LENDING CLUB CASE STUDY

Group Members:

Prasanna Rahavendra A
Kunal Mahale

This case study was done as part of the Statistics Essentials Module of the Executive PG Programme in Machine Learning & AI - November 2022 - IIIT, Bangalore and Upgrad.

PROBLEM STATEMENT & OBJECTIVES

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.

Performing EDA on the dataset and understanding how each of the attributes behave will help the company in reducing the risk of any loss.

STEPS IN PROBLEM SOLVING

Understanding the dataset

- Looking at the data dictionary to understand the meaning of each of the columns.

Data Pre-processing

- Data Cleaning
- Deriving new columns from existing ones.

Univariate Analysis

- Analysing and understanding all the categorical and numerical columns.

Segmented Univariate Analysis

- Grouping the data and analysing each segment of the data separately.

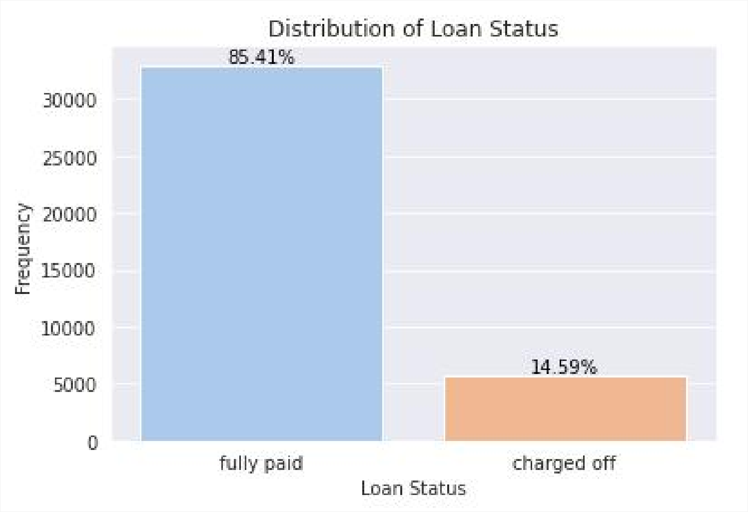
Bivariate Analysis

- Understanding the relationship between two columns.

Insights

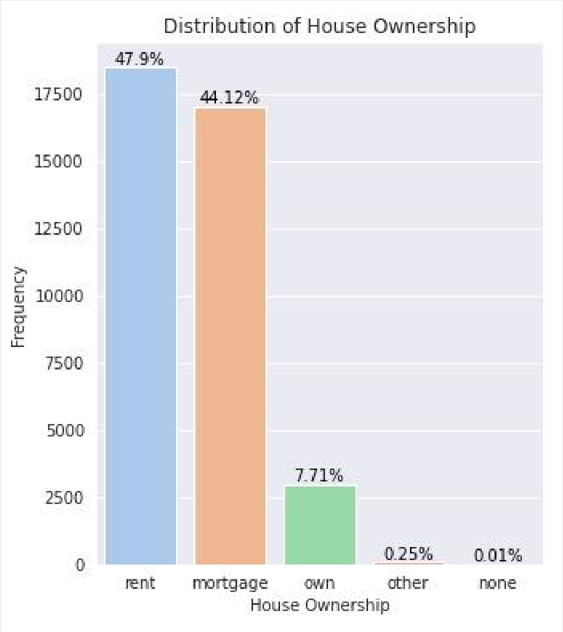
- Obtaining insights from the analysis.
- Recommendations to reduce loss to the company.

UNIVARIATE ANALYSIS – LOAN STATUS, TERM & HOUSE OWNERSHIP



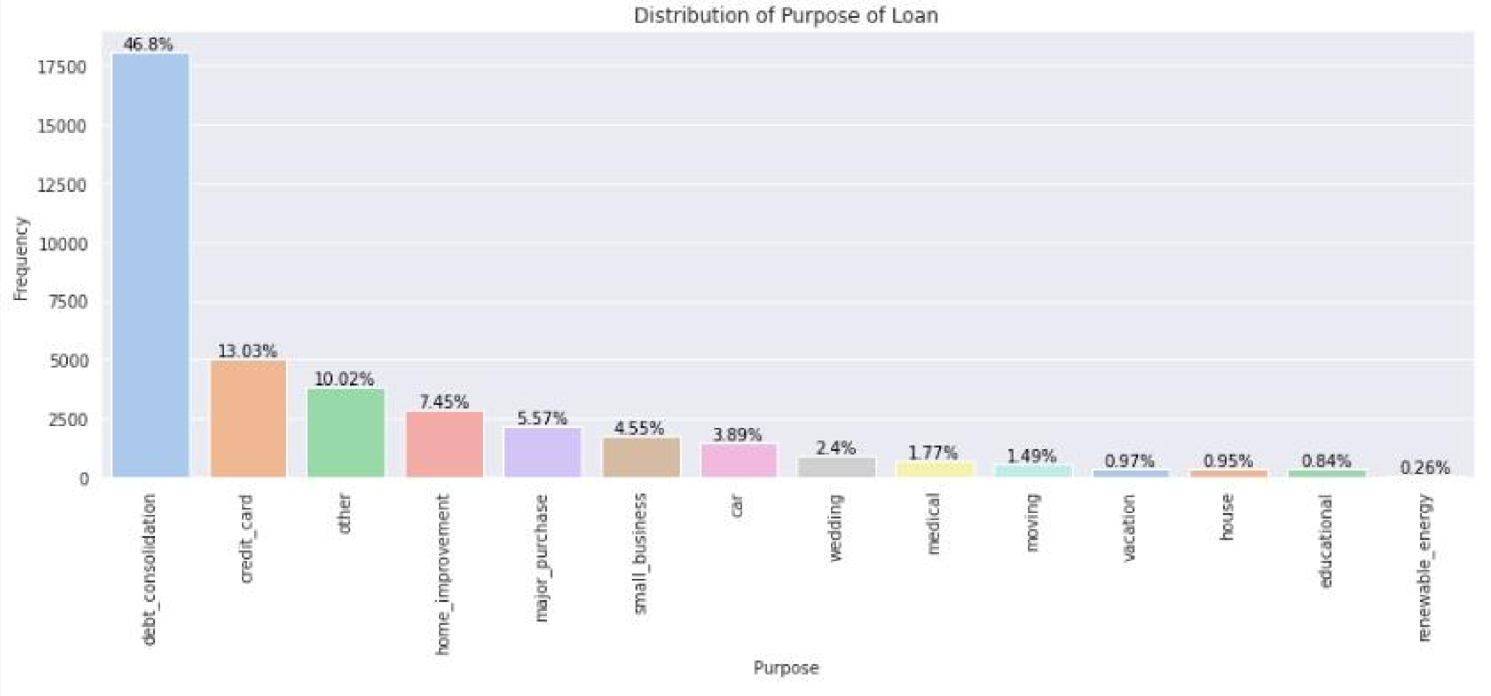
Around 85% of the loans were fully paid and ~14% defaulted.

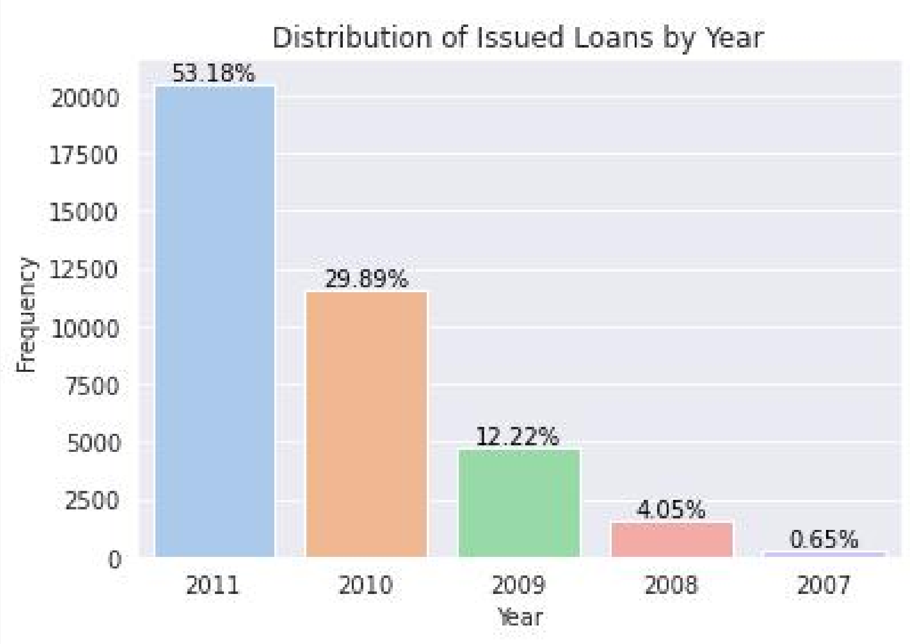
most borrowers (~75%) prefer a term of 36 months more than 60 months.



Most of the borrowers have a rented house or mortgage.

UNIVARIATE ANALYSIS - PURPOSE OF LOAN & ISSUED YEAR

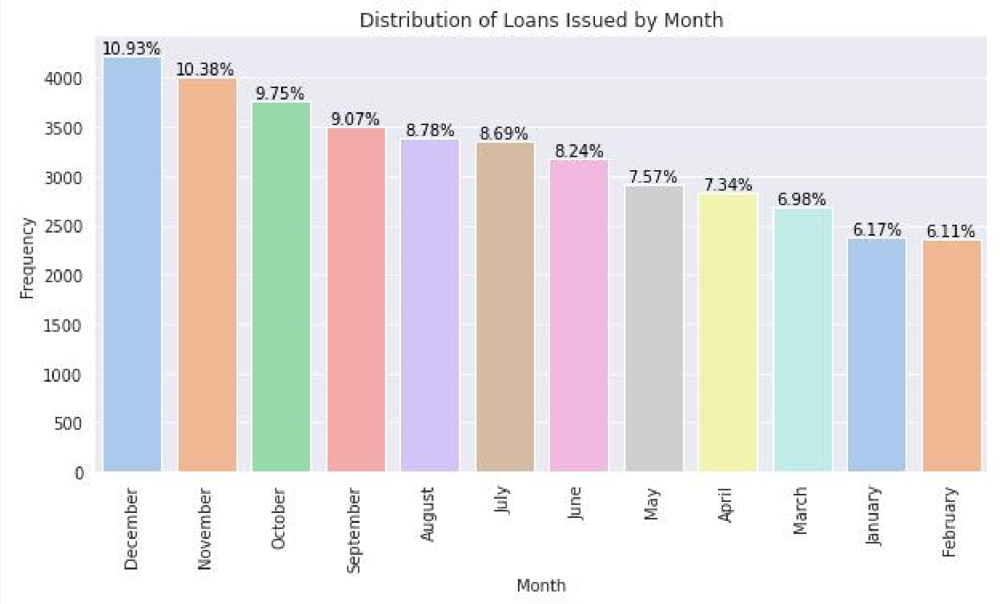




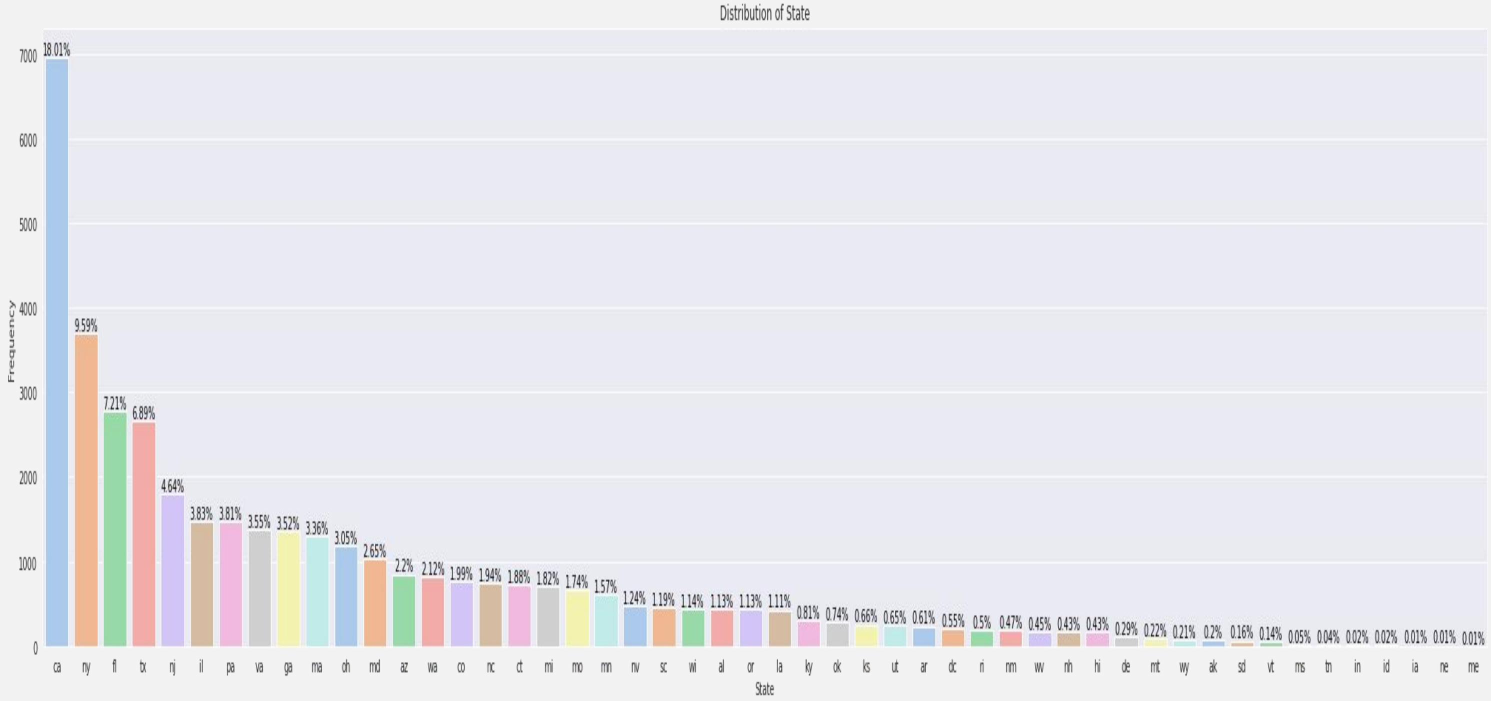
- Most loans were taken for dept consolidation i.e. taking a loan to pay off many other loans.
- Considering other purposes, people mostly get credit card loans followed by other (purpose unspecified), home improvement and so on.

The number of loans given increases every year. Most loans were given in the year 2011.

