



Lending Club – Case Study (Lending Risk Analysis)

TEAM NOMADS

Santosh A P
Saman T
Afroz A K
Vivek D

Contents

Business Understanding

Executive Summary

Top Finding and Recommendations

Key Assumptions and Scope

Data Sourcing & Cleaning

Data Analysis and Comprehensive Visuals

Business Understanding

Business Understanding

Lending Club - 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. Data about the Loan application is maintained by the company.

Lending loans to 'risky' applicants could result in credit loss for the company. The idea of this exercise to determine the main drivers for charged off(defaulted loans). These drivers can then be used to workout the risk mitigation actions , thus reducing the credit loss for the company in longer run, while sustaining the business.

Goal

- Mitigate high risk lending
- Reduce Credit Loss for company

Objective

- Analysis of Charged Off Loans
- Comparative analysis of defaulter Vs good borrowers
- Pattern and behavioral analysis of borrowers

Tactic

- Exploratory data analysis of loan borrowers

Executive Summary

Executive Summary

Key Drivers behind Charged Off loans

- Unsecure loans - small businesses/Educational Loans
- High Debt Ratio (dti) and high utilization of Revolving Balance
- High interest rate and Installment
- High ratio of open / total accounts
- Funded amount < Loan amount

- High default from geographical area Nevada State
- Unknown(non-rental & not own house) home ownership
- Increasing trend of defaulters(2007 onwards). Highest in 2011

Employment

- Non Disclosure of Employers Detail
- Non Disclosure of Employment Length

Credit Numbers

Credit History

- Age of Credit History
- High loan inquiries
- Less financially active
- High public records including bankruptcies

Others

Income

- Low Income
- Income verification status

Top Findings and Recommendations

- **Funded Amount < Customer applied amount** are marginally likely to **Charge Off**.
- Customers who have **longer commitments opted charge off** due to varying financial circumstances over the time.
- **High Interest Loans are Risky** - due diligence in vetting the application is required. An applicant with High DTI is risky.
- Loans with **high DTI and higher interest levied are more likely to default**. Also, the verification status has no bearing in this scenario. The vetting and interest rates will need to be reviewed.
- **Too high or too Low Credit History increases the Risk of application**. Too High credit History – indicates higher applicant age. Low Credit History meant low visibility of the financial profile.
- **Home ownership with OTHER** has **high bearing on charged off loans**. This category will need to be corrected to more apt one. Additional vetting in such cases is recommended.
- **Staunch indicators of defaults** - High derogatory comments, public bankruptcies, high credit revolving balance utilization, high open accounts on credit file.
- In geographical terms, suburbs from states are more likely to default. **Nevada has highest recorded default percentages**.
- The level of disclosure of employee name and employment length has impact on defaults. **Default rate will increase if the details are not disclosed**. Again, employment vetting process will need to be reviewed.

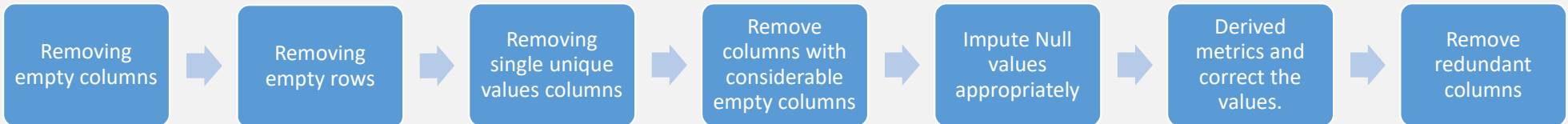
Key Assumptions and Scope of Analysis

- The observations considered for the analysis are loan applications **during the period 2007 - 2011**.
- The observations are the loans that are accepted by LC and eventually either **fully paid or charged off**.
- The observations for the **currently active loans are not considered** in the analysis
- There are two variables of interest in the data set – state code and zip code – As entered by the application while applying for the loan. There are some inconsistencies in these variables. It is assumed, that the **zip code entered is correct and state code is derived** from the zip code for correcting the values.
- **Deep Text Analytics** has been considered **out of scope** of this EDA.
- The **'purpose'** as entered by the customer has been **fine lined and mapped**, before analyzing, using the 'description' that customer has entered. It is assumed that the description entered is more genuine than the purpose. For example, loan with purpose 'Wedding' but with a description of *'Borrower added on 08/26/11 > I am in the process of purchasing an engagement ring.
 Borrower added on 08/26/11 > I am also going to consolidate \$15,000 of my credit card debt.
 Borrower added on 09/08/11 > After all my fixed expenses including the \$600 loan payment, I will have approximately \$650 a month for miscellaneous expenses and savings.
*

NOTE: The highlighted statement implies, that the loan is used for debit consolidation rather than the actual purpose stated.

