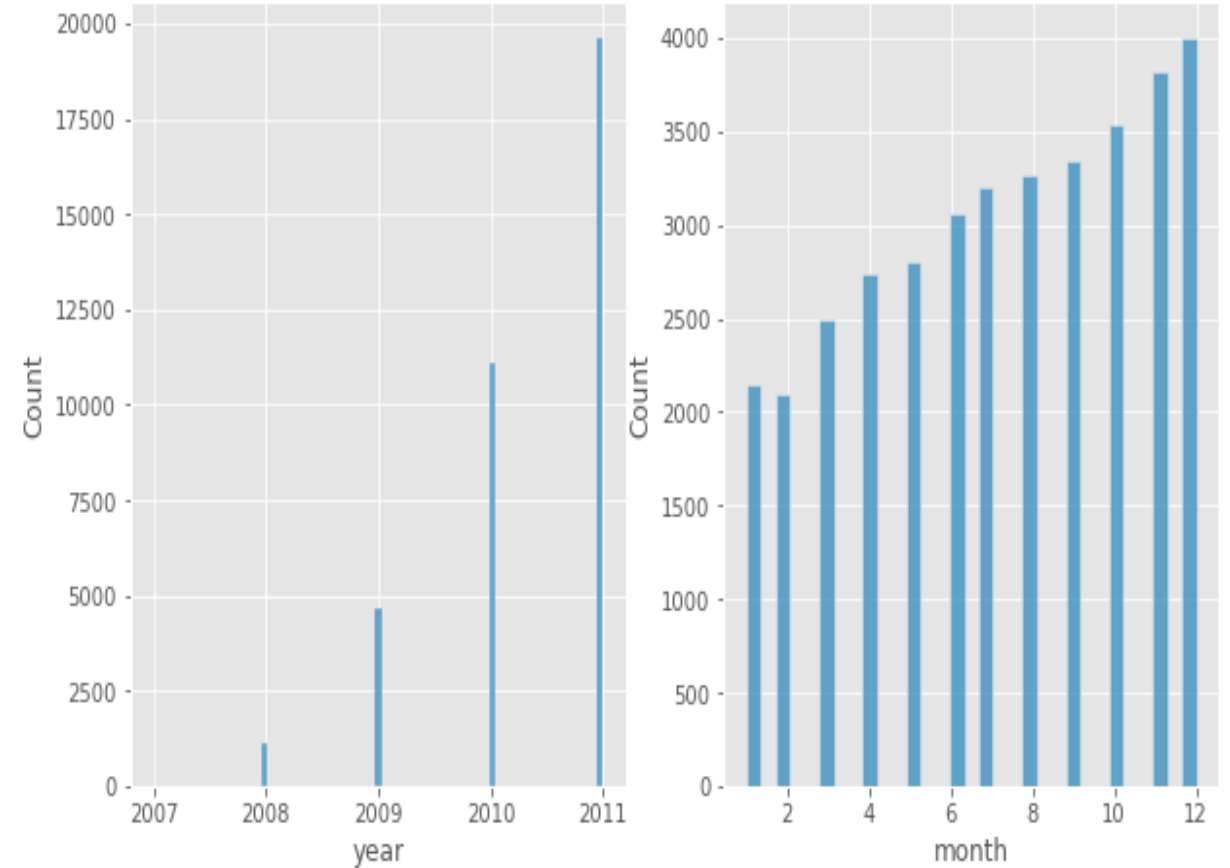
Most people took the loan have rented Home or Mortgage