UNIVARIATE ANALYSIS – ISSUED MONTH & STATE OF BORROWER

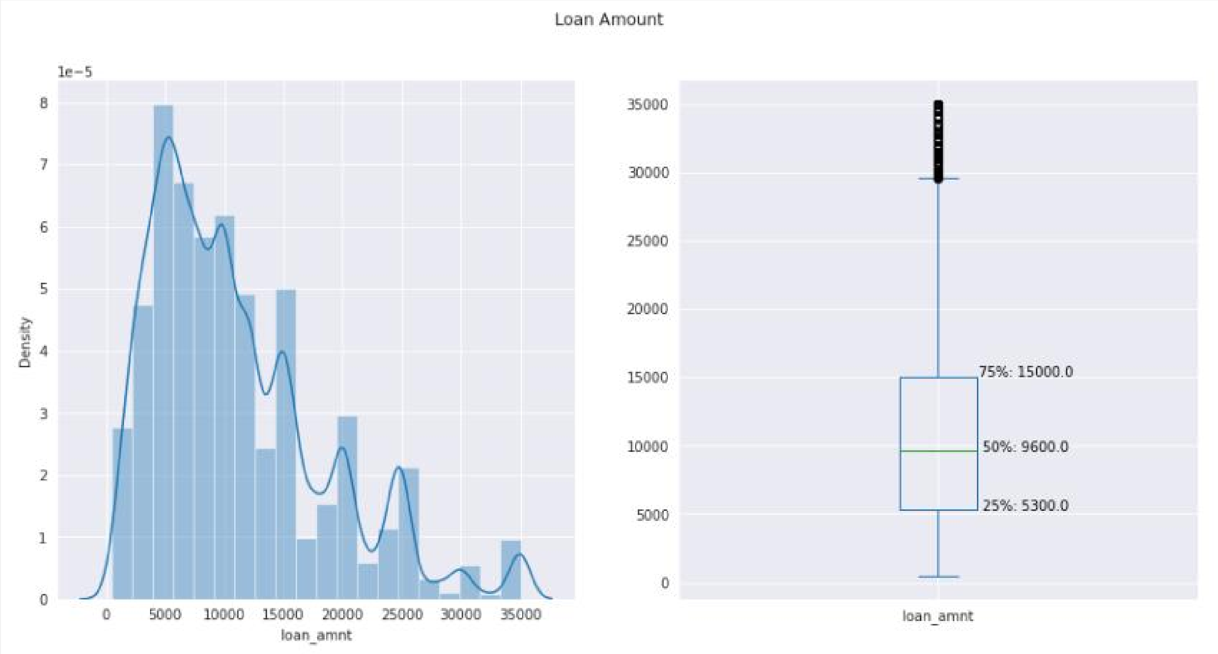


Most loans were given during the final months of a year (October, November & December) as they are festive seasons and people tend to spend more during those months.

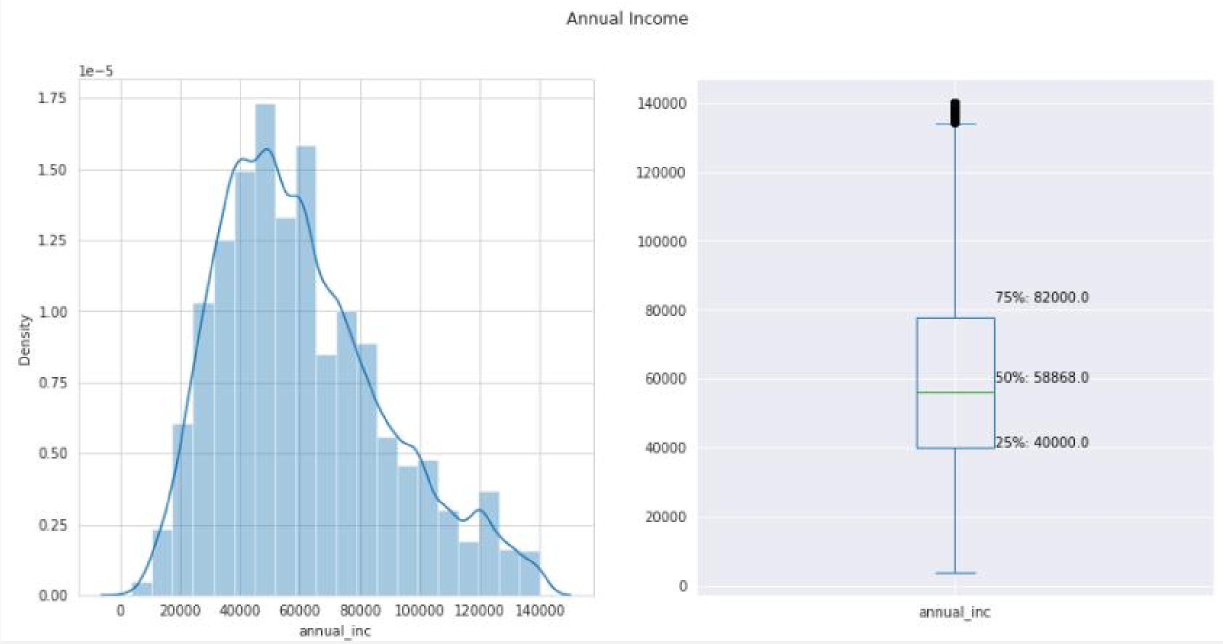


Most of the borrowers are from California (CA) followed by New Your (NY), Florida (FL) and so on.

UNIVARIATE ANALYSIS - LOAN AMOUNT & ANNUAL INCOME

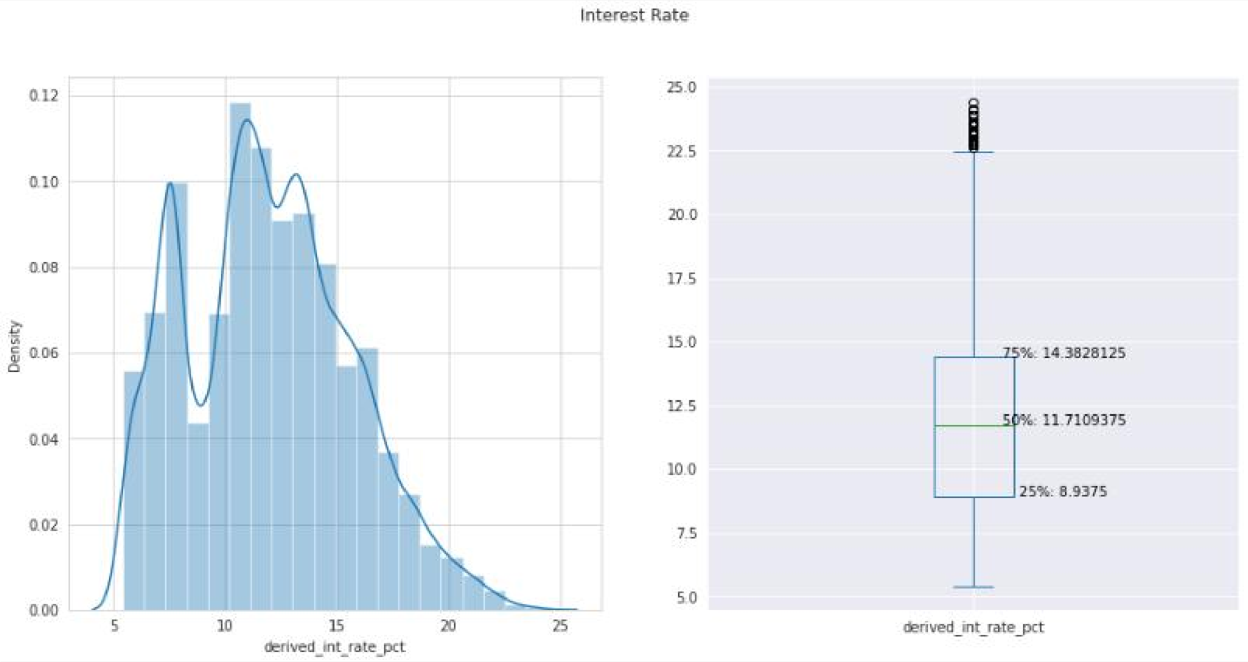


- Most loan amounts were approximately between 5000 and 10000. The median is 9600.
- In the histogram, there are many spikes near round numbers. This might be because, most people prefer rounded numbers more.

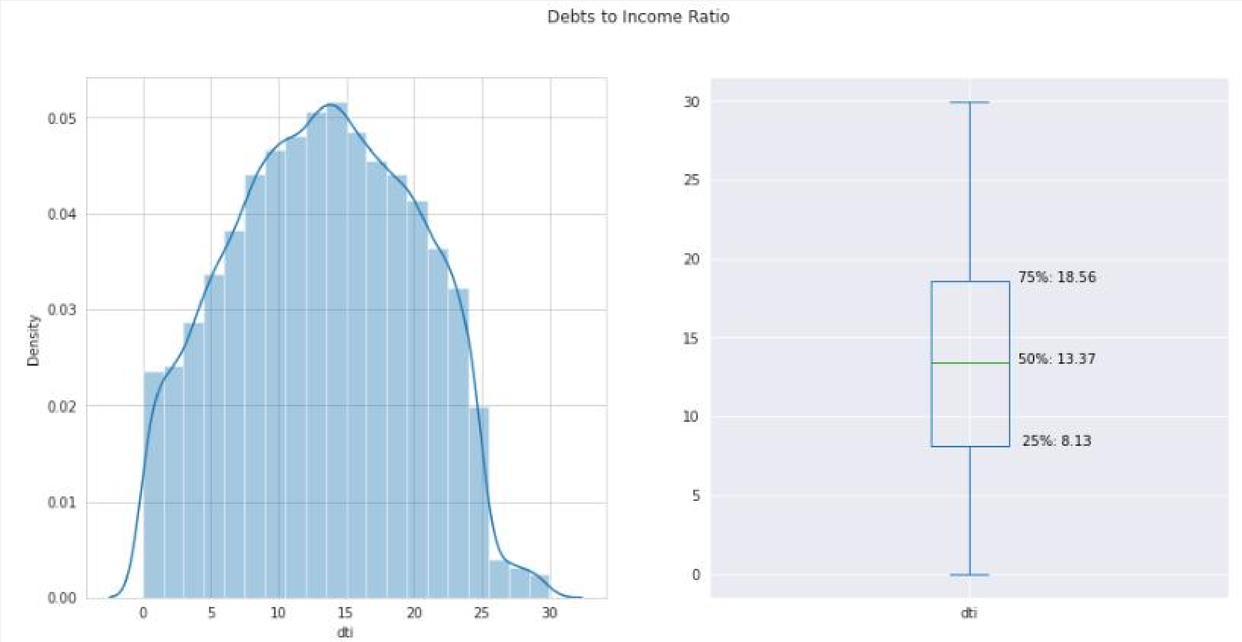


- An income of 58868 (median) is an representation of the borrowers.
- 95% of the people have an income below 140004.

UNIVARIATE ANALYSIS – DTI & INTEREST RATE

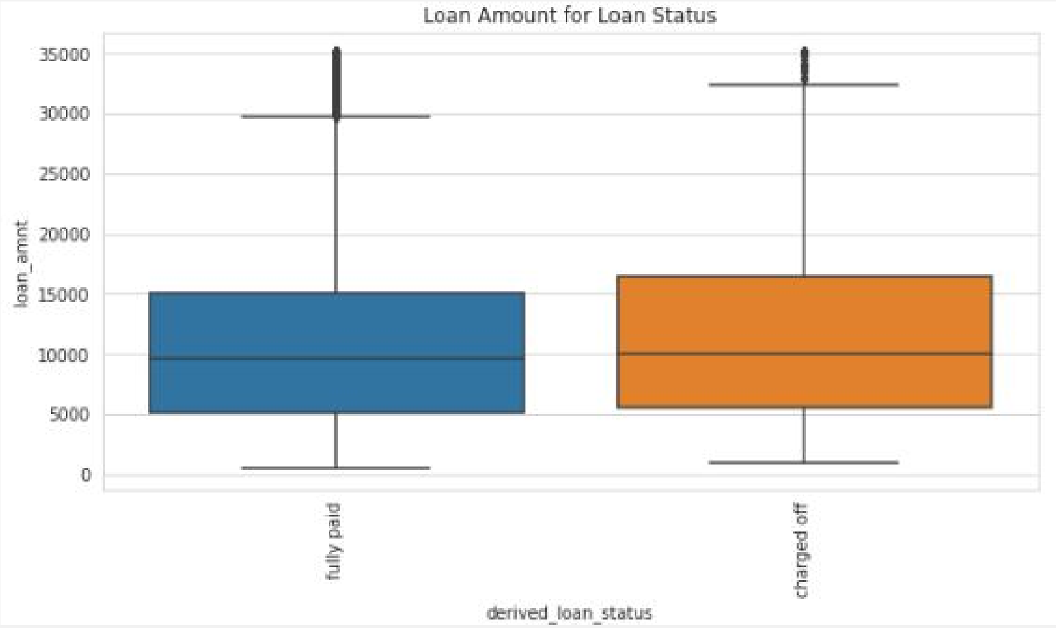


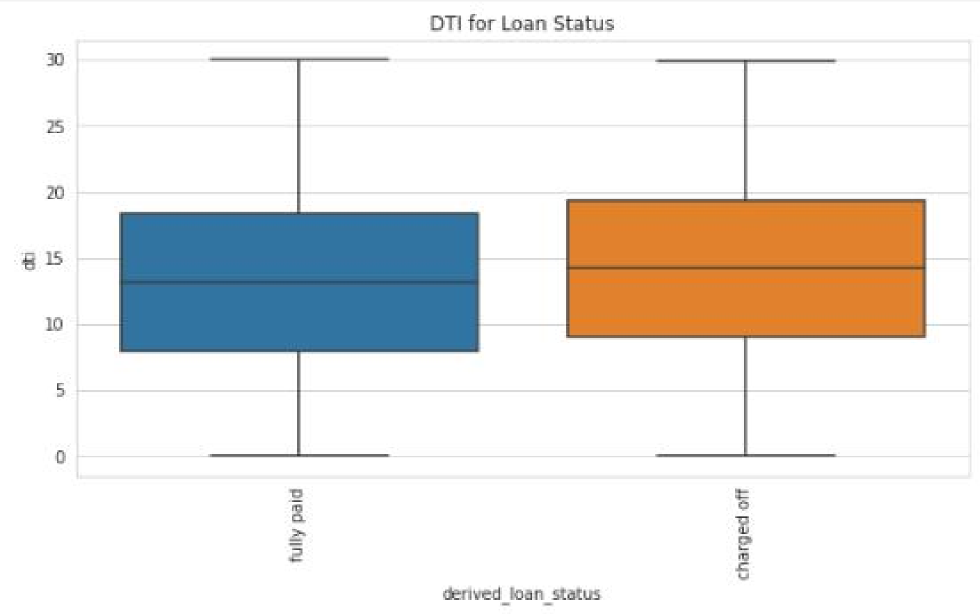
Most loan interest rates were between ~9% and 14%.

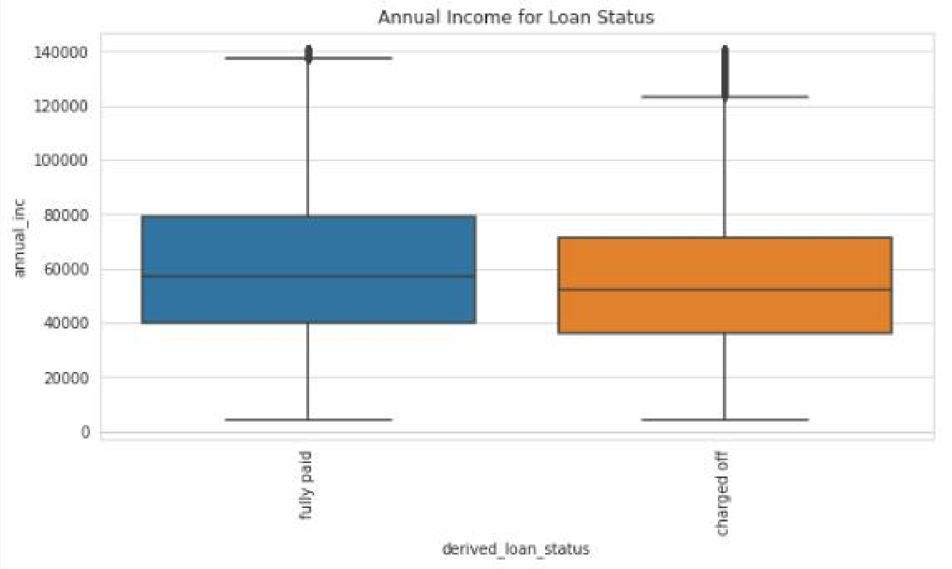


- greater the DTI lesser is the chance of getting a loan.
- All the approved loans have a DTI of less than 30 and most lie in the range of 8 to ~19.