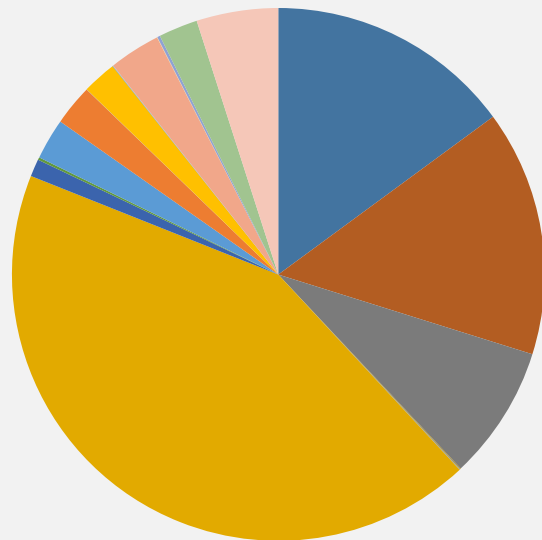
Data Sourcing & Cleaning

Data Sourcing and Cleaning



Derived Columns

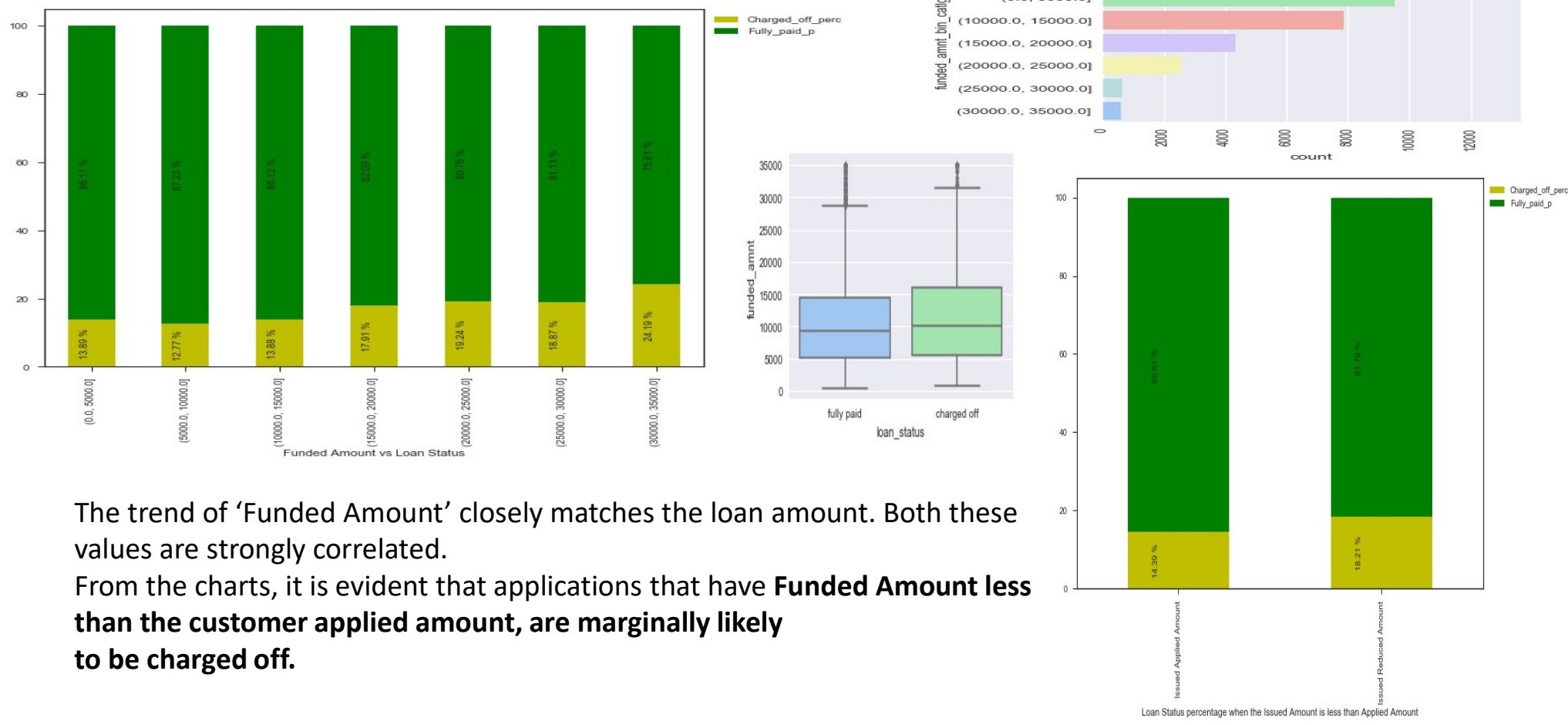
- annual_inc
- annual_inc_log2
- dti
- emp_length
- emp_length_disclosed
- emp_title_provided
- home_ownership
- income_verification
- inq_last_6mths
- installment
- int_rate_f
- issue_date_dt
- loan_amnt