Leanding Club – Data Analysis

state distribution

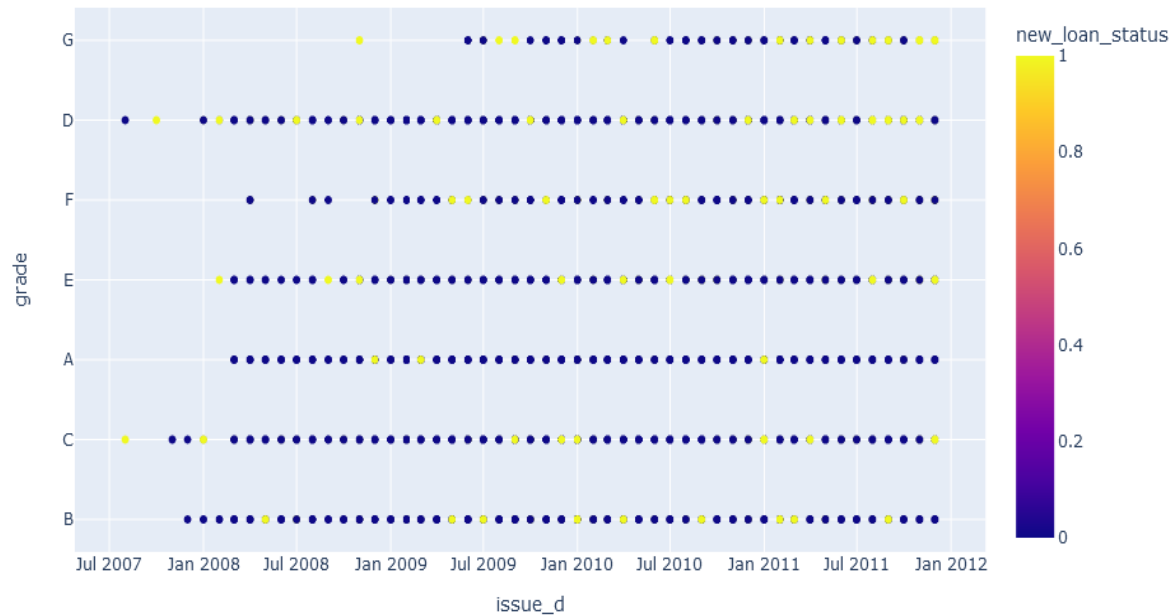


Most of the loan applicant from CA and NY state

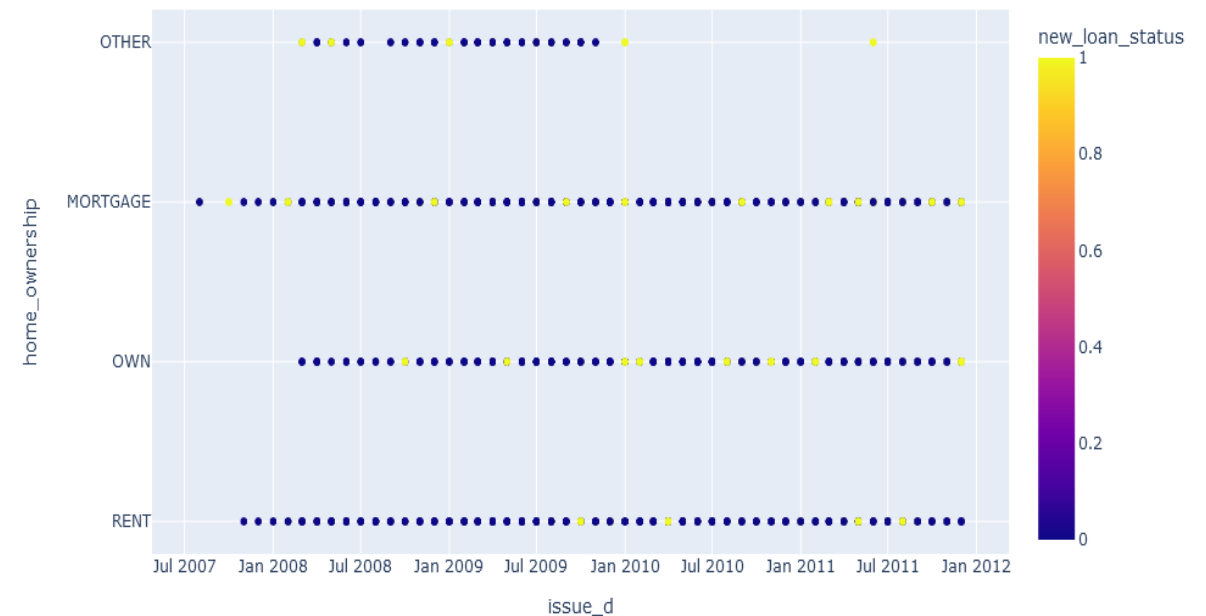


Maximum loan taken on year 2011 and it is continuously increasing

Leanding Club – Data Analysis



It is observed Grade D – Defaults are increasing in recent year.

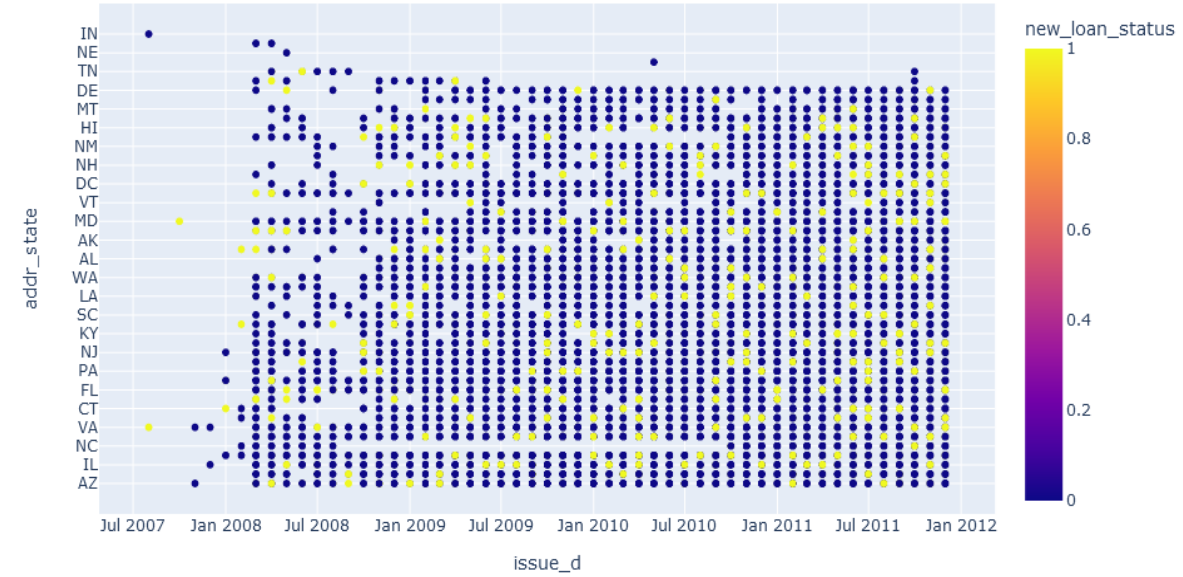


Early years till End of Jan 2009, Rented Home ownership did not have any defaults.

Leanding Club – Data Analysis

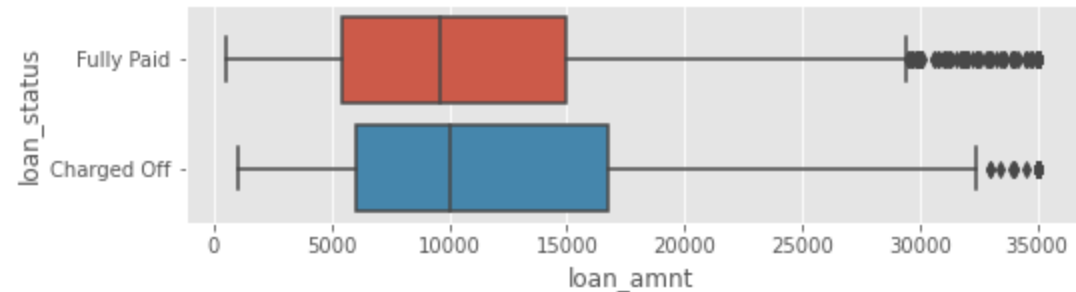


Purpose – Credit Card, Wedding and Moving seen consistent full payments.

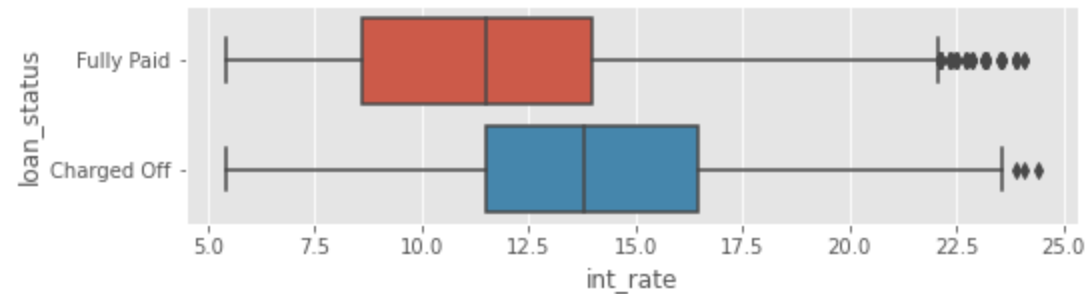


Defaults are spread across states over the years.

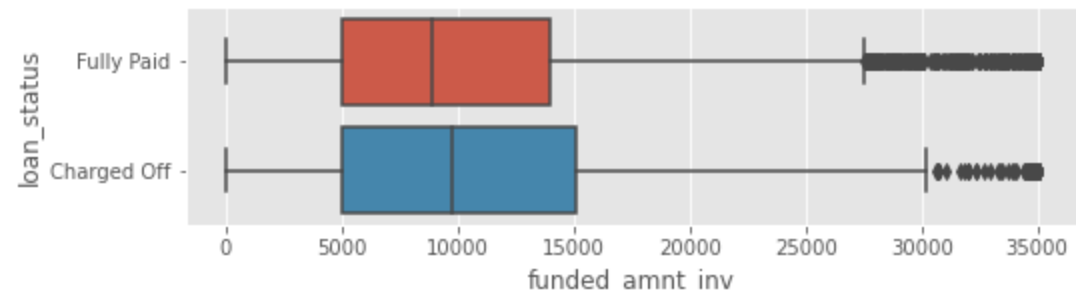
Lending Club – Data Analysis



1. Loan amount is more than 10000 can be defaulter
2. Amount less than 6000 has very less chances to be defaulter
3. Higher the amount cause more defaulter.