Segmented Univariate Analysis of Loan Status



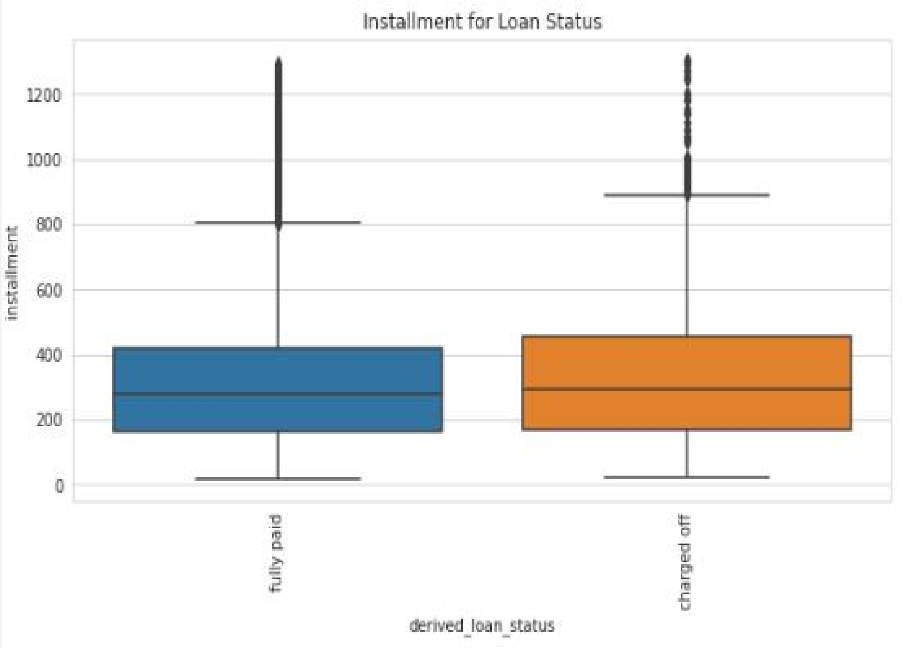


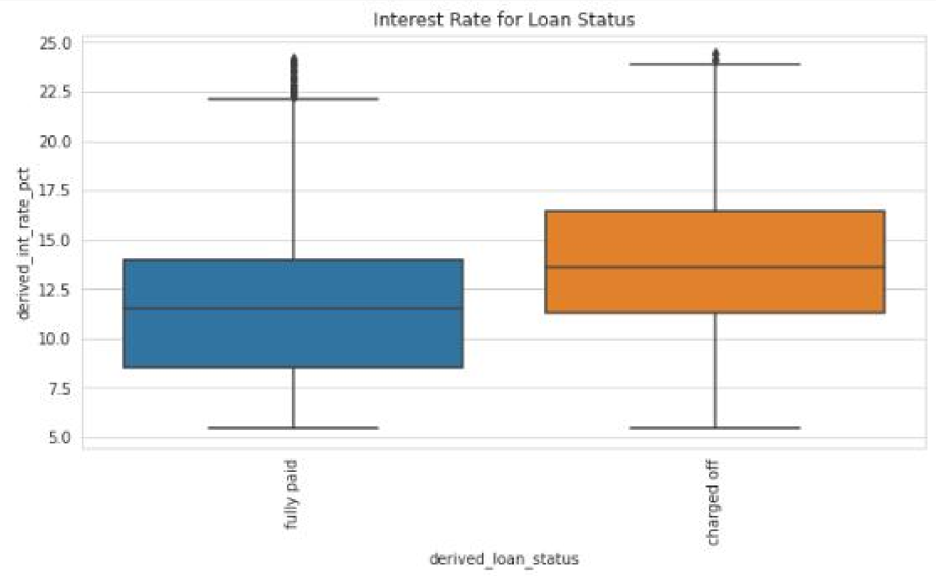


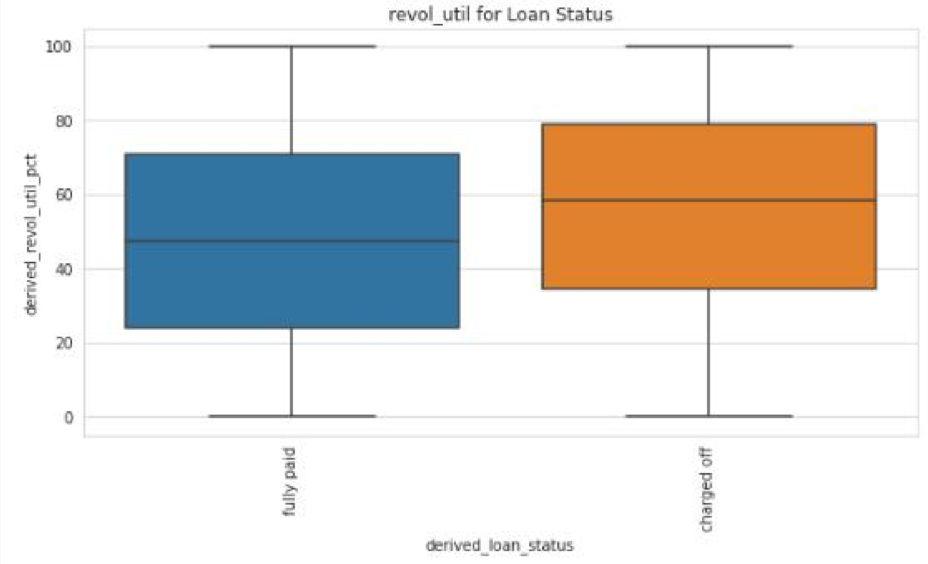
Insights:

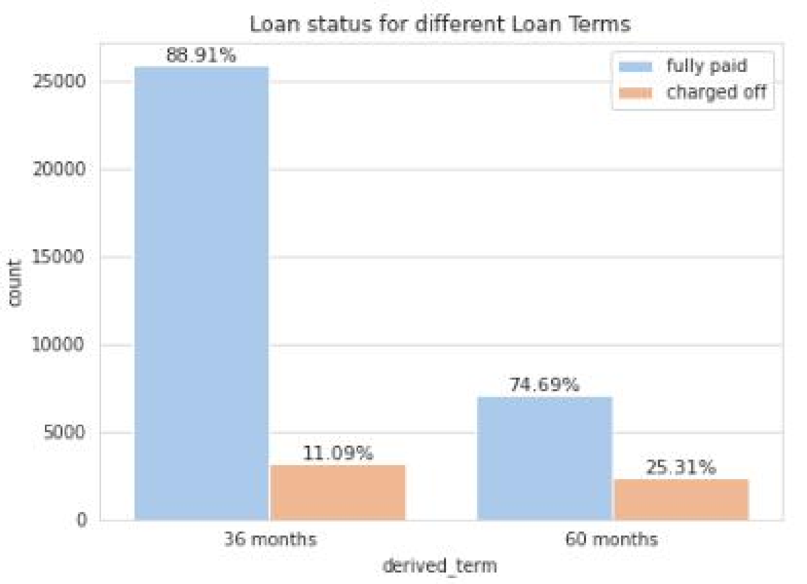
- Loan amount of fully paid loans seems to be a bit lower than charged off loans.
- Loans that have been fully paid seems to have a lower DTI compared to loans that were charged off.
- Annual income of borrowers who have fully paid the loans seems to be higher than borrowers whose loans charged off.

Segmented Univariate Analysis of Loan Status





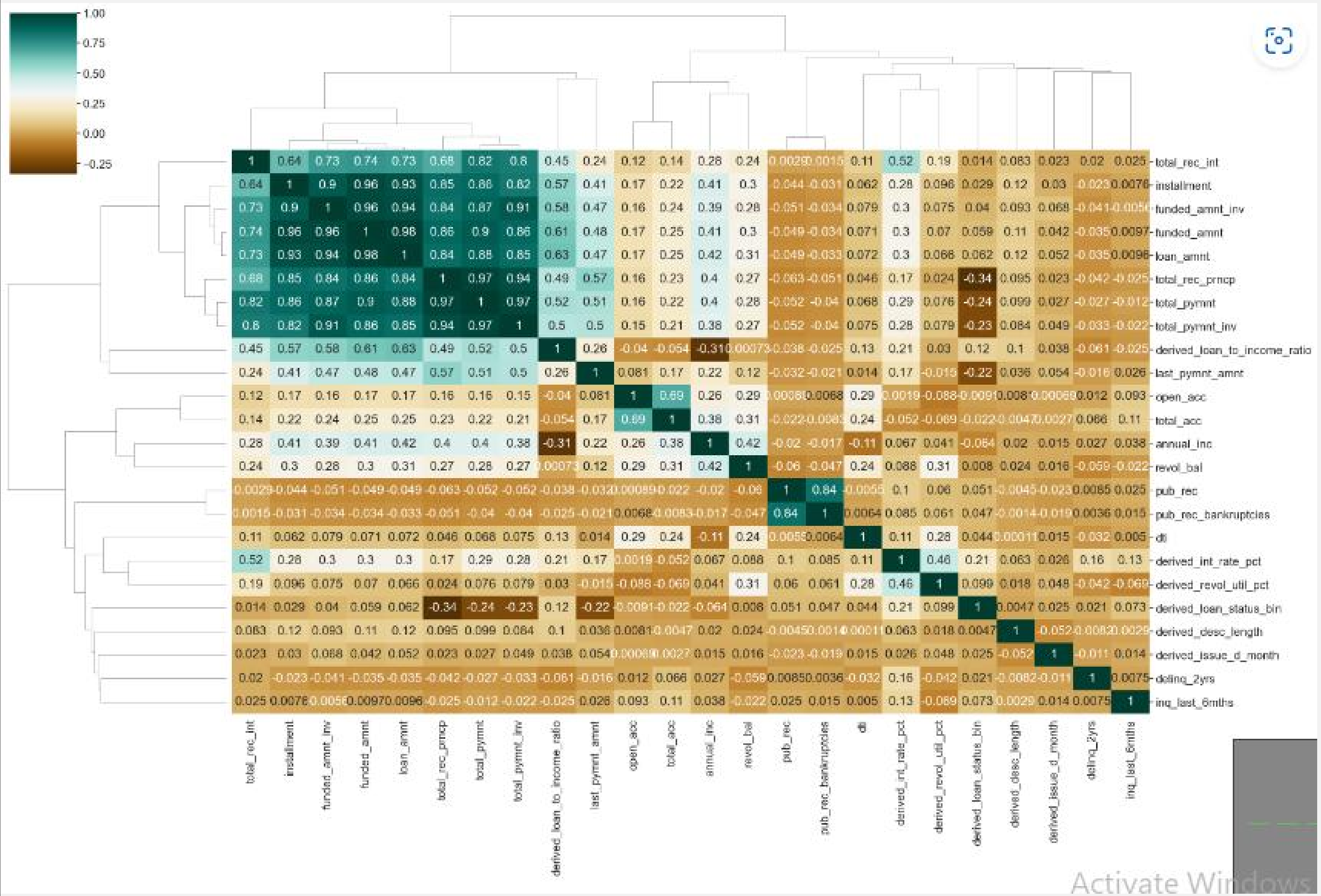




Insights:

- Fully paid loans have lower monthly installments compared to charged off loans.
- Loans that have been fully paid seems to have a lower interest rate compared to loans that were charged off.
- Higher the revolving utilization rate, lesser is the chance of the loan getting paid back fully.
- Borrowers who had chosen 60 months term tend to be defaulters (25.3%) more compared to borrowers who choose 36 months term (11.1%).

BIVARIATE ANALYSIS : CORRELATIONS

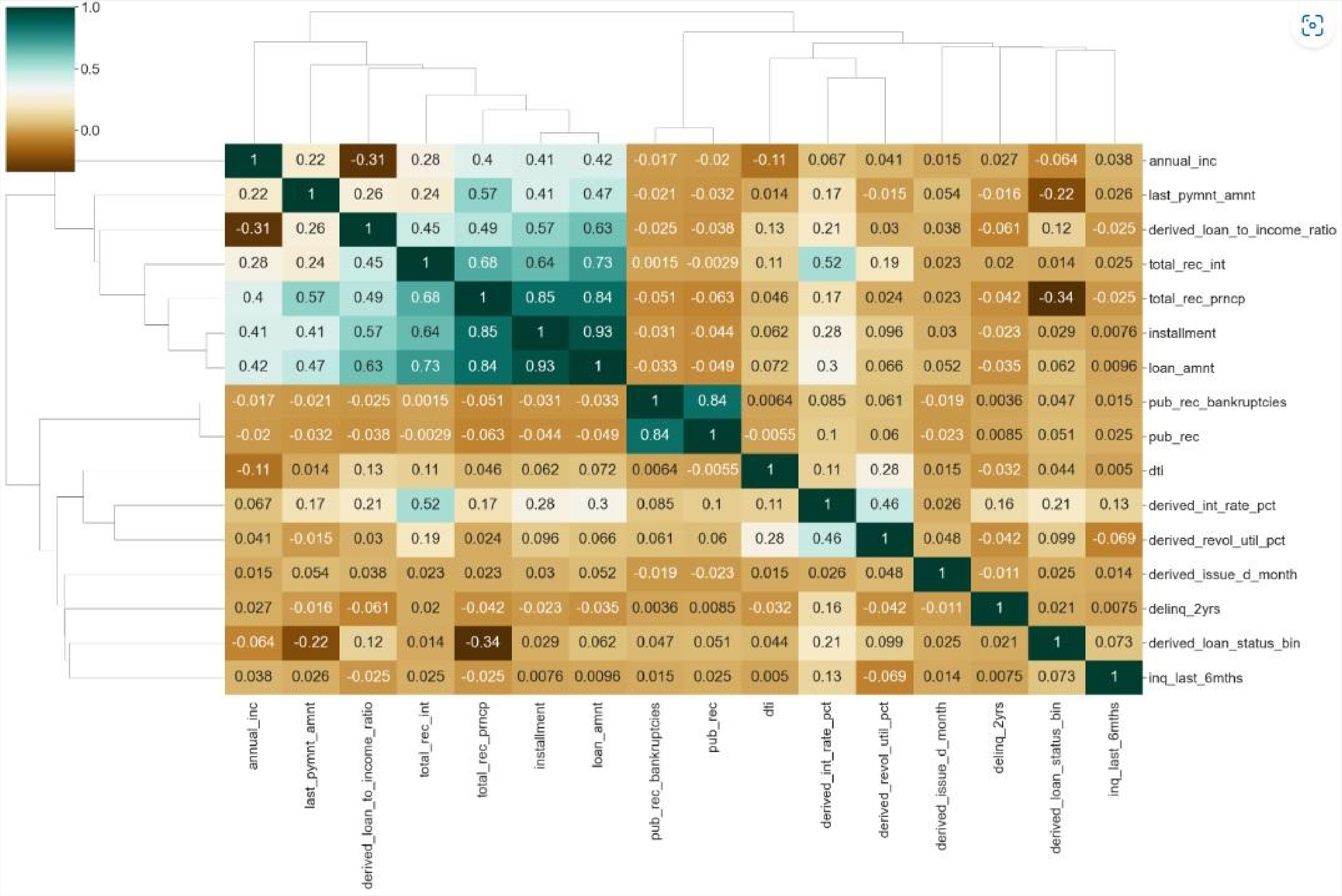


Insights:
Highly correlated groups :

- .[loan_amount, funded amount, funded amount by investors, total payment to investors, total payment] l
- public records and public records of bankruptcies
- open account and total open account

Current Loan Related Features:
1. annual income :
- .+ revolving credit balance
- + total payments
- + total accounts & open accounts
- - dti
- - loan status
2. interest rate :
   + total recovered interest
   + revolving utilization of credit
   + loan amount
   - total accounts
3. total recovered principal
   + annual income
   + revolving cred bal
   - public records & public records of bankruptcies, delinq_2yrs