Column Name	Column	Column Description	Missing	Unique counts
annual_inc	float64	Annual income provided by the borrower during registration.	0	5210
annual_inc_log2	float64	Derived Column. Log2 of the annual income	0	5210
dti	float64	Ratio of Total debt (excluding Auto and Mortgage debt) to the total monthly income	0	2853
emp_length	object	Length of employment of the customer	0	12
emp_length_disclosed	object	Derived column. Holds custom bool value for employee length disclosure		2
emp_title_provided	object	Derived column. Holds the values of employee disclosure.	0	2
home_ownership	object	Home ownership status as entered by customer	0	4
income_verification	object	Derived column. Contains Income verification	0	2
inq_last_6mths	int64	The number of times, the customer has inquired for loans	0	3
installment	float64	The installment per month, customer has to pay for the loan	0	14399
int_rate_f	float64	Derived column. Holds interest rate	0	370
issue_date_dt	datetime64[ns]	The date when the loan was issued by LC	0	55
loan_amnt	int64	The amount for which the loan was applied	0	863
funded_amnt	int64	The amount that was funded at the time loan was issued	0	863
loan_status	object	Status of loan. Will be either fully paid or Charged off	0	2
open_acct_ratio			0	723
pub_rec	int64	Number of public derogatory records on customer credit file	0	5
pub_rec_bankruptcies	float64	Number of public bankruptcies recorded	0	4
purpose_mapped	object	Derived column. Holds the purpose for which the loan was applied.	0	10
revol_util	float64	Percentage utilization of the revolving balance on the credit accounts.	0	1088
source_verification	object	Status of source verification	0	2
delinq_2yrs	int64	Number of 30+ days past-due incidences of delinquency in the borrower's credit file	0	12
state_code	object	Validated state code of loan applicant	0	55
zip_code	object	Zip Code of the US state	0	831
term	int32	Tenure of the loan account	0	2
time_delta_int	int64	Derived column. Credit History of the applicant. It is the difference of the earliest credit line of the customer and date when the LC loan was issued.	0	1730