1. Interest rate more than 13.79% can be defaulters.
2. Interest rate less than 11.49% has very less chances to be defaulter.
3. High interest rate cause more defaulters.

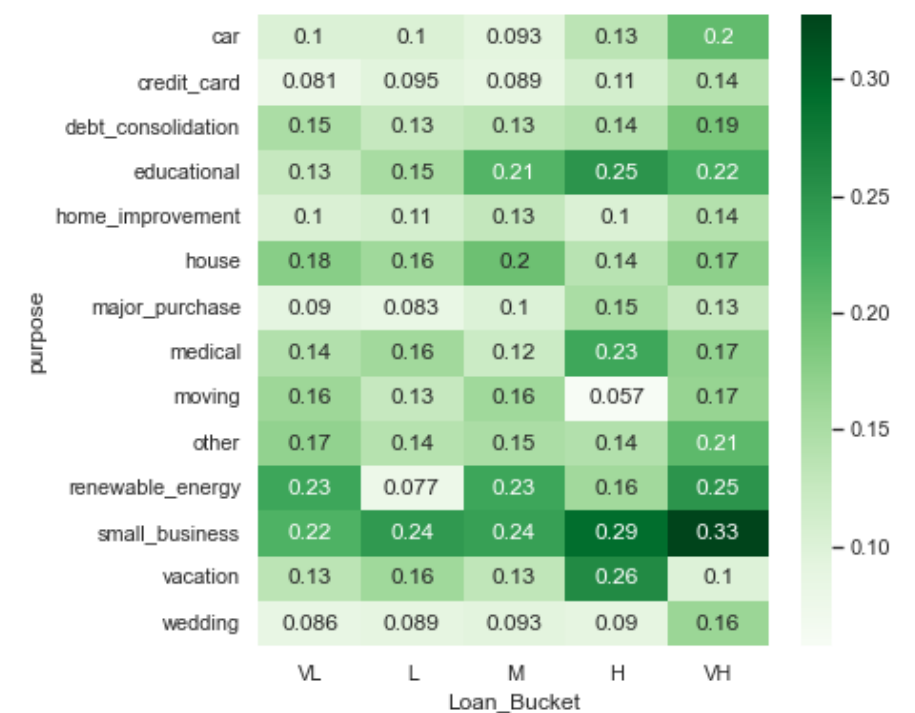
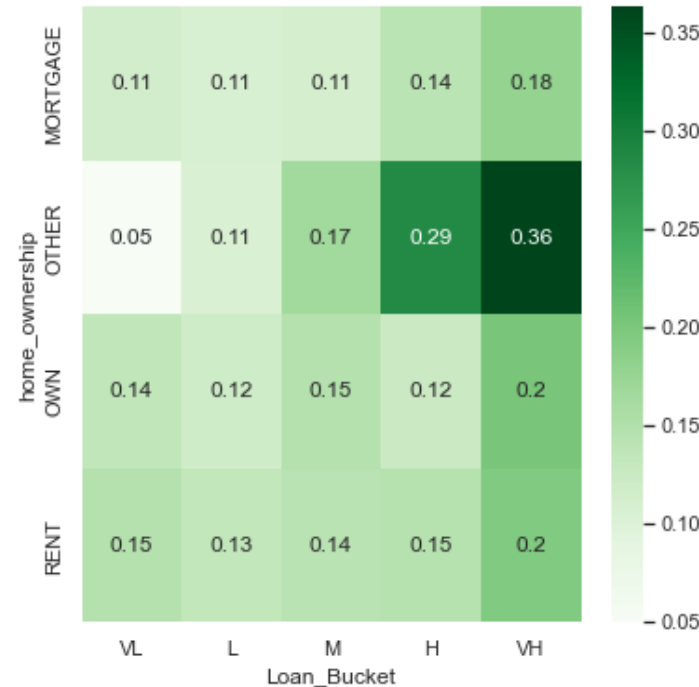
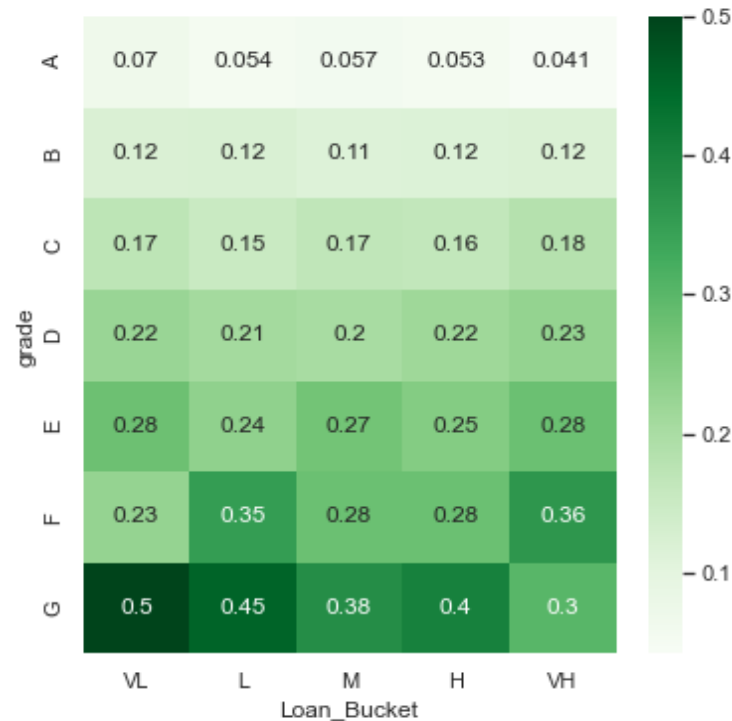


1. Amount committed by investors is more than 9725 can be defaulter
2. Less than 5000 has very less chances to be defaulter

Leanding Club – Data Analysis

Loan Amount Buckets/bins -

1. Very Low amount - VL -0% to 20%
2. Low - L - 20% to 40%
3. Medium - M - 40% to 60%
4. High - H - 60% to 80%
5. Very High - VH 80% to 100%



1. Lower grades E,F,G defaults are increasing with loan amount
2. For grade G 20-40 bucket has more defaults
3. Grade F has more defaults with Very High loan amounts.