BIVARIATE ANALYSIS : CORRELATIONS



Class : Credit history related variables

Debt to income ratio [dti]:

+ open accounts
+ revolving credit utilization %
+ total accounts
+ revolving balance

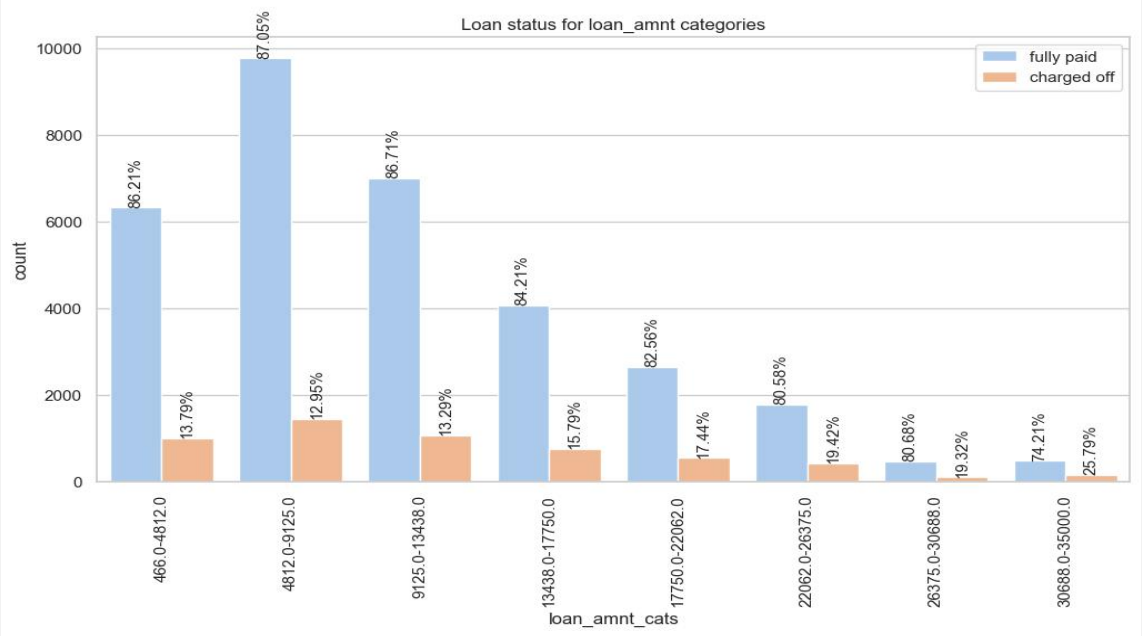
-annual income
-delinq_2yrs

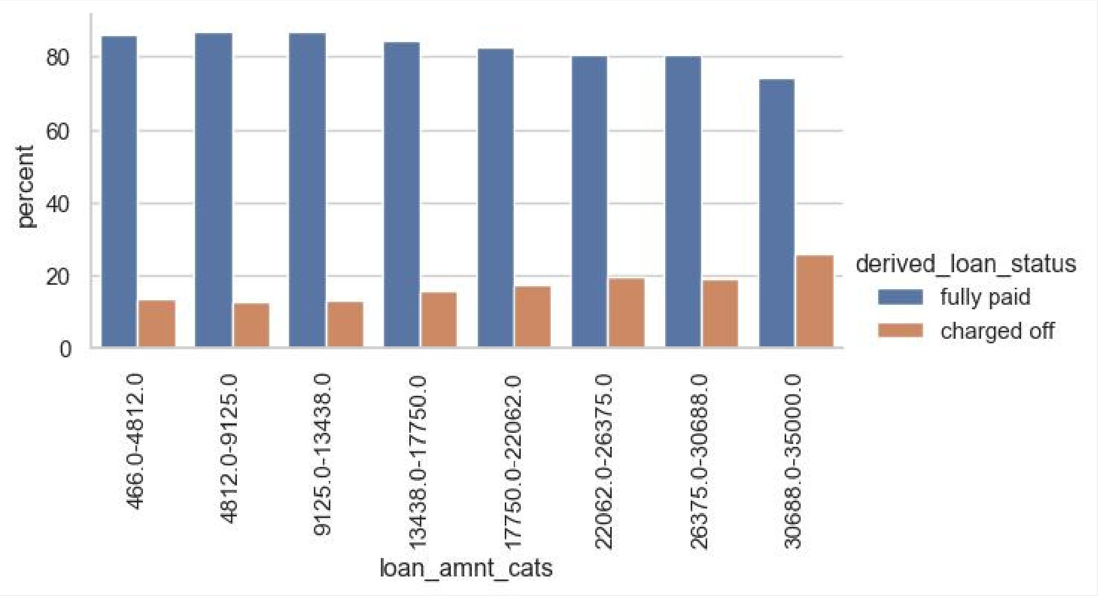
delinq_2yrs :

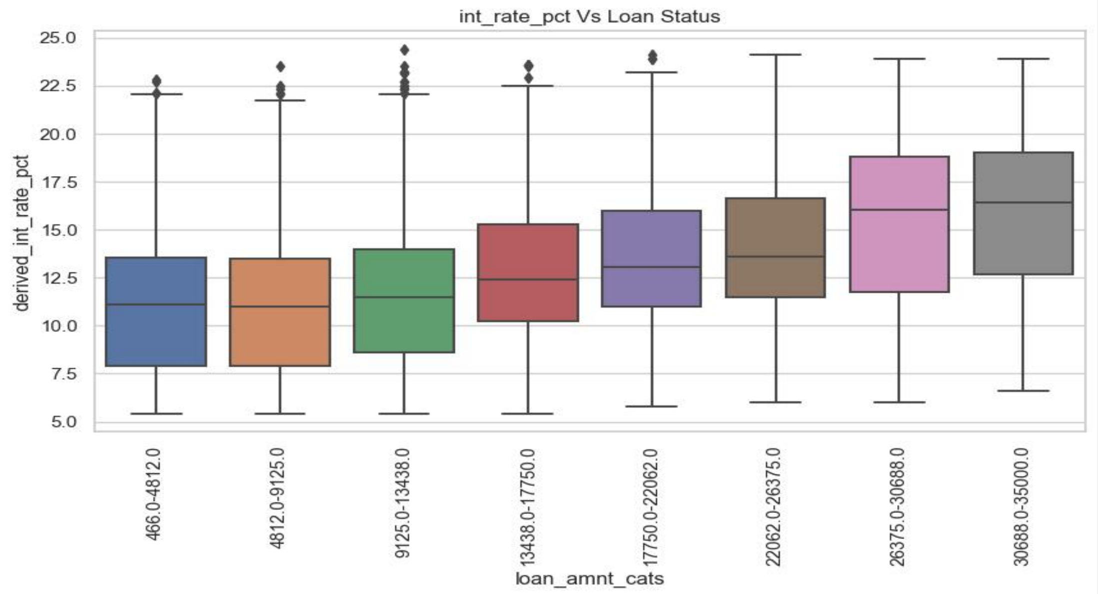
+ interest rate

- loan amount
- revolving credit balance

Bivariate Analysis: Impact of ‘Loan amount’ on Loan Status



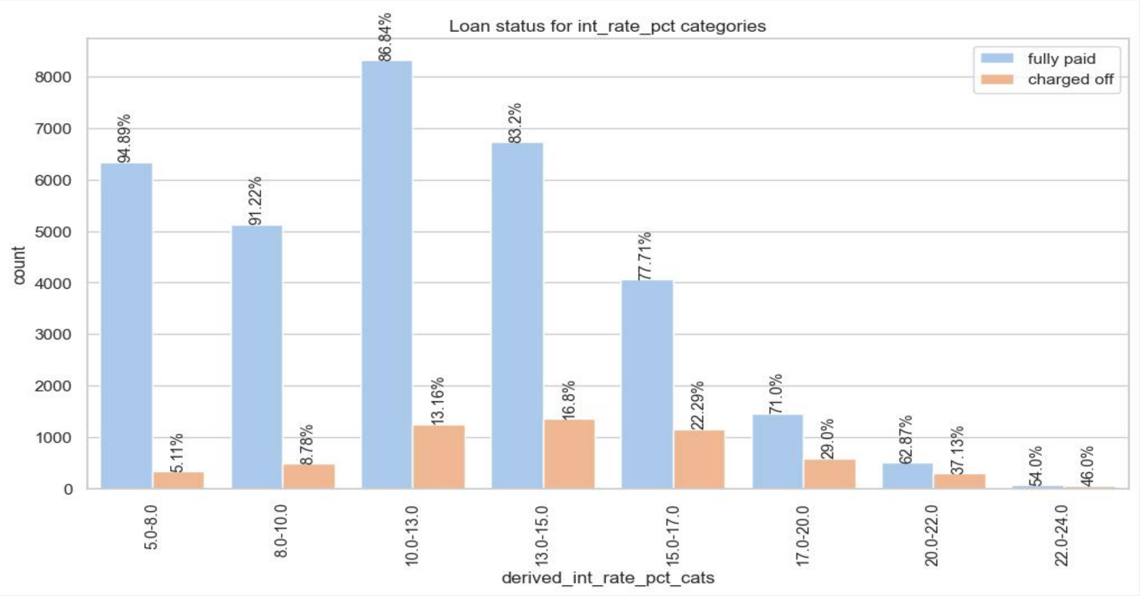


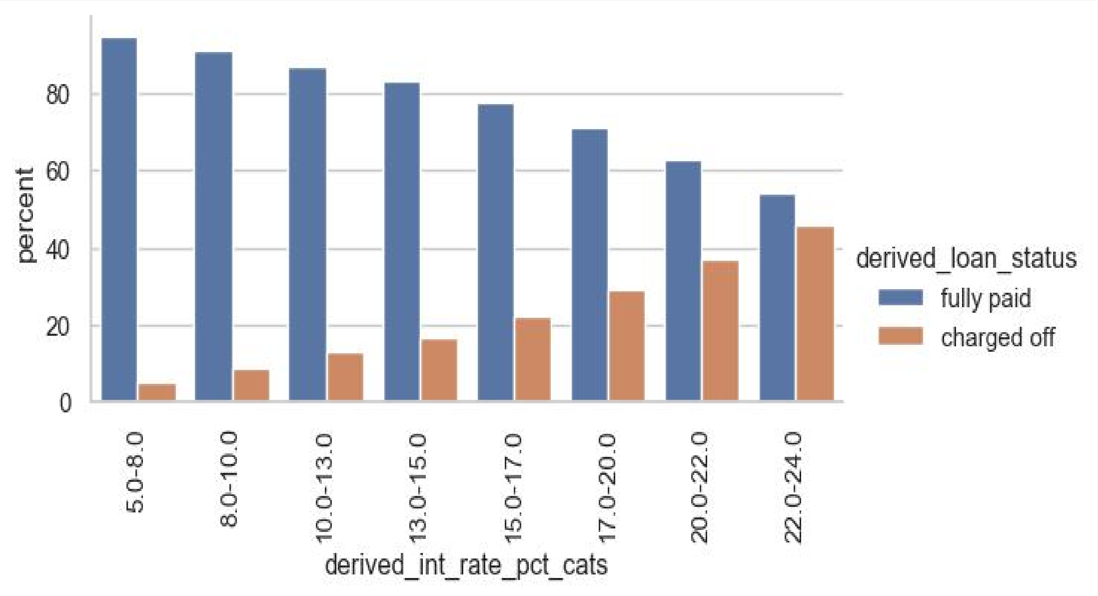


Insights:

- Loan amount is divided into multiple categories.
- More number of loans are approved from 500-13K
- Default rate increase as the loan amounts increase
- For the defaulted loans, the interest rate increase as the loan amount increases
- % of defaults increase as loan amount & interest rates increase

Bivariate Analysis: Impact of 'Interest rate' on Loan Status



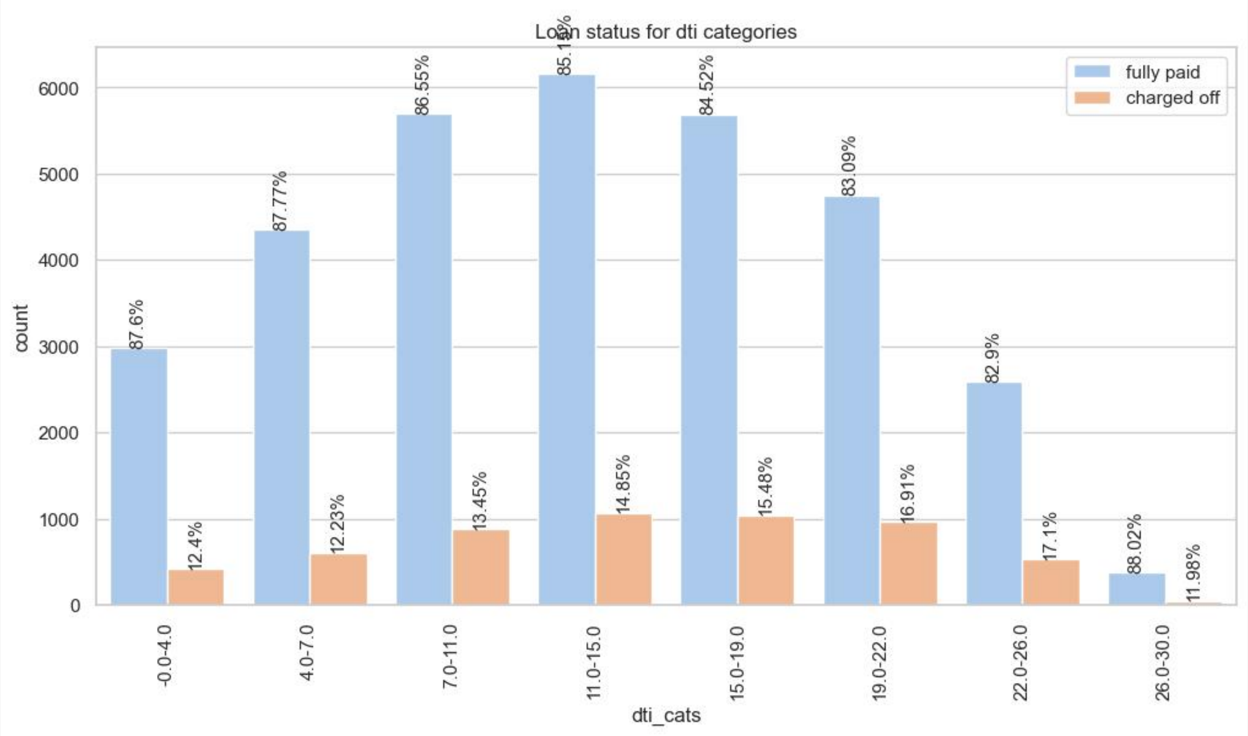


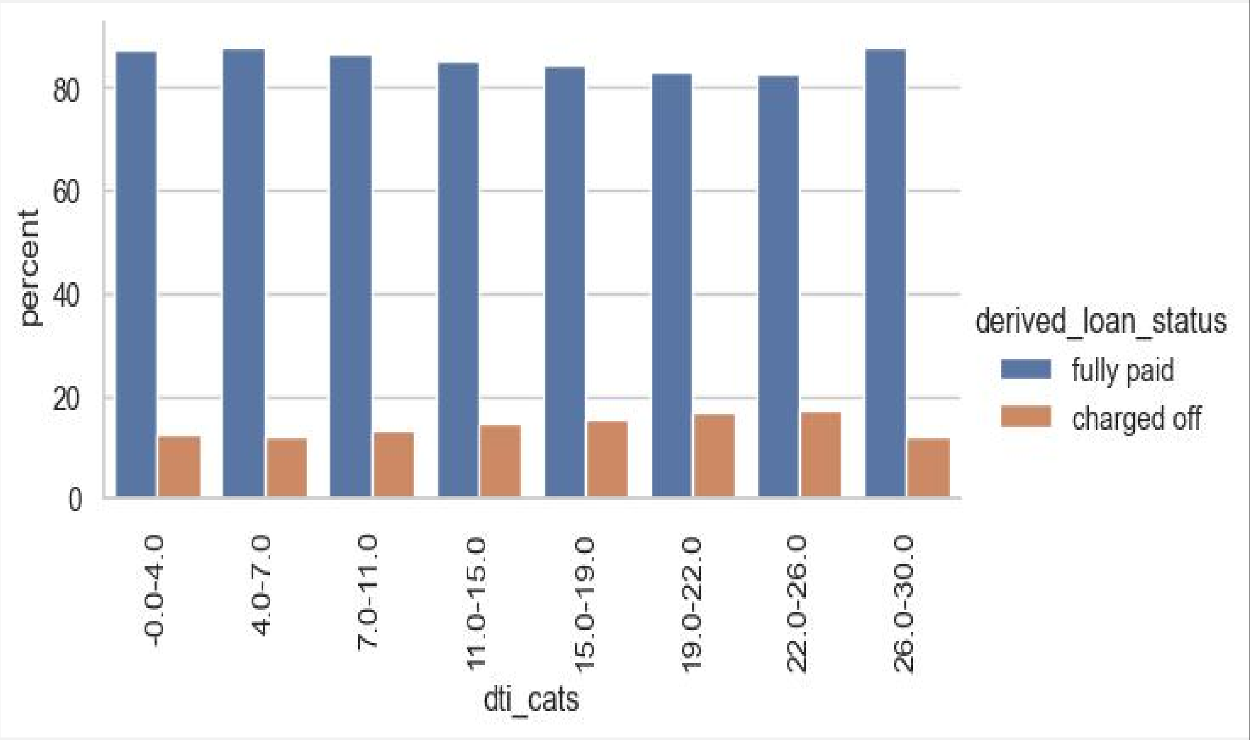
Insights:

- Higher interest rates == higher loan defaults
- More loans are distributed between 5-17% of interest rates
- Default rate increase as the loan amounts increase
- loans w/ higher interest rates are more likely to be charged off, strong trend
- loan w/ int rates till 10 % have less chance of getting charged off (<= 8%).

derived_int_rate_pct_cats	derived_installment_to_monthly_salary_ratio	dti	annual_inc	loan_amnt	derived_loan_to_income_ratio	derived_loan_status_bin
5.0-8.0	0.053	12.124	64099.556	8244.422	0.150	0.051
8.0-10.0	0.062	12.406	63475.352	9475.992	0.172	0.088
10.0-13.0	0.067	13.488	64826.192	10921.260	0.189	0.132
13.0-15.0	0.069	13.996	63570.863	10915.771	0.191	0.168
15.0-17.0	0.072	14.006	65725.492	12348.698	0.206	0.223
17.0-20.0	0.078	14.168	74067.405	16324.867	0.241	0.290
20.0-22.0	0.086	14.500	82066.647	19859.499	0.267	0.371
22.0-24.0	0.093	15.109	87846.174	22504.500	0.277	0.460

Bivariate Analysis: Impact of 'DTI' on Loan Status

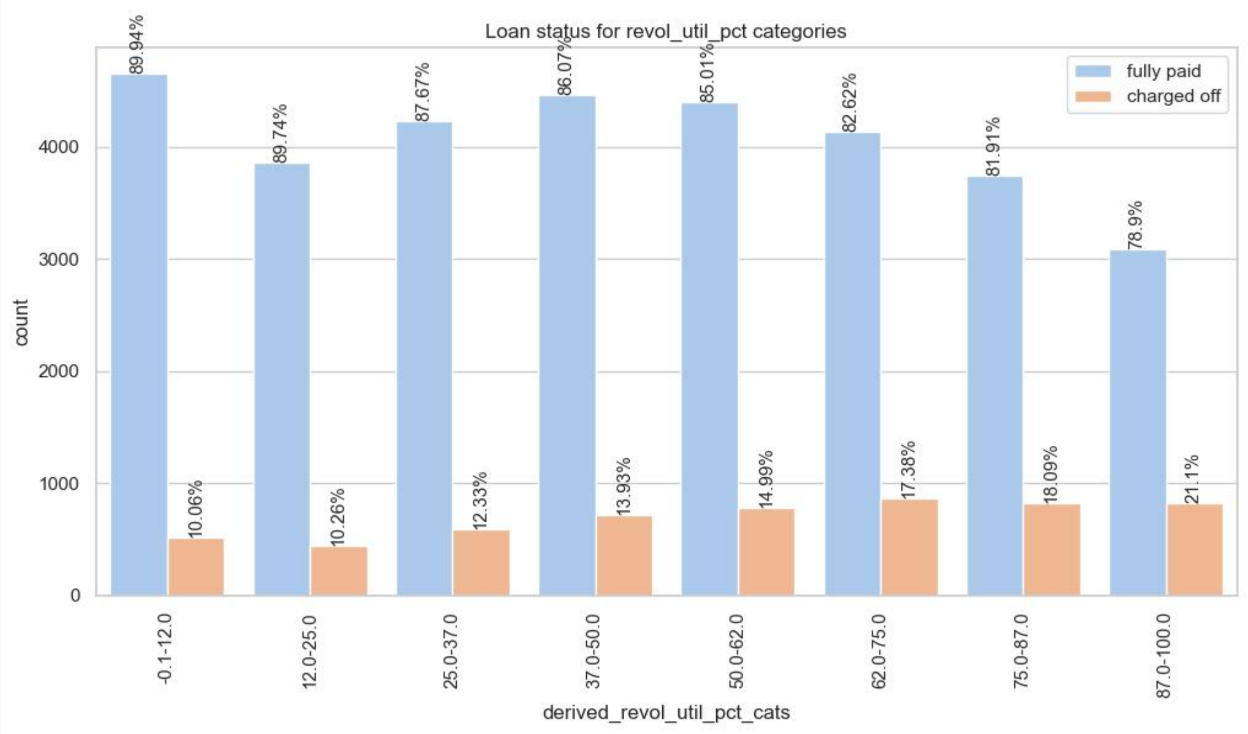


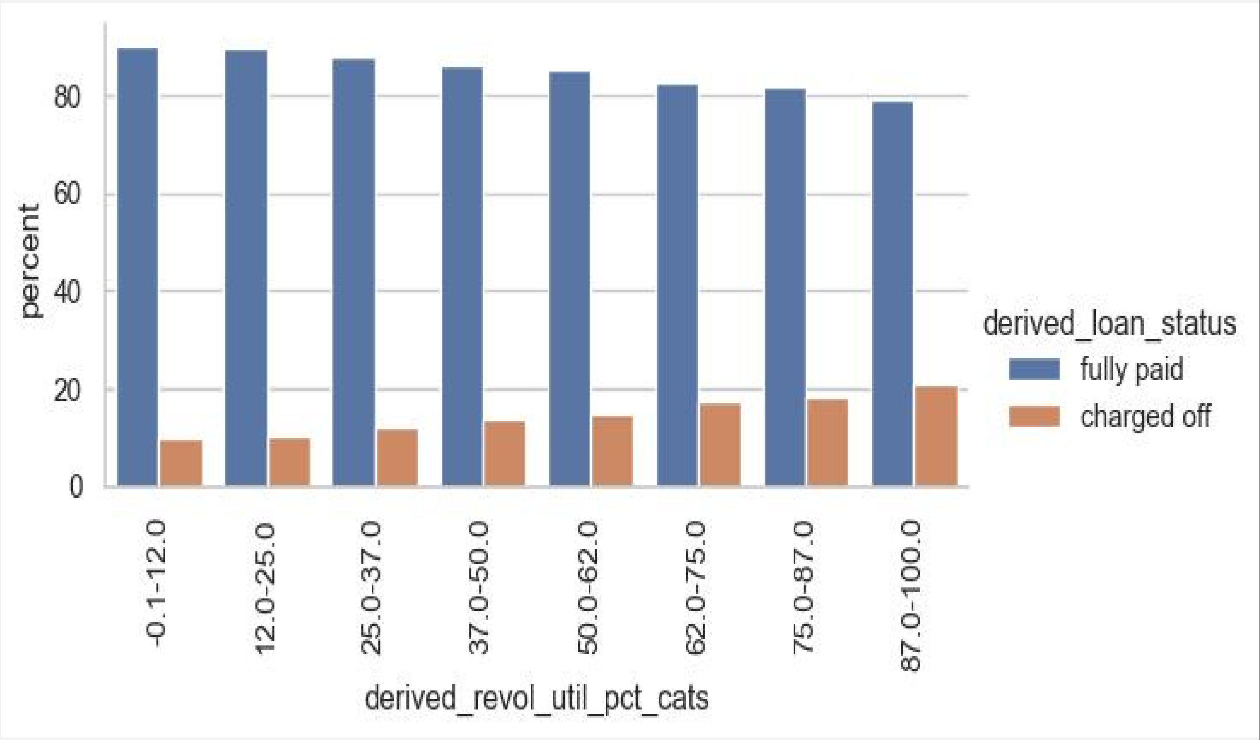


Insights:

- borrowers with dti of 7-10 are getting loan more than rest
- there is a slight increase in the % of loan getting 'charged off' from dti score of 7-26
- borrowers w/ dti score of 26+ and 0 to 7 are less likely [10-13%] to charge off the loan

Bivariate Analysis: Impact of ‘Revol util’ on Loan Status

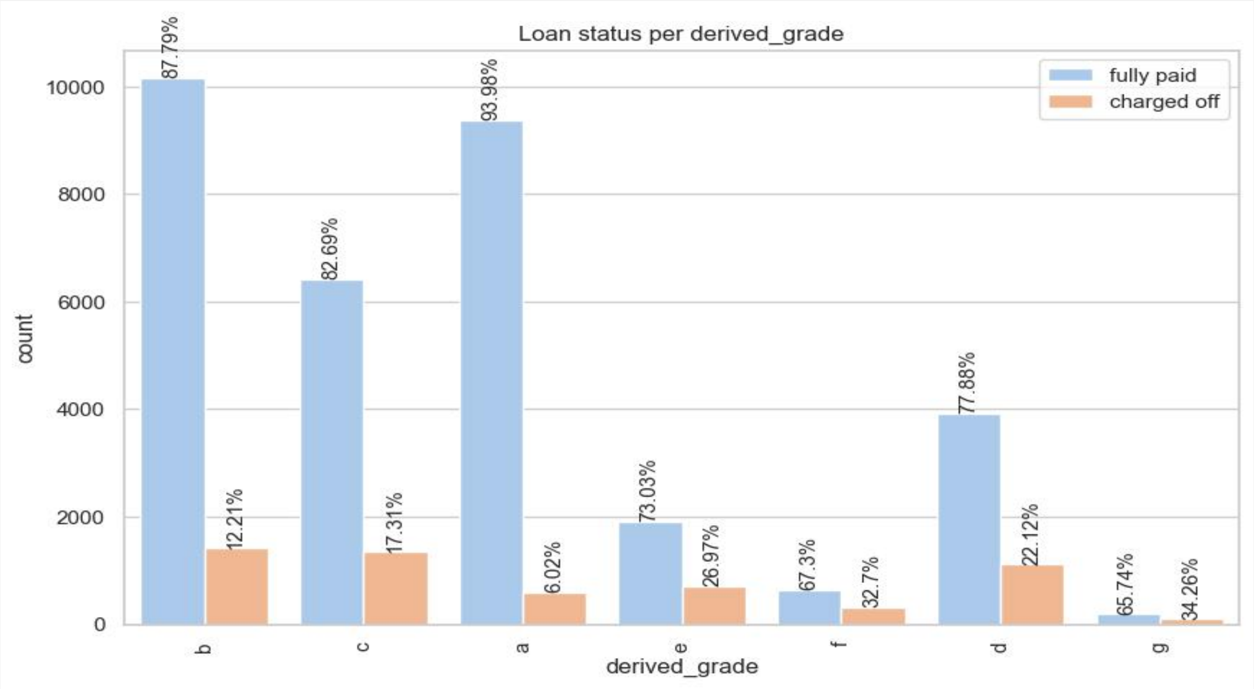


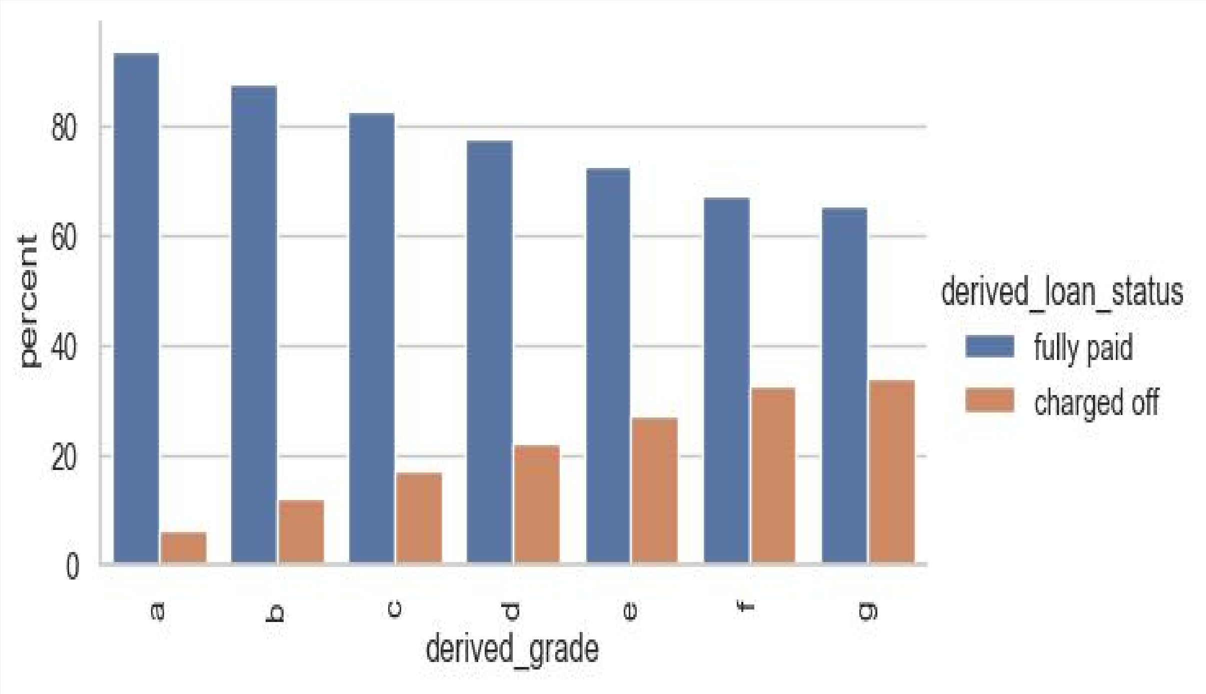


Insights:

- number of borrowers increase with the rate of utilization of available revolving credit
- rate of defaulting sligtly increase as the rate of revol util increases, which seems obious as people tend to use more credit and eventually charge of some loans.