Data Analysis/Comprehensive Visualizations

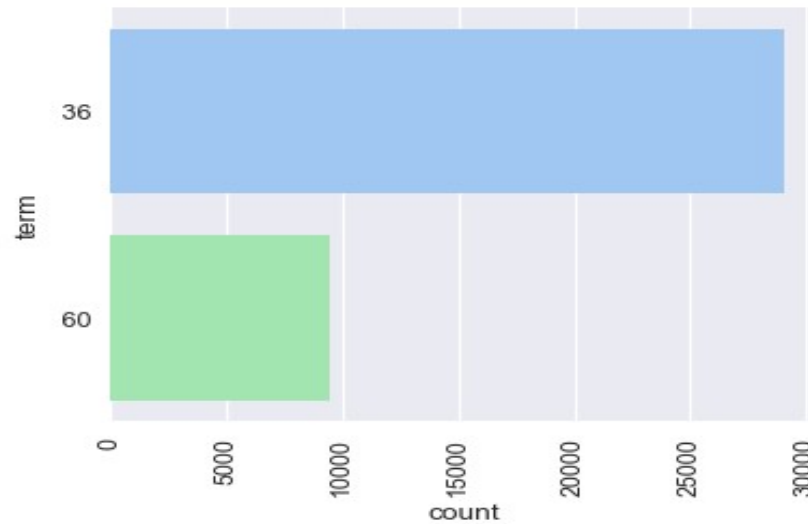
Funded Amount vs Loan Status



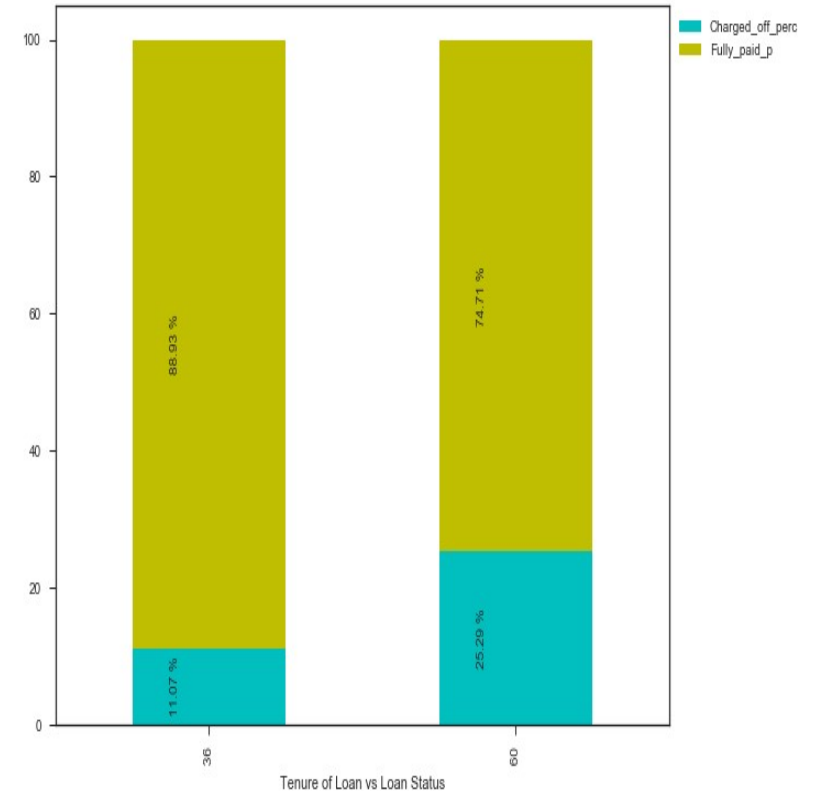
The trend of 'Funded Amount' closely matches the loan amount. Both these values are strongly correlated.

From the charts, it is evident that applications that have **Funded Amount less than the customer applied amount, are marginally likely to be charged off.**

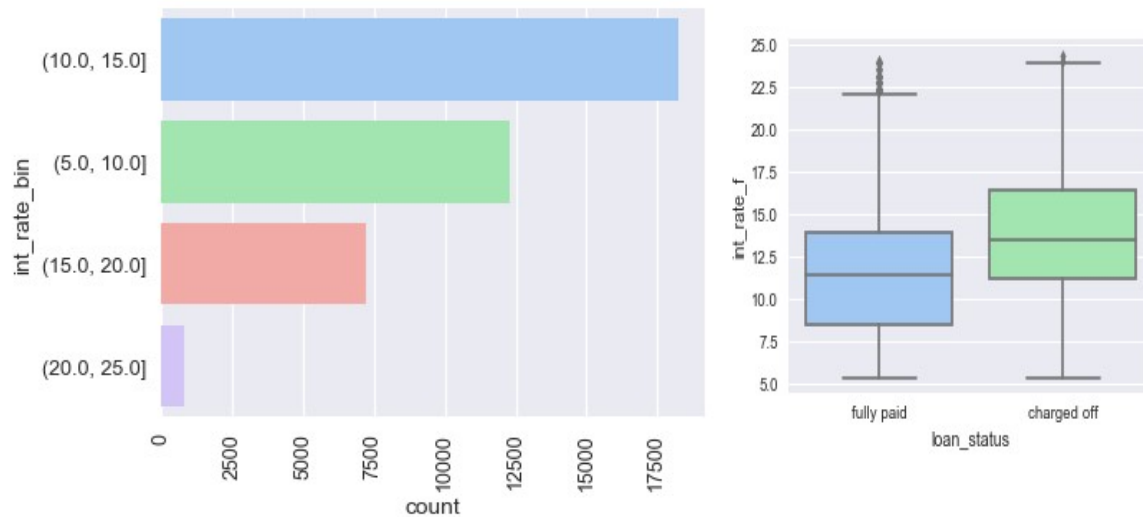
Term - Tenure of the loan



Loans with longer tenure have high percentage of Charge Off applications. This indicates that **customers who have longer commitments opted charge off** due to varying financial circumstances over the time.