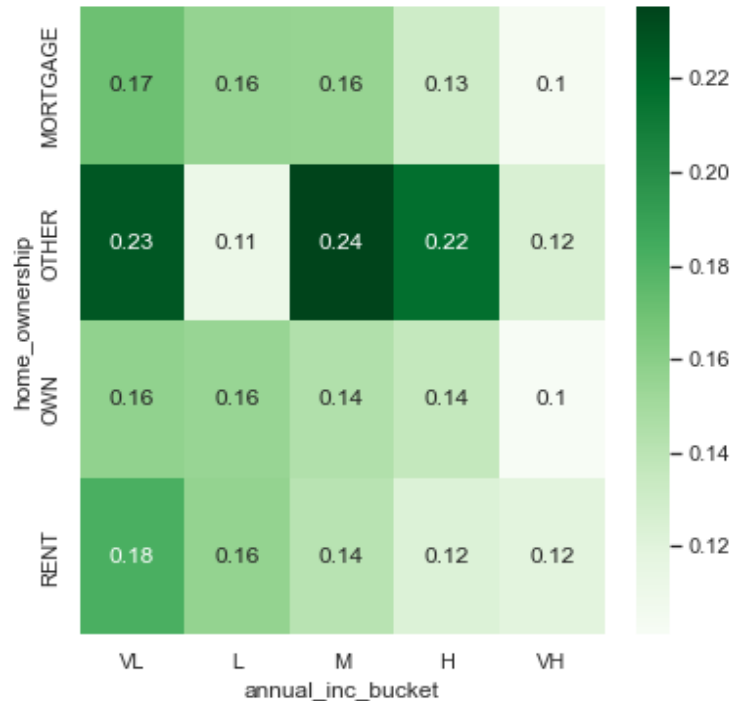
"Other" category Lower Loan amounts - less defaults. however as the loan amount increase the default as well increase

Small business having high defaults overall and defaults increase as the loan amount increase

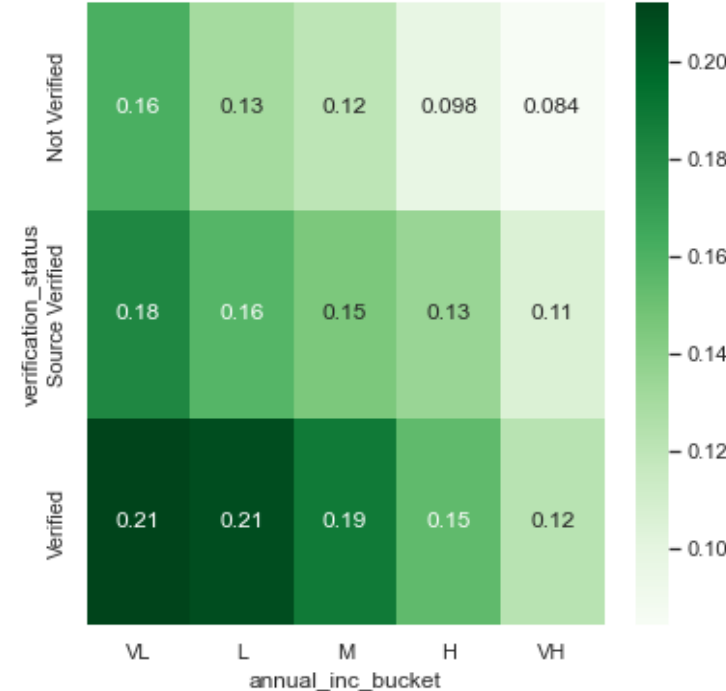
Leanding Club – Data Analysis

Annual Income Buckets/bins -

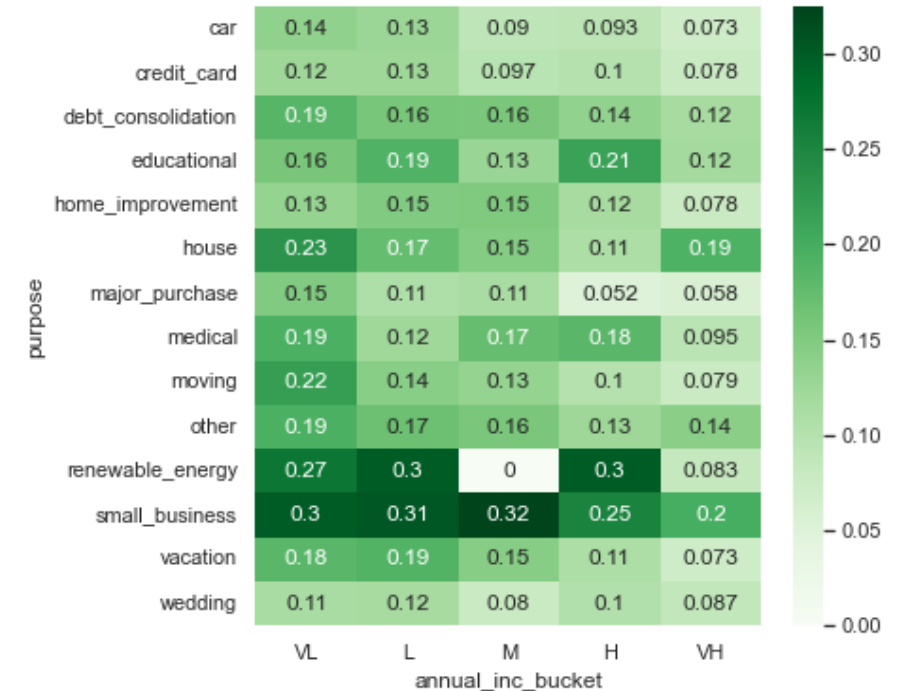
1. Very Low - VL -0% to 20%
2. Low - L - 20% to 40%
3. Medium - M - 40% to 60%
4. High - H - 60% to 80%
5. Very High - VH 80% to 100%



Home ownership type - Mortgage has more default compare to OWN or RENT

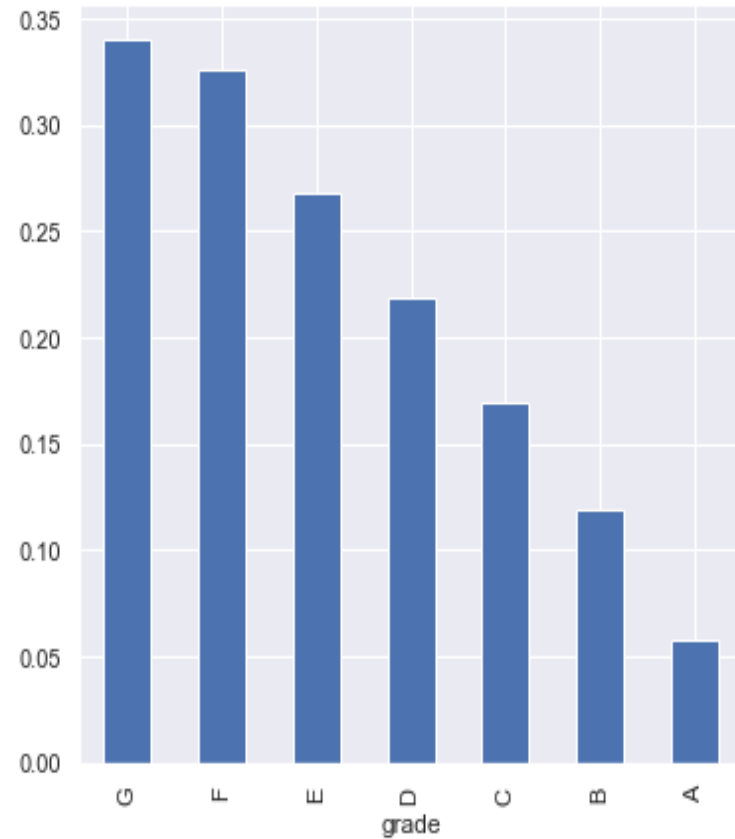


Very low and Low income buckets has more defaults irrespective of the verification status