Bivariate Analysis: Impact of Loan Grade on Loan Status



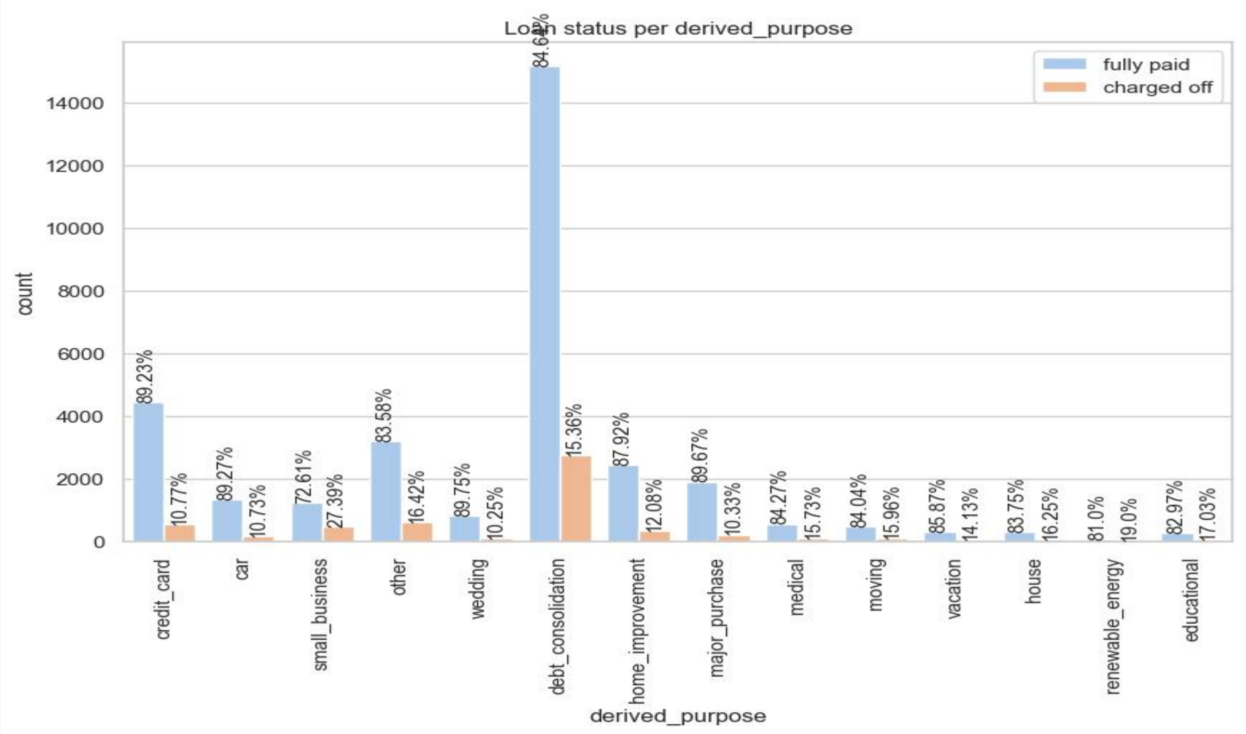


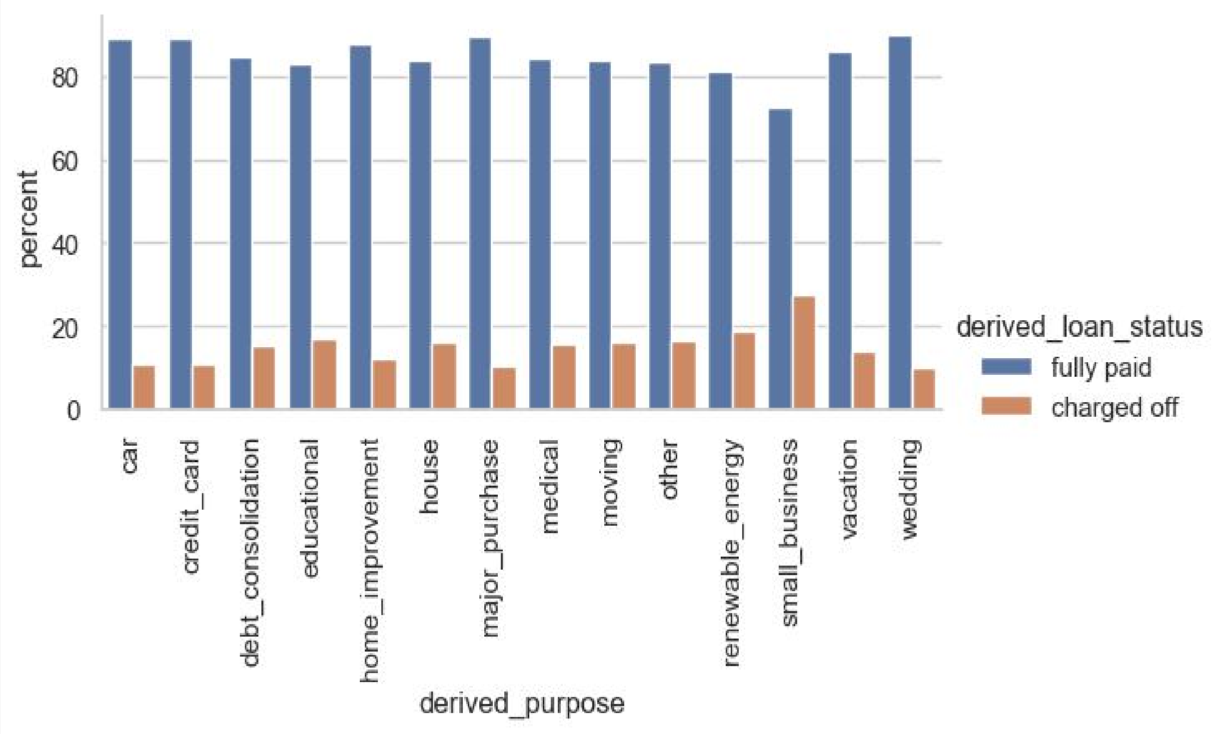
Insights:

- 1. loan grades of a-d are very frequent, e-g grades are comparatively rare
- 2. increase in loan grade significantly increases default risk

- Imp References :
- 1. wrt https://www.lendingclub.com/foliofn/rateDetail.action, interest rates increase from a-g in increasing order
- 2. As per table 3, higher loan amount gives some benefit in the loan grade
- 3. Higher term [ie 60 months], provides 8-4 points benefits in interest rate

Bivariate Analysis: Impact of ‘Purpose’ on Loan Status

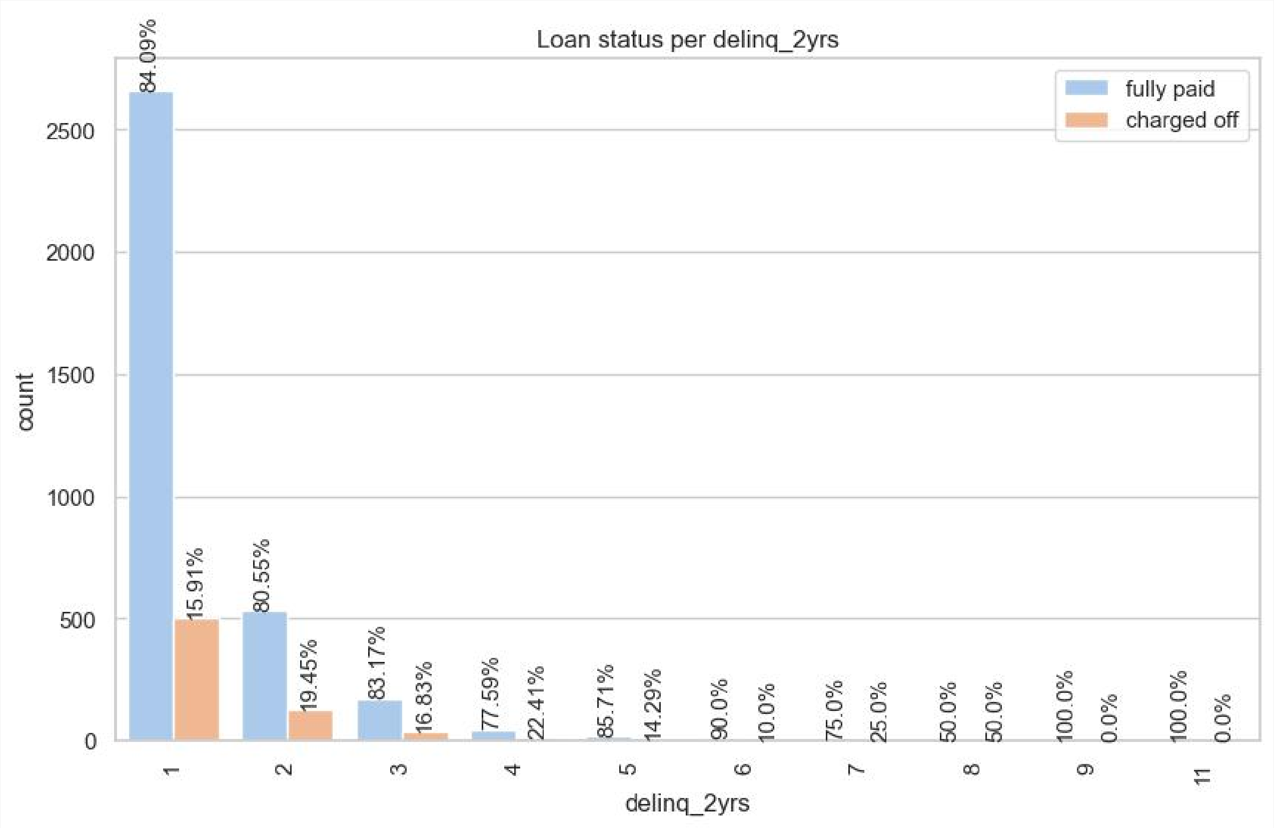


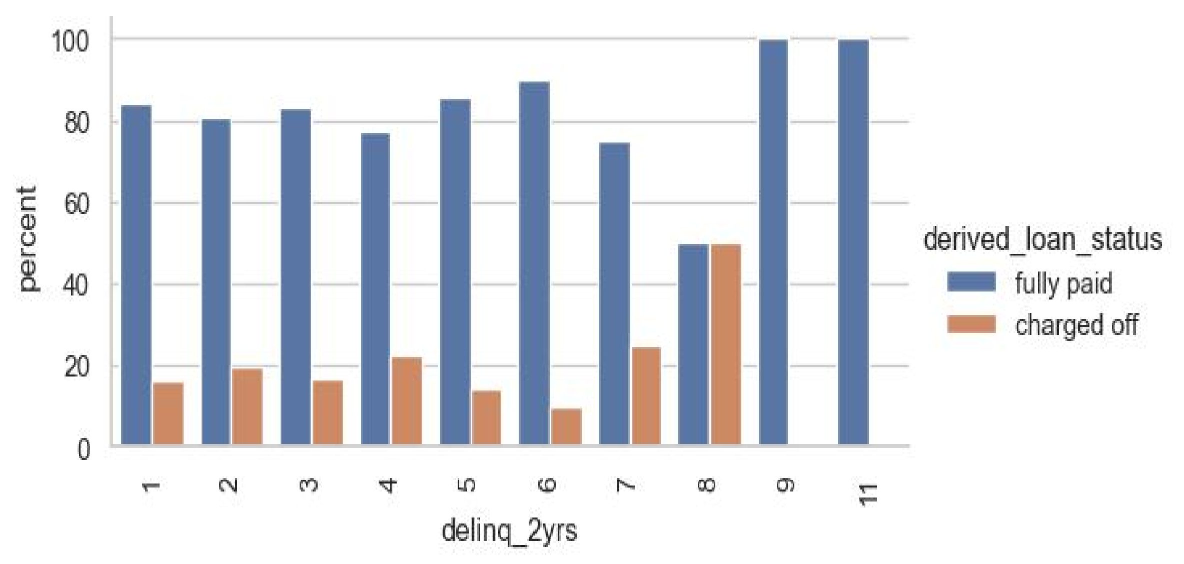


Insights:

- 1. Most of the loans are for debt_consolidation, followed by credit card and other
- 2. Categorywise Risk of defaults lies
- 1. small business [27%]
- 2. renewable energy [19%]
- 3. educational [17%]
- 3. Frequency based Risk of default lies with :
- 1. debt_consolidation [2k+]
- 2. other [400+]
- 3. credit card [400+]