Interest Rate of the loan

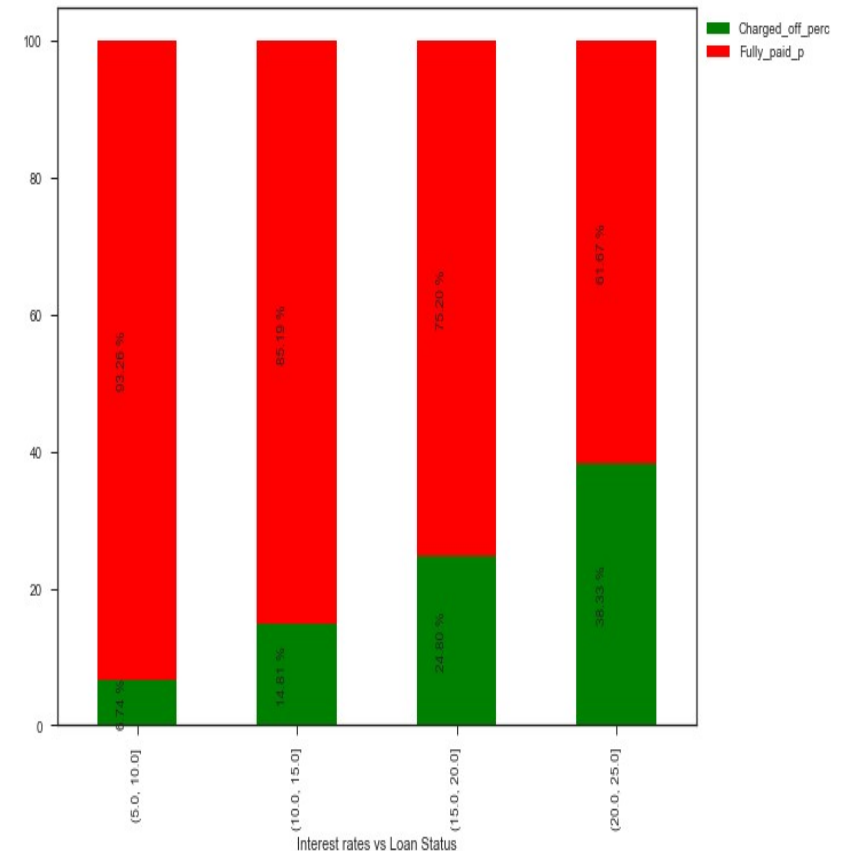


In general, the higher the interest rate is, the higher is the percentage of Charged off Loans.

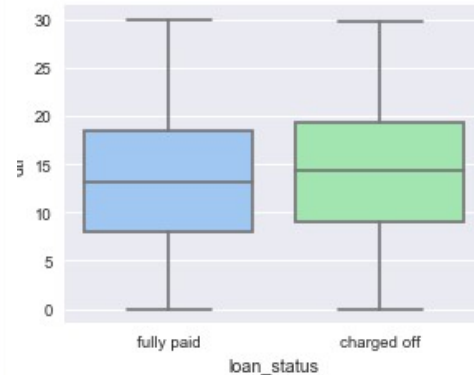
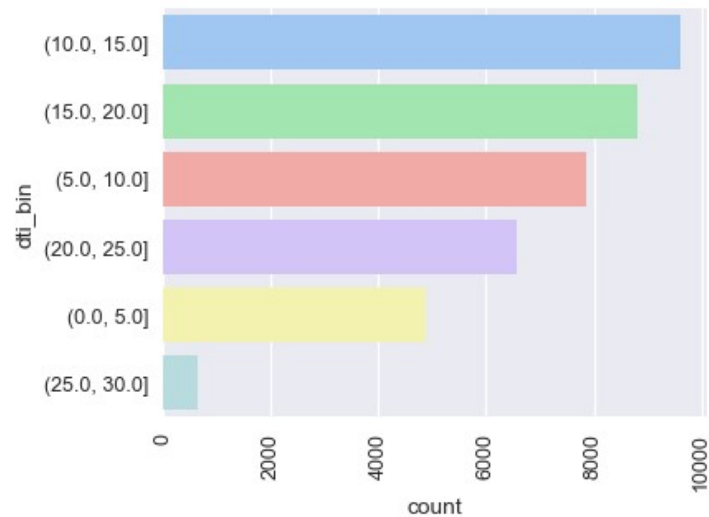
Loans with in interest rate 10 -15 are most frequent, followed by those with 5-10 and 15-20.

However, the percentage of charged Offs is high in 10-15, followed by 5-10.