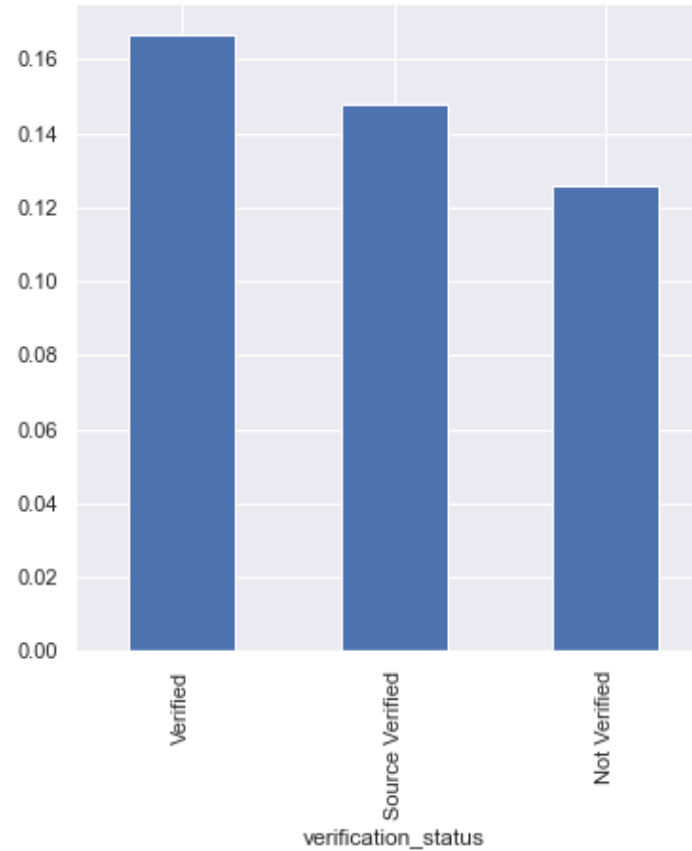


1. Small Business with Medium annual income bucket has more defaults compare to all.
2. Loan purpose - house has more default even if the annual income is very high

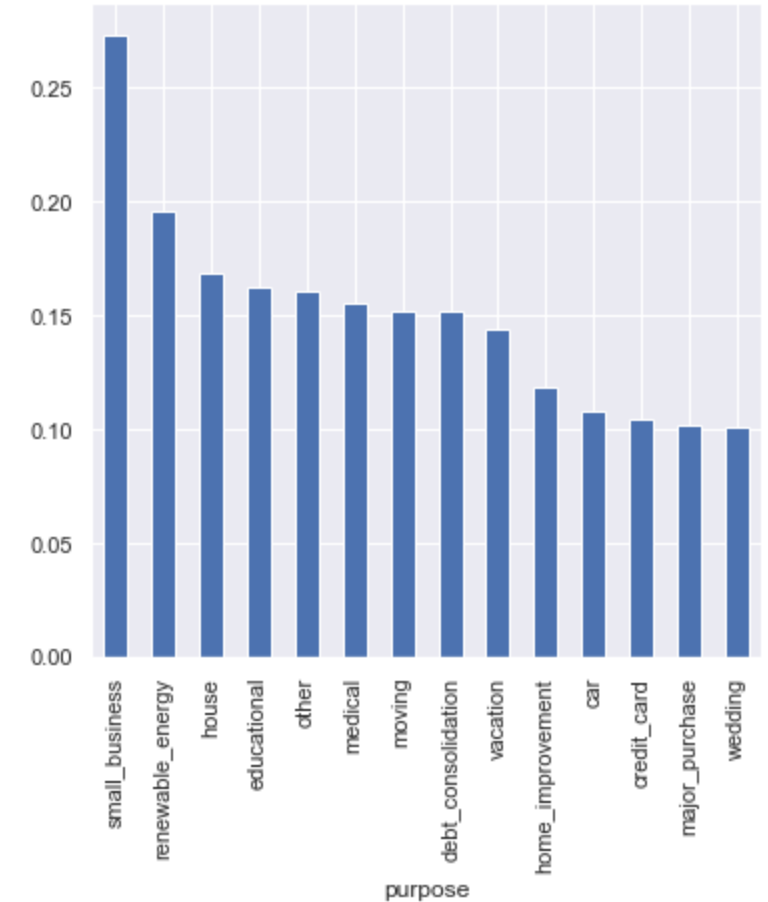
Leanding Club – Data Analysis



Defaults are increasing with the Grade.



Verified loans having more Defaults

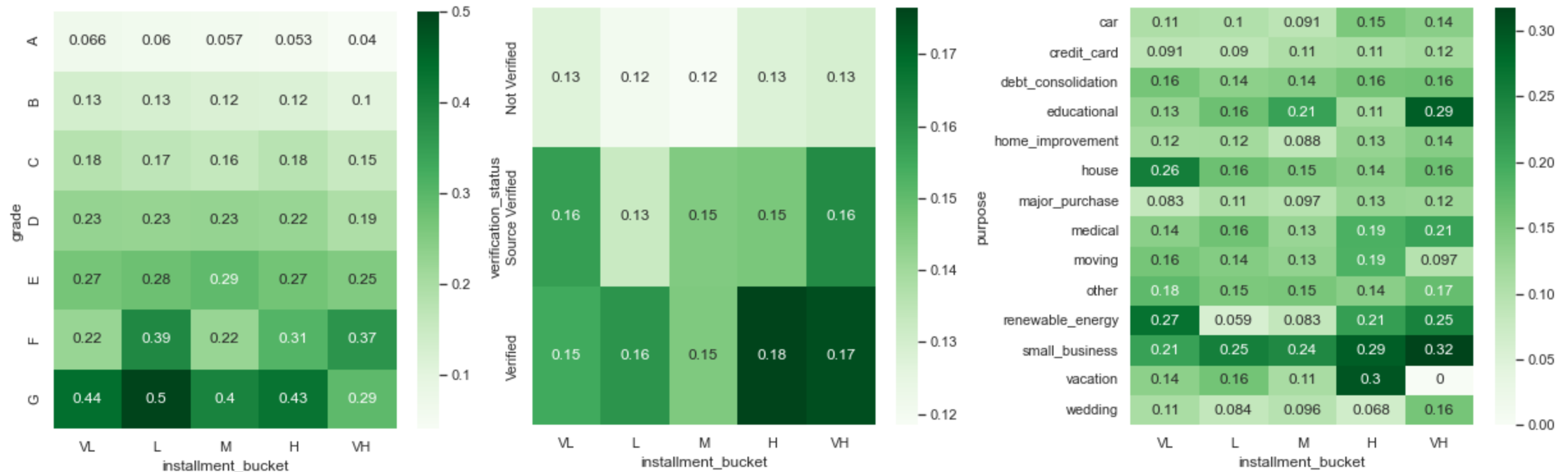


Small business and Renewable energy has more defaulters.