Bivariate Analysis: Impact of ‘delinques’ on Loan Status

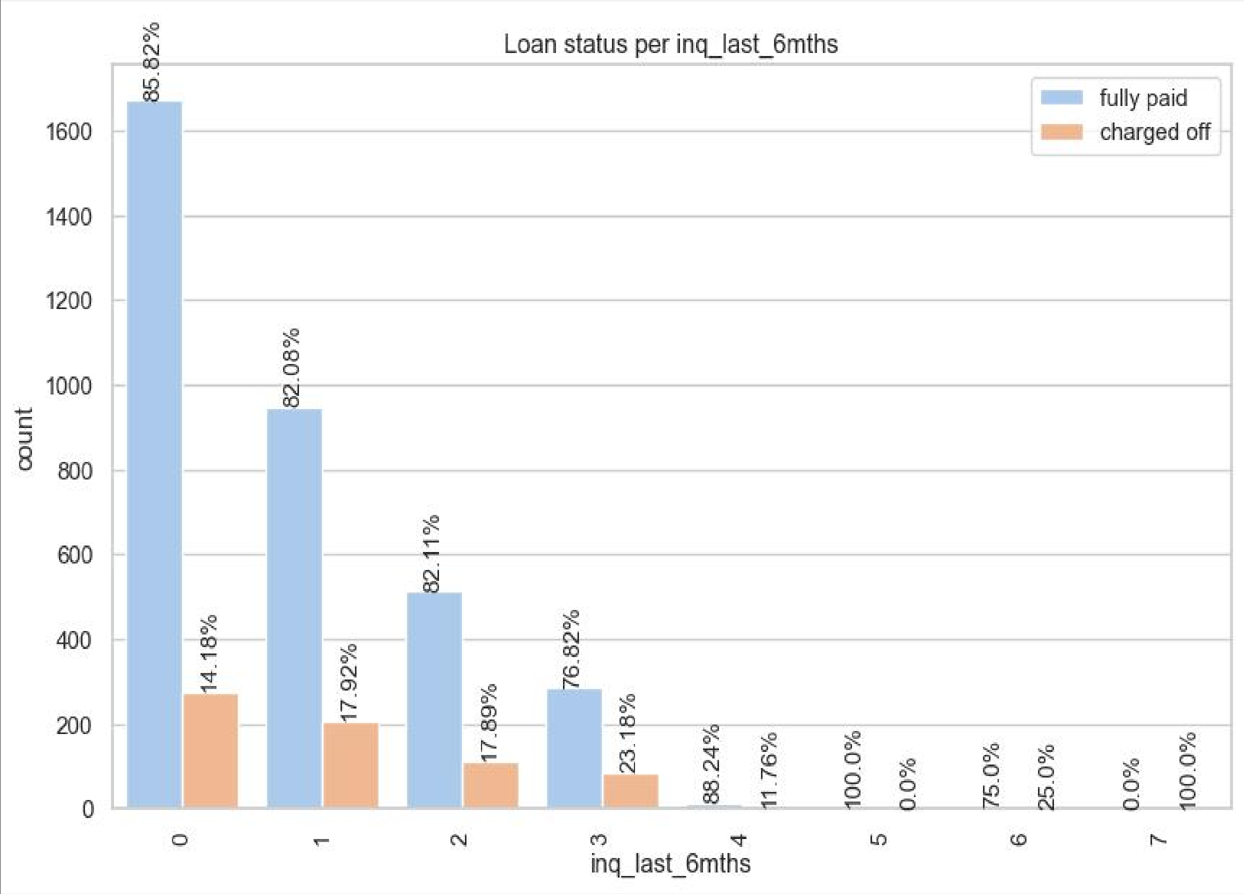


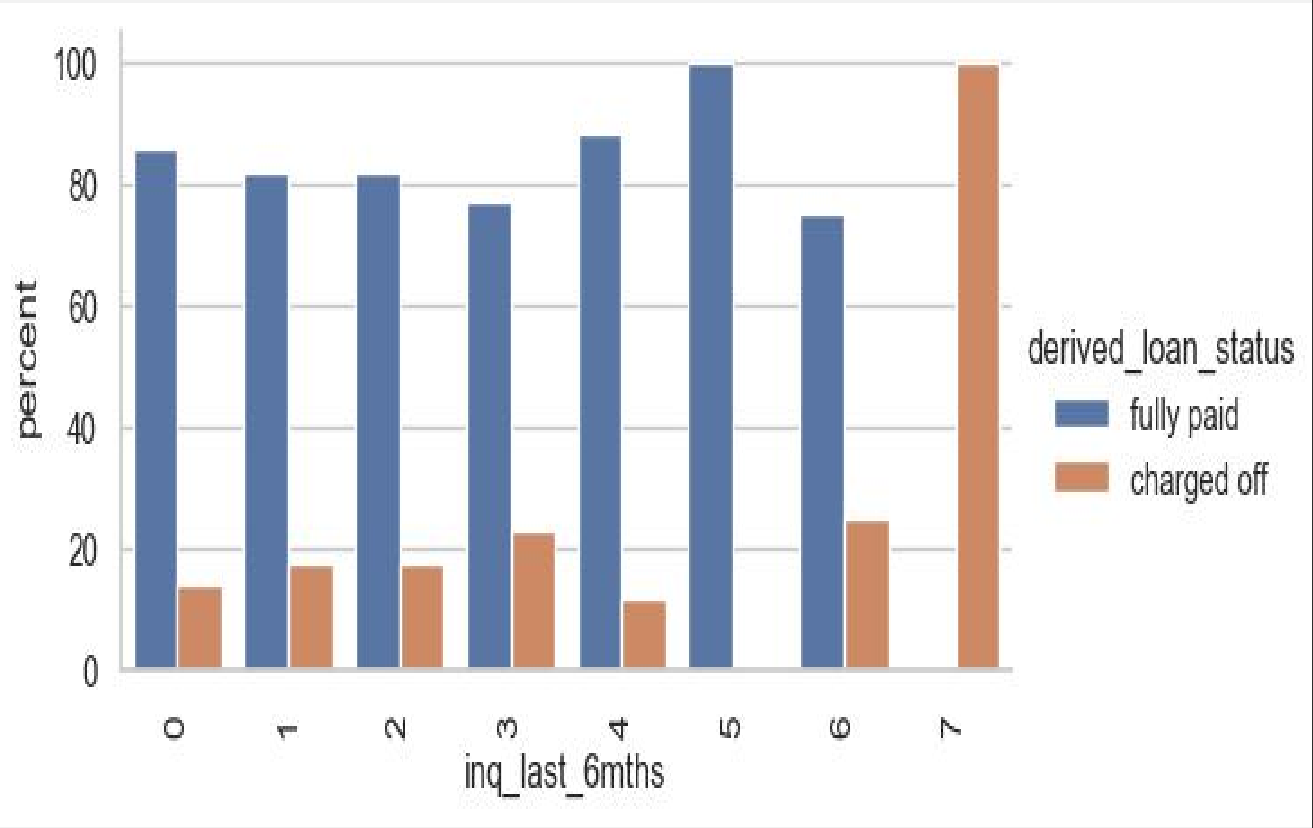


Insights:

- 1. There are borrowers having delinq history, though the proportion is small.
- 2. Among the borrowers with delinq history, 1 & 2 occurences are more frequent than others
- 2. risk of default increases as delinq occurrences increase

Bivariate Analysis: Impact of 'Enquiries For Credit' on Loan Status

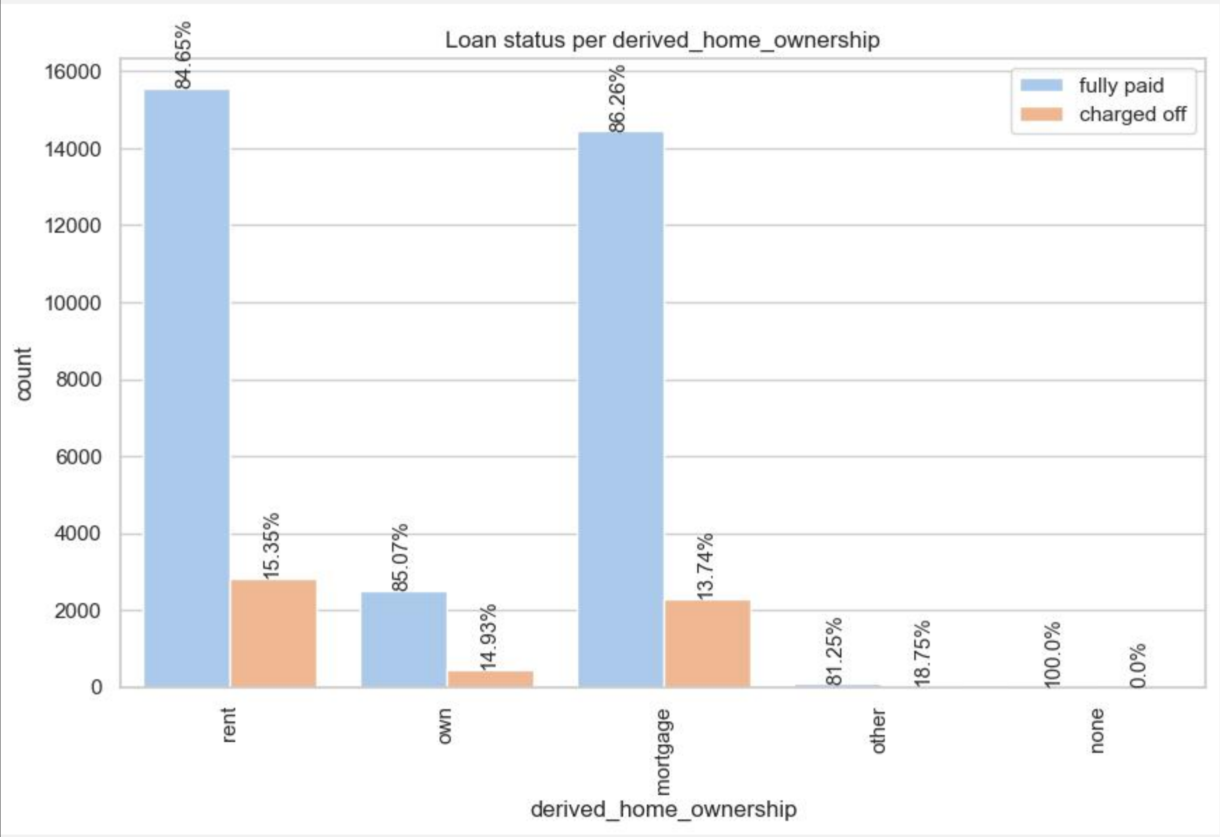


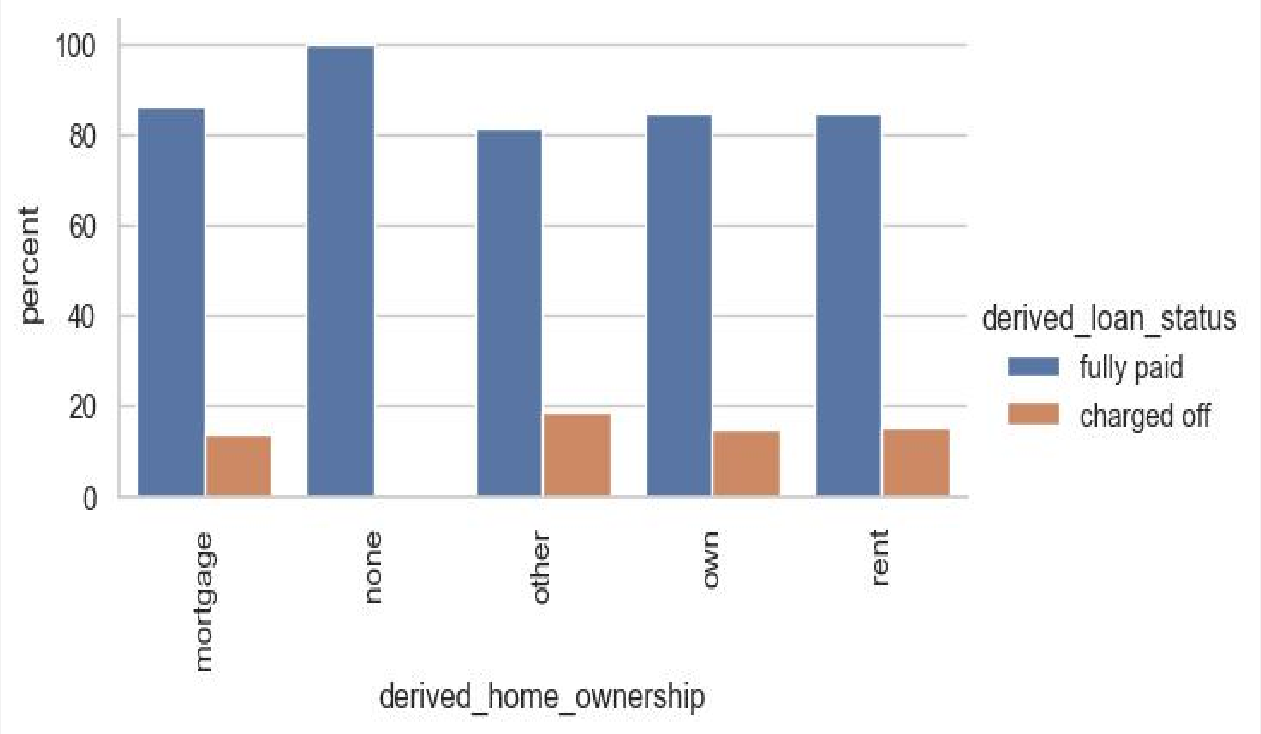


Insights:

- 1Risk of default generally increases as number of inquiries for credit increase

Bivariate Analysis: Impact of home ownership on Loan Status

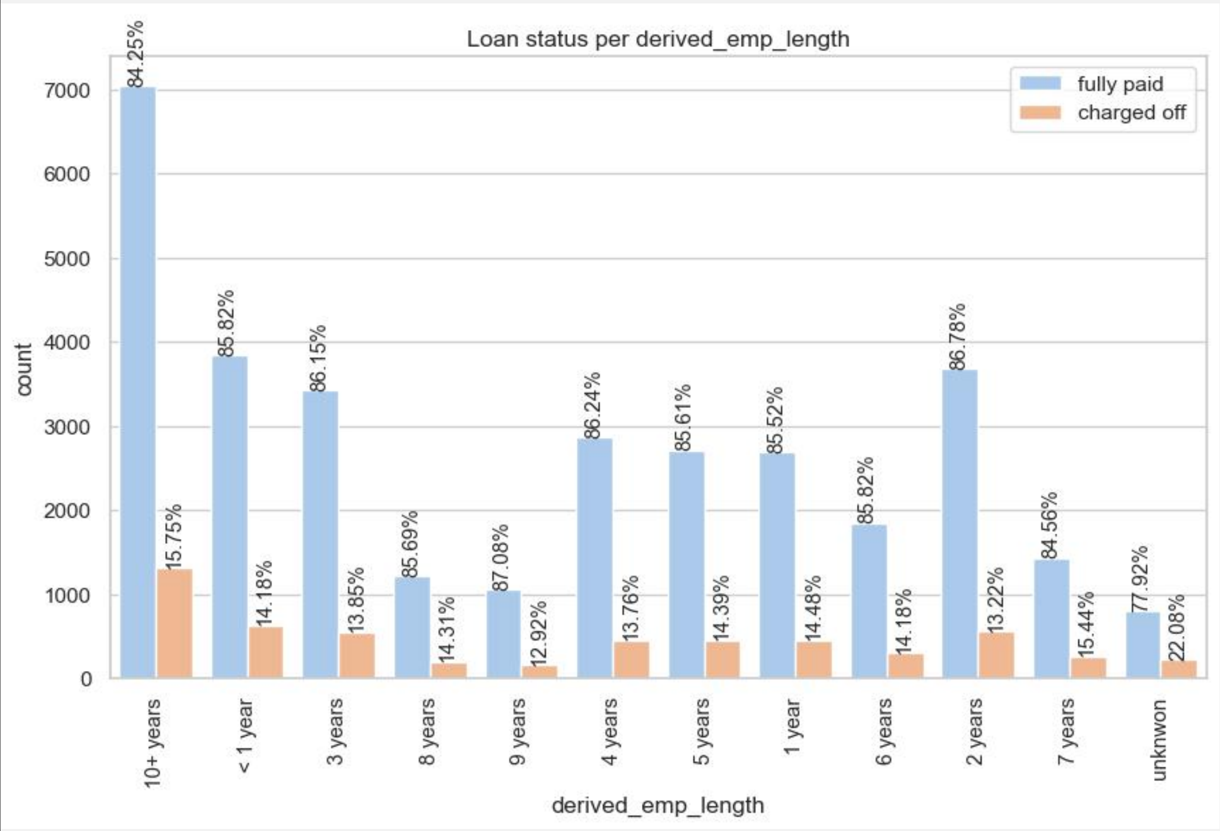


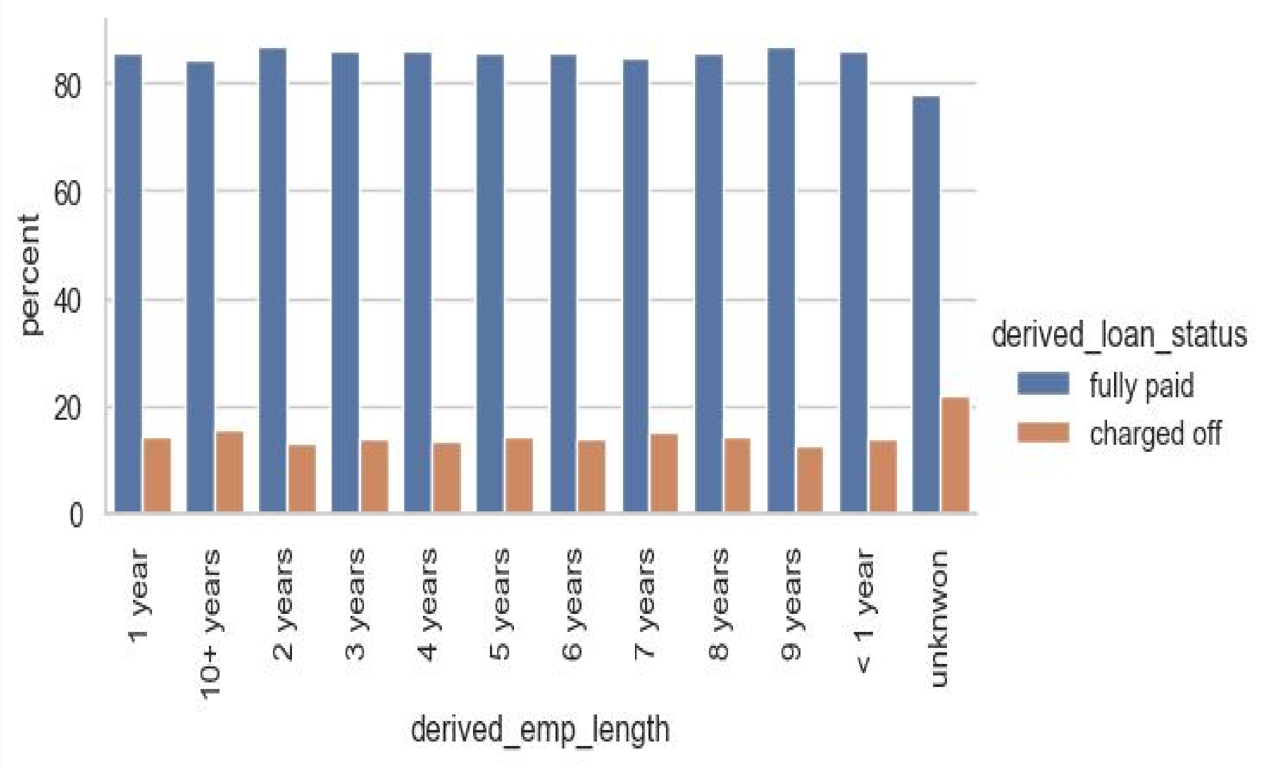


Insights:

- 1. borrowers with Rented house or mortgage, tend to be more frequent type of borrowers
- 2. Frequency wise, higher risk of default lies with rented followed by mortgage
- 3. category wise, 'other' typed borrowers have higher risk of default [19%], followed by
- rent [15 %]
- own [15 %]
- mortgage [13 %]

Bivariate Analysis: Impact of Employment length on Loan Status

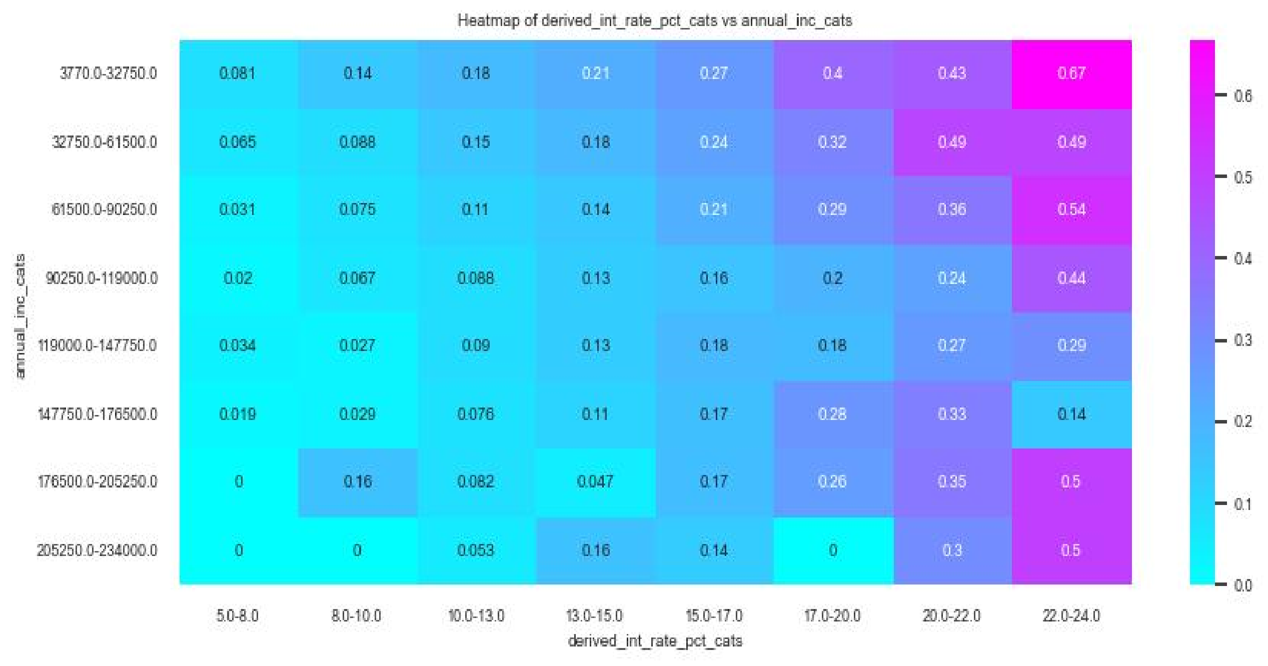


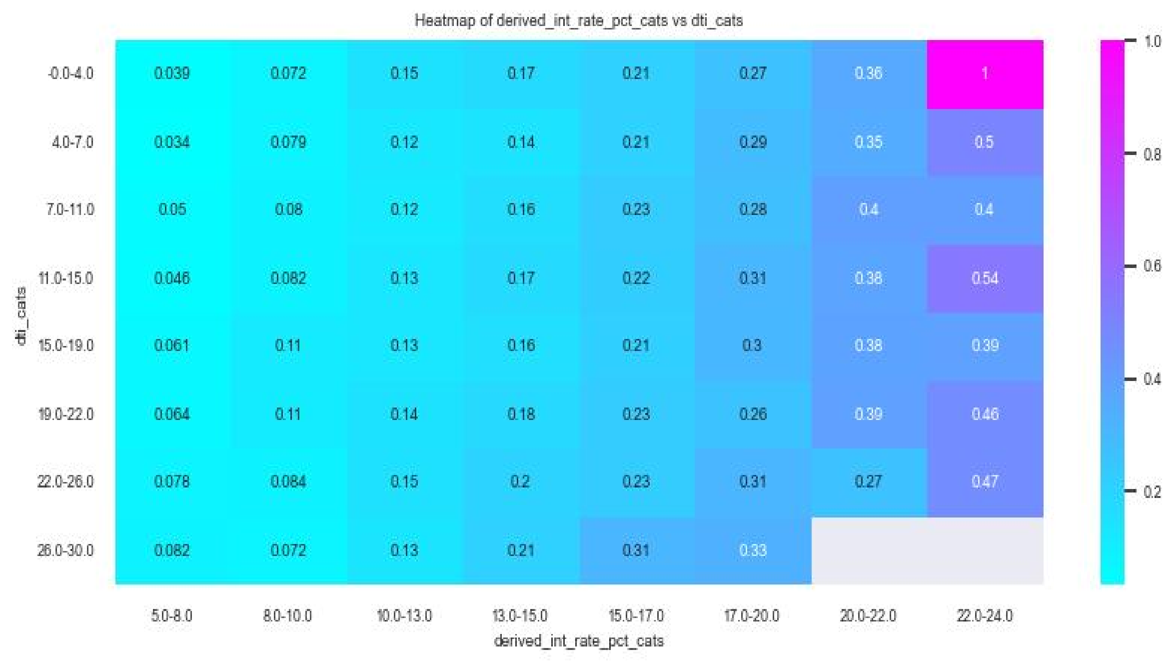


Insights:

- 1. As the number of years of exp increases, freq of credit availed descrease. though freq of 10+ years exp is maximum.
- 2. No specific trend in the default behavior except there is 21% chance of default in case of 10 yrs of exp.
- 3. Unknown emp length has most likelihood of defaults [22%]

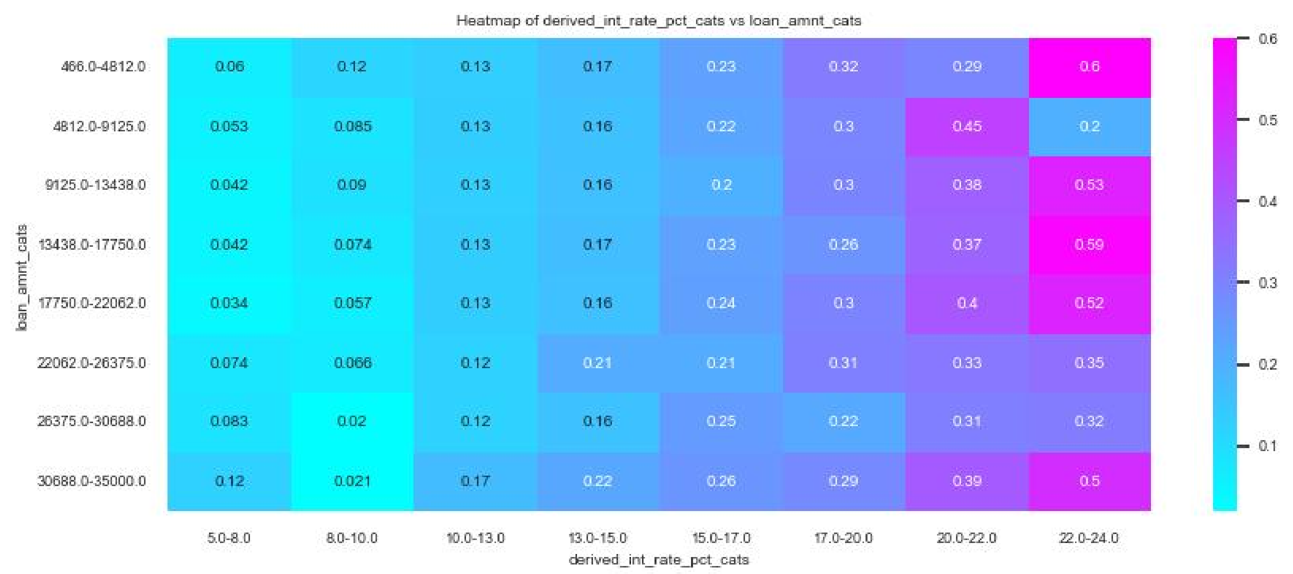
Bivariate Analysis: Comparing interest_rates





Insights:

- 1. higher interest rates at any loan amount are causing defaults
- 2. higher interest rates at any dti slab are causing defaults.
- 3. higher interest rates at any income slab are causing defaults

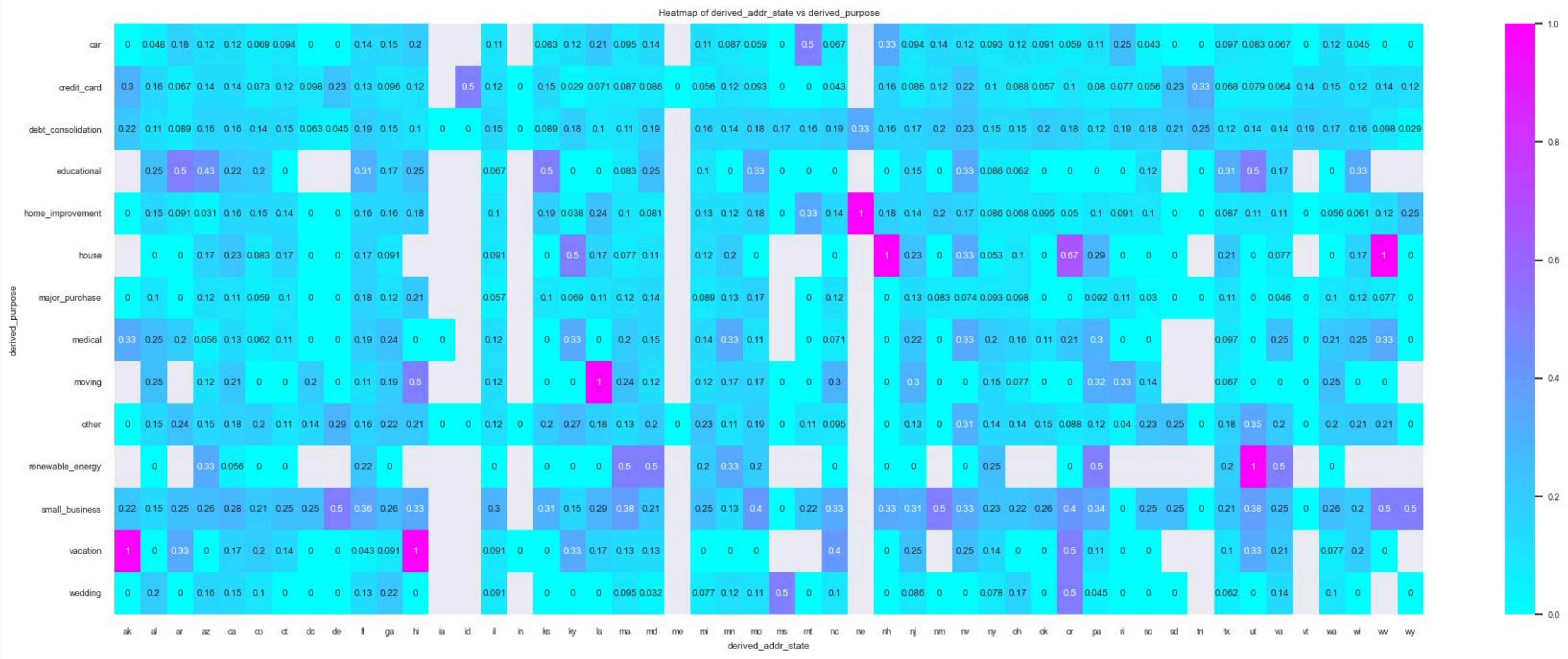


Bivariate Analysis: Comparing DTI

Insights:

- 1. small business tend to have highest rate of defaults across all dti cats
- 2. educational loans tend to have higher rate of default even at the lowest DTI category [0-4] and also
- in [19-26] DTI scores
- 3. moving, other, renewable enrgy, educational, debt consolidation show higher chances of default at higher DTI rates
- lower and middle income borrowers show high default rates across all DTI categories
- At highest int rates even the lowest DTI category is not able to cope up

Bivariate Analysis: Comparing States to Purpose



Insights:

- 1. small business tend to default the loans across all states, followed by other and debt consolidations
- 2. greater than 50 % defaults are observed :
- small business loans in DE, NM, WV, WY
- vacation loans in AK, HI, OR
- renewal energy loan : UT
- wedding loans in MS, OR
- education loans in AR, KS, UT

OBERVATIONS : DEFAULT CASES HAPPEN WHEN

- **Features Related to Borrower' s Credit History & Profile :**
  - *lower* Annual Income [below 90K]
  - *higher* Employment length : Unknown / Missing employment history [22 %]
  - Home ownership *isin* (other, Rent, owned)
  - Address State *isin* (NE Nebraska, NV Nevada, SD South Dacota AK Alaska and FL Florida)
  - *higher* DTI [above 11 %]
  - *higher* Revolving credit util [above 37%]
  - *higher* Credit inquiries in last 6 months [above and equal to 1]
  - *higher* Delinq occurrences in last 2 years [greater than equal to 1]
  - *higher* Number of Public record of bankrupcies [greater than equal to 1]
  - *higher* Number of derogatory public records [1 or 2]

OBERVATIONS: DEFAULT CASES HAPPEN WHEN

- **Features Related to current Loan application :**
  - *higher* Loan amount [> 13k]
  - *higher* Interest Rate [above 13%]
  - *higher* Installment [above 338]
  - *higher* Loan Grade [on and above grade C], same observations through Loan Sub Grades
  - Loan purpose *isin* [Small Business(27%), Renewable Energy(19%), Educational loan(17%)]
  - Verification Status *isin* [Verified (17%), source verified (15%), not verified (13%)]
  - Loan Issue Month *isin* [Dec, May, Sep, Oct)
  - Loan repayment Term *is* [60 months term 25%]
- **Impact analysing Multiple Features**
  - High Loan amount and High Interest rates for Low Annual income borrowers
  - High Installments with Lower Annual Income and Longer Term
  - High Installments with higher Interest Rates for longer Term
  - High DTI individual with Lower income

RECOMMENDATIONS

- Borrowers with **Lower Income** should be provided loans considering
  - Loan amount should be based on interest rates, term, installments, dti and delinq history so that
    - loan amount should not be too high [possible loss of business but Risky]
    - OR the installments are payable by the borrower
  - should be reviewd for DTI, delinq history, revolving util and inquiries for credit on a regular basis
- Different **Loan purposes** have different behavior and they should be evaluated based on
  - Annual Income, Credit history (DTI, revol_util, delinq, credit_inquiries and public records), Borrower specific information (Borrowers State, employment length), Loan Grade

RECOMMENDATIONS

- **Verifications** ie both 'Verification' & 'Source Verification' need some more improvements as the default rate is still high.
  - Considering that verifications may have done in the cases of risky loans and those cases may have been addressed for the involved Risk, either by increasing the interest rates / reduing loan amount / adjusting the term, these adjustments needs a review whether the risk is getting mitigated. As we still see the defaults in these scenarios.

- **High Risk loan** can be reduced on by
  - Reducing loan amount / adjusting the installment, interest rate and term to compensate the risk
  - asking for co-borrowers / collateral
  - getting Investors / Credit Default Swaps

THANK YOU