This analysis evidences, that there is **some risk in issuing the High Interest Loans** and additional due diligence is required while processing loan applications



DTI - Debt to Income Ratio

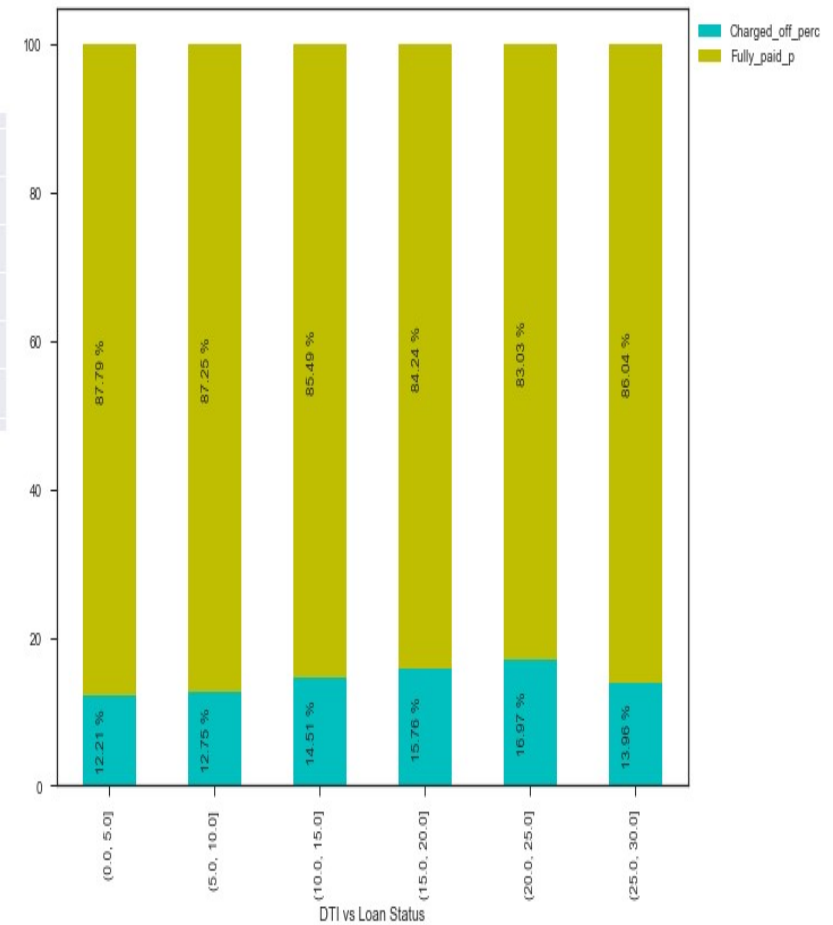


The percentage of charged off loans is increasing as the 'dti' increases.

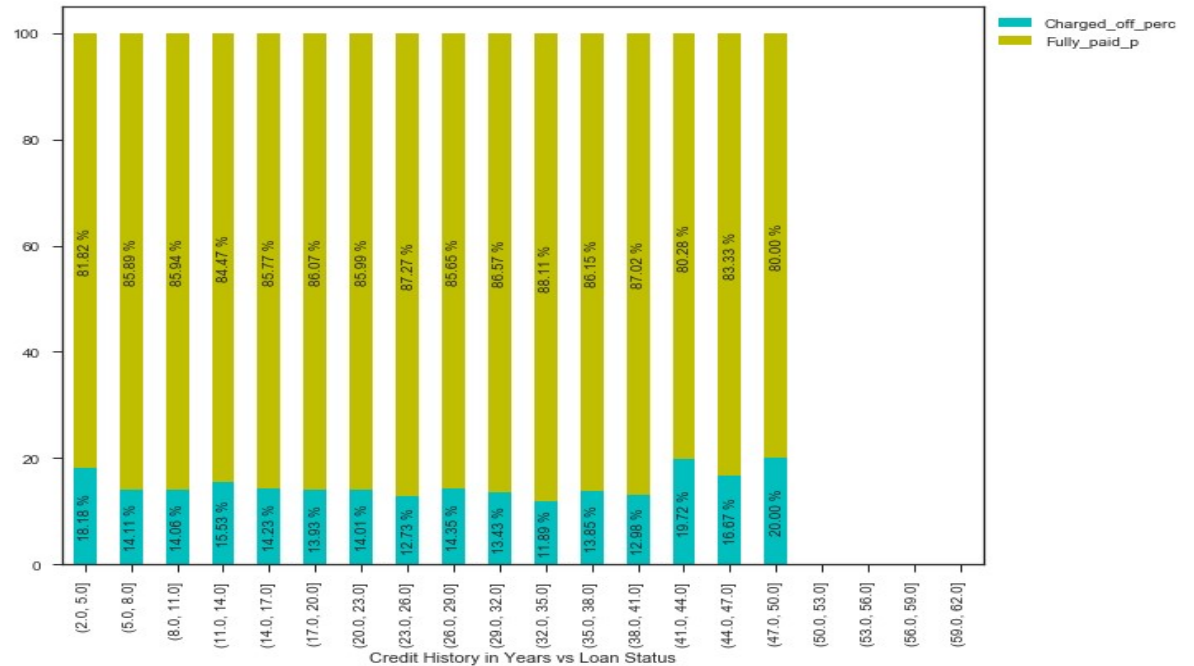
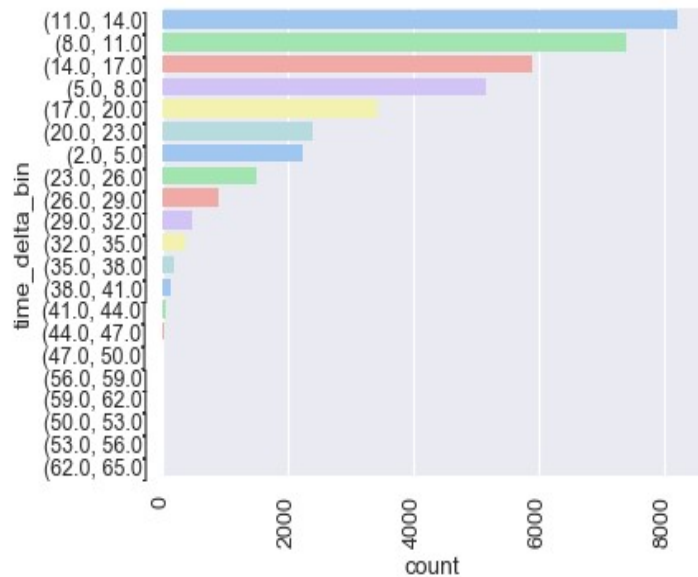
Also, from box plot, it is evident, that 'dti' is higher in charged off loans.