Leanding Club – Data Analysis

Installment Buckets/bins -

1. Very Low - VL -0% to 20%
2. Low - L - 20% to 40%
3. Medium - M - 40% to 60%
4. High - H - 60% to 80%
5. Very High - VH 80% to 100%



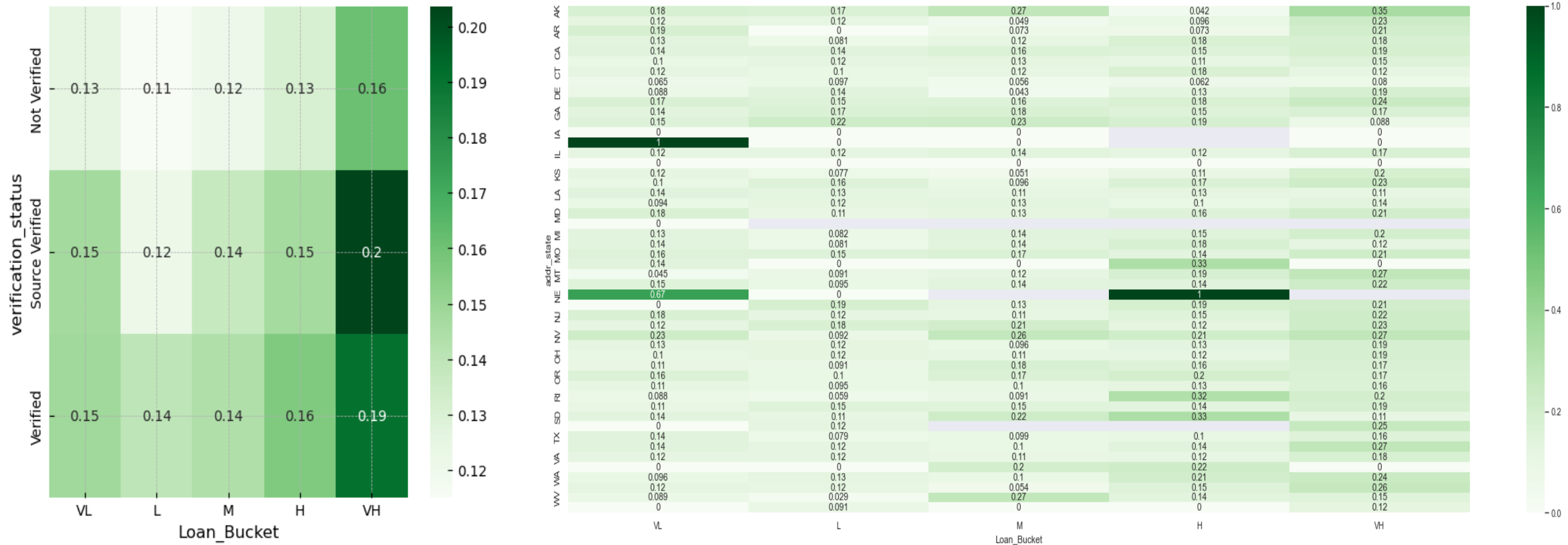
Low installments tend to have more defaults. mainly seen in Grade G and F

Higher installments with the verified source has more defaults

1. Vacations with higher installments has defaults.
2. small business and educational with higher installment bucket has more defaults
3. House and renewable energy with very low installment bucket has more defaults.

Leanding Club – Data Analysis

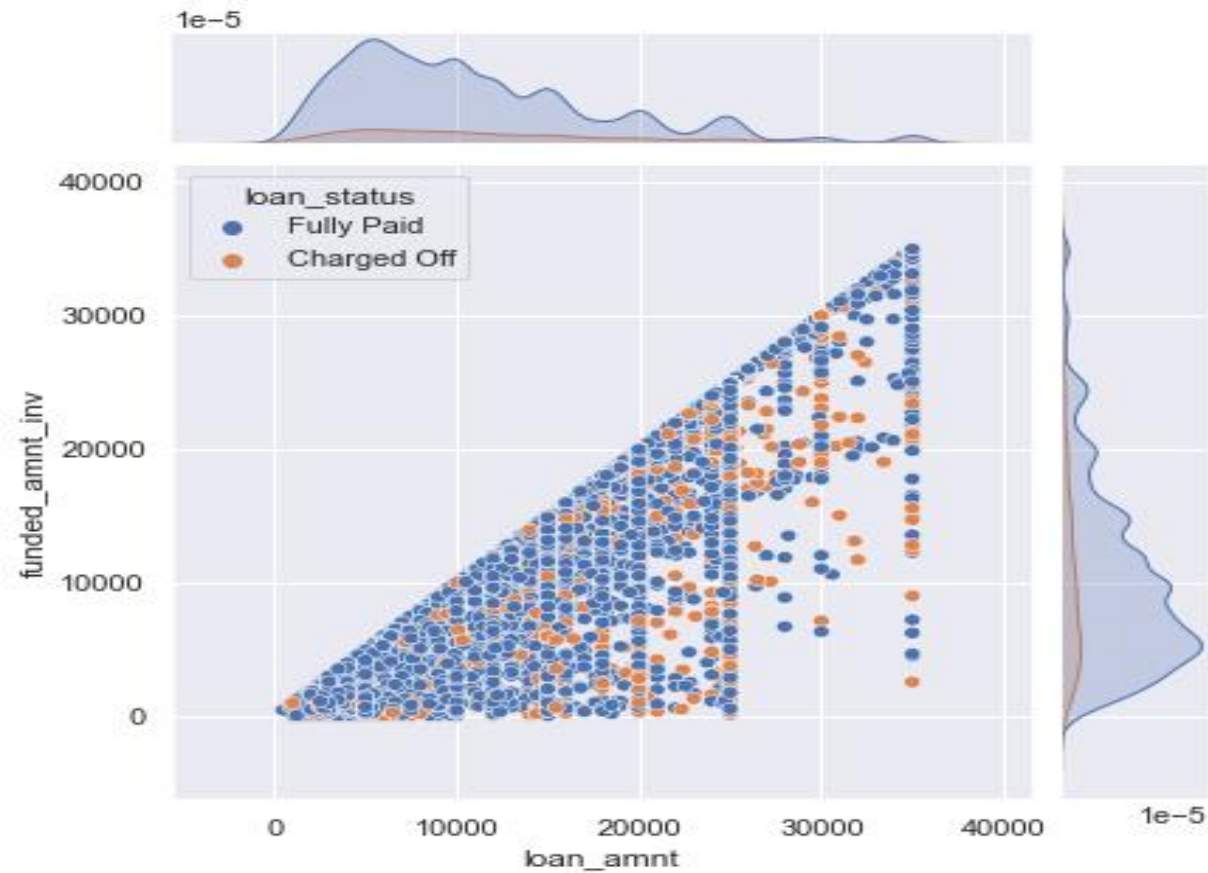
- Loan Amount Buckets/bins -**
1. Very Low amount - VL -0% to 20%
 2. Low - L - 20% to 40%
 3. Medium - M - 40% to 60%
 4. High - H - 60% to 80%
 5. Very High - VH 80% to 100%