While 'dti' in 10-15 is frequent observation, the percentage of charge offs are higher in 15-20.

This evidences, that **high 'dti' meant a risky loan application.**



Credit History



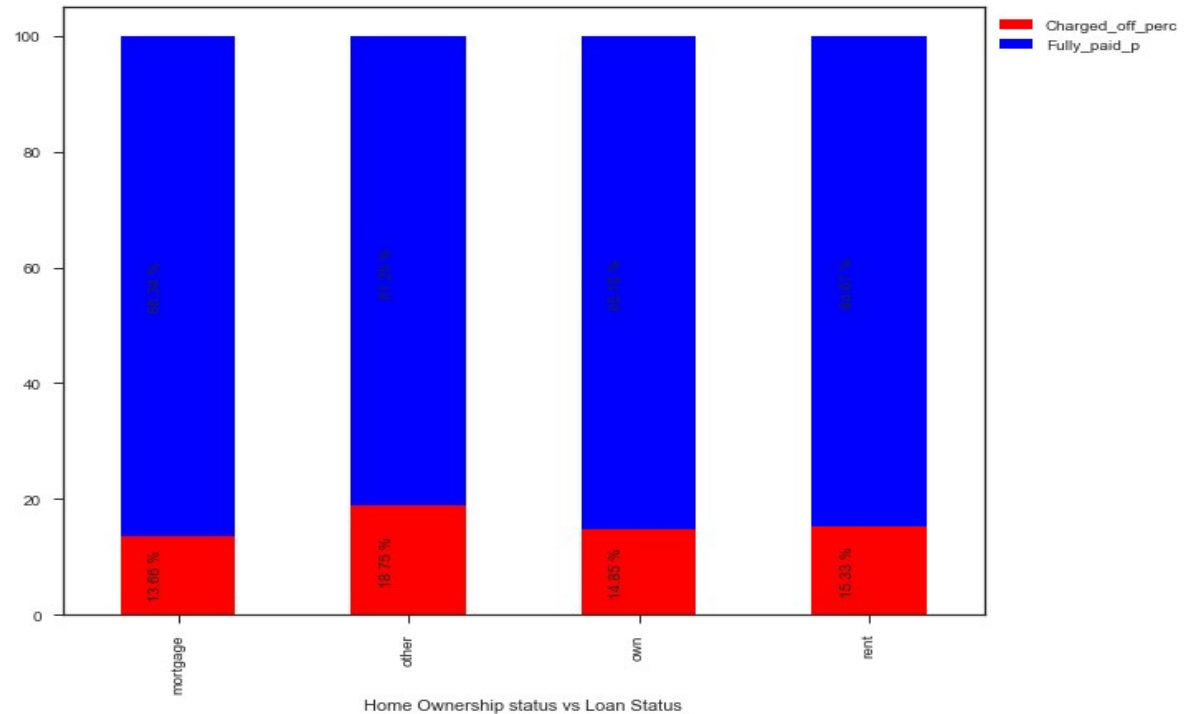
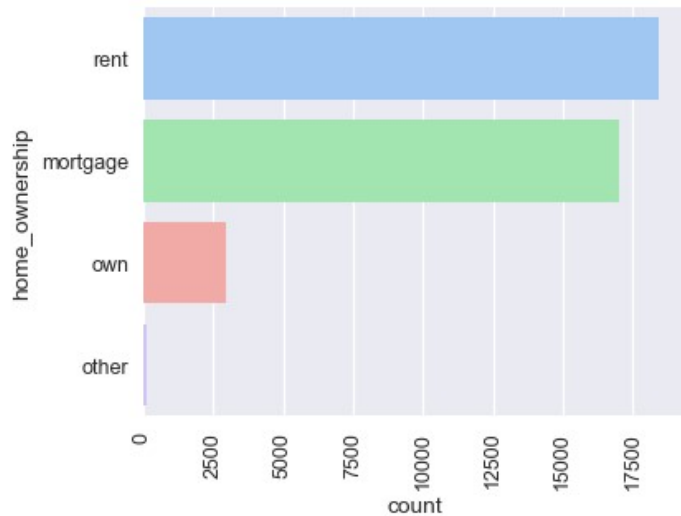
There is a mixed trend in the charged off loans when plotted against the credit age.

In general, a very **low credit age meant, the financial profile of the applicant is not clear.**

A high credit age meant, the financial profile of the applicant is superseded by the age of the application.

This trend is clearly visible in the above percentage stacked plot.

Home Owner Ship

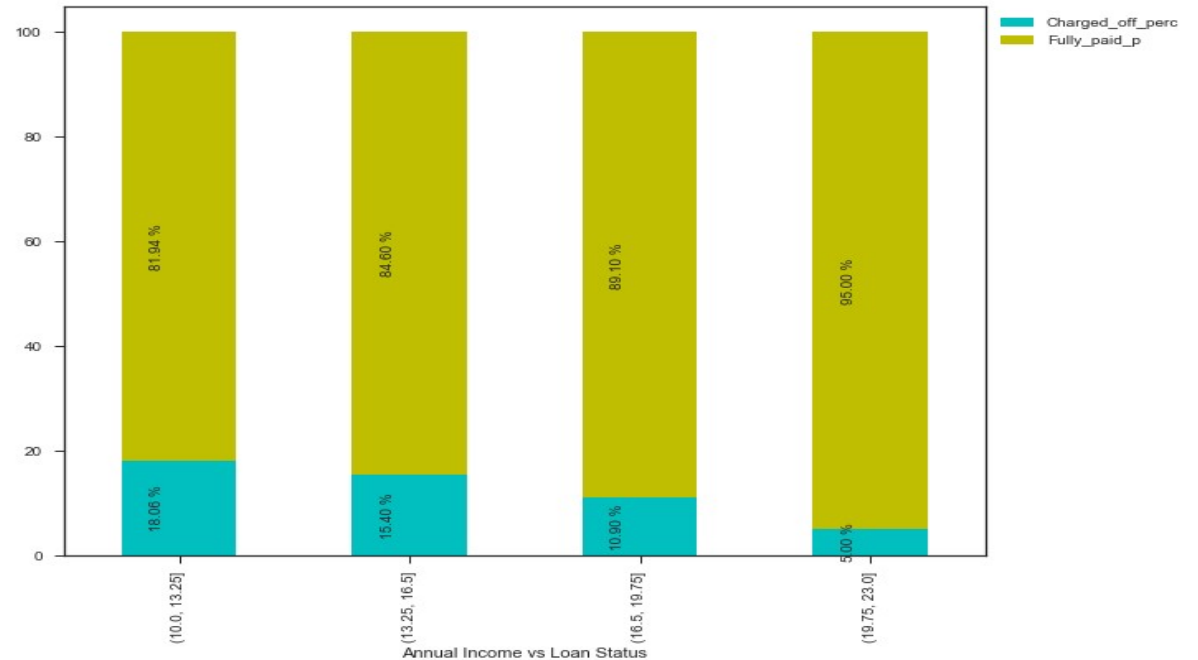