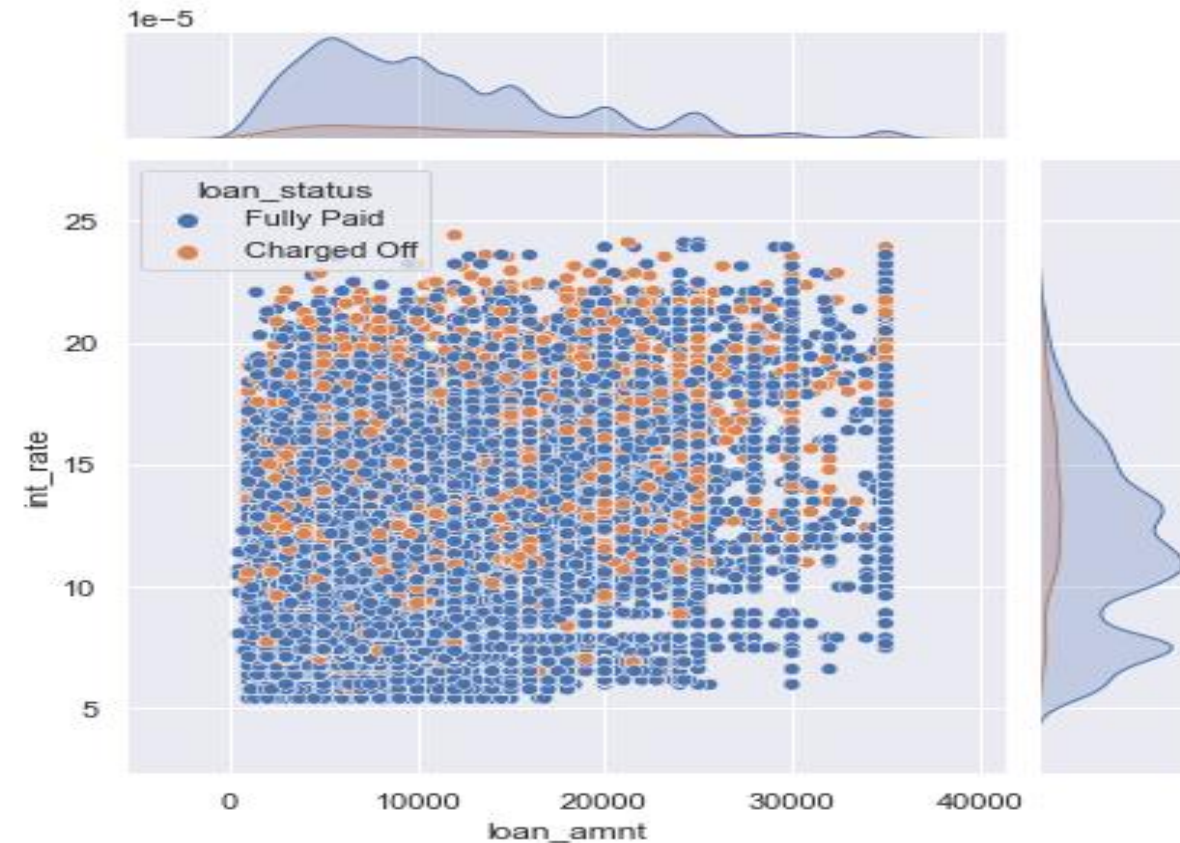
Verified sources with higher loan amounts has high defaults

NE – State seen more defaults with Very low or High loan amounts as well.

Lending Club – Data Analysis

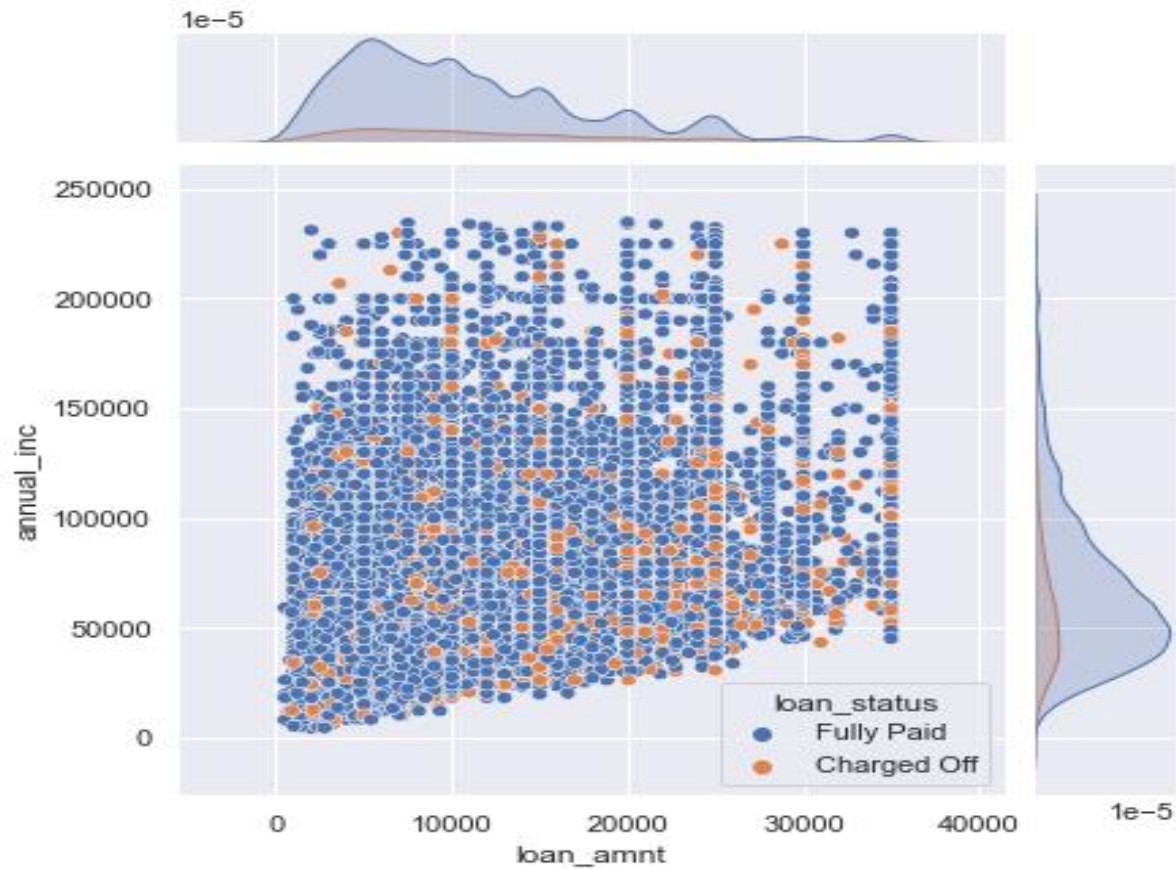


Defaults are more scatter. However the fully paid loans are more densely packed at lower amounts.

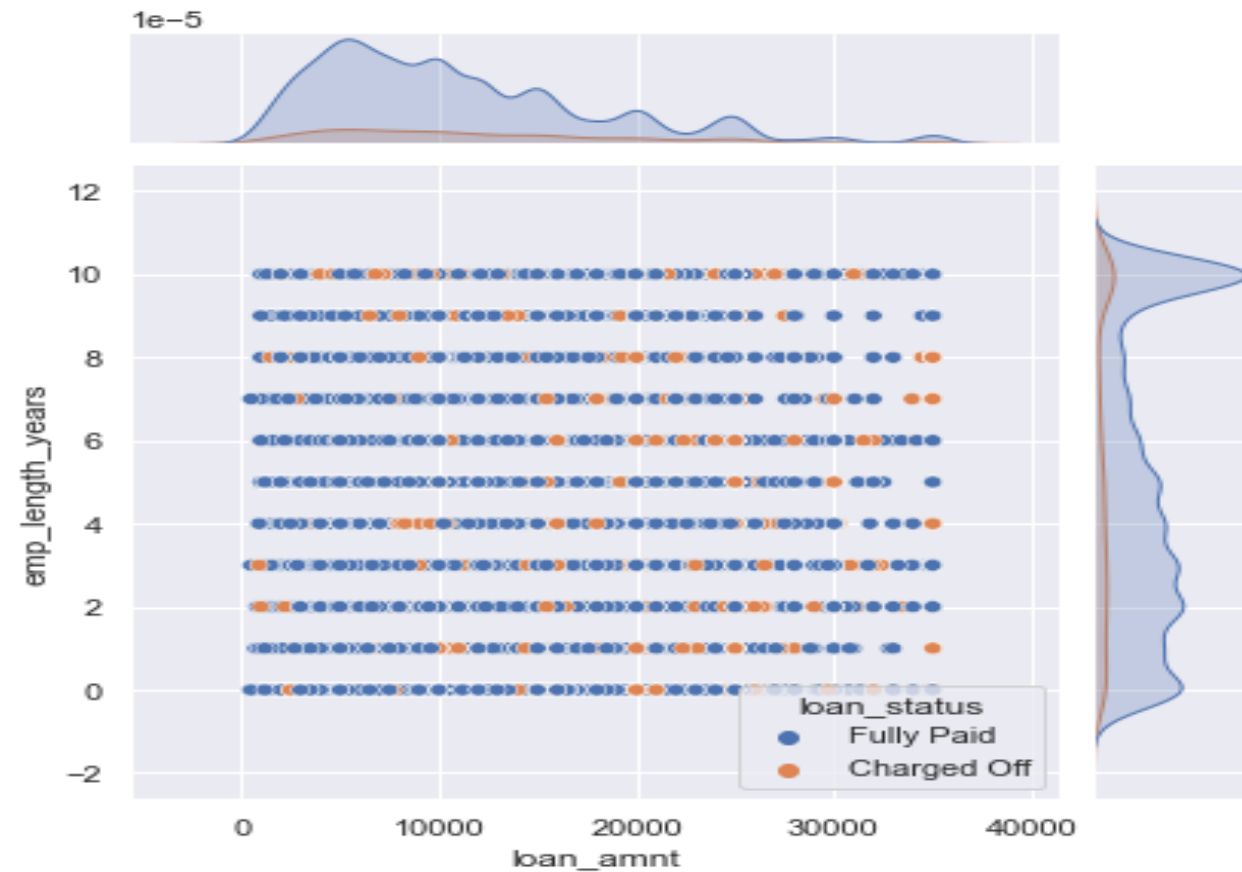


Defaults are more scatter with interest rate and loan amount. However with less loan amount and lesser interest rate, repayment is better.

Lending Club – Data Analysis



Loan amount and Annual income is strongly correlated. And Loan amount is skewed to left, as more low loan amount.



Employment length , Loan amount is fully scatter with Fully paid and Charged off loans.

Conclusions

These points highlight the factors associated with increased default risk, helping lenders make more informed decisions when evaluating loan applications.

- Higher the amount cause more defaulter, more than 10,000 have higher chances to be default and amount less than 6000 has very less chances to be defaulter also Funded amount higher than 9,725 associate with high default rates.
- Loan term with 60 have greater chances to be default than term with 36.
- High interest rated cause more defaulters, Interest rates more than 13.79% can be defaulters and interest rate less than 11.49% has very less chances to be default.
- Lower loan grades E, F, G have higher default rates, defaults increasing as the loan amount rises within these grades.
- Verified applicants with higher loan amounts tend to default more frequently also Loans for verified applicants with high installments show increased default rates.
- Loans taken for small businesses, renewable energy, and educational purposes have higher default rates especially when loan amounts increase, also Loans for vacations with higher installments also have a higher Possibility for default.
- Applicant's from NV and TN state have higher default rates.
- Ownership with mortgages tend to default more often compared to those who own or rent their homes.
- Lower installments, particularly in grades F and G, are associated with higher default rates also Higher installments in loans for small businesses, education, and renewable energy also correlate with higher default rates.
- A higher DTI, especially up to 30%, indicates a higher default.
- Regardless of verification status, lower-income buckets show more defaults, indicating income level itself is a critical factor.
- No clear pattern is observed for employment length, suggesting it's not a strong indicator of default.
- Mostly do not have significant public records of bankruptcies, but those who do have a slightly increased risk of default.

Recommendation

- **Limit loan amount :**

- Implement eligibility criteria for loans exceeding 10,000, as higher amounts are linked to increased default rates and Encourage applicant to select 36-month terms for higher loan amounts.
- Establish loan amount Criteria based on annual income levels to prevent applicant from taking extra loan that can cause their repay capacity.
- Lesser Loan amounts tend to be repaid more often than higher loan amounts.

- **Purpose Specific Criteria :**

- Small business and educational with higher instalment bucket has more defaults and Small Business with Medium annual income bucket has more defaults compare to all.
- Monitor purpose of loan as House and renewable energy with very low instalment bucket has more defaults and Vacations with higher instalments has more defaults.

- **Improve Grade Criteria :**

- Implement lending criteria for lower grades (E, F, G), maximum loan amounts for these grades to reduce exposure to high-risk loans and Notability Grade F has more defaults with Very High loan amounts.

- **Specify DTI ratio :**

- Implement rule for DTI ratio, DTI levels should be a lower, up to 30% indicates a higher risk.

- **State Considerations :**

- guidelines for borrowers in states with higher default rates, such as (NV) and (TN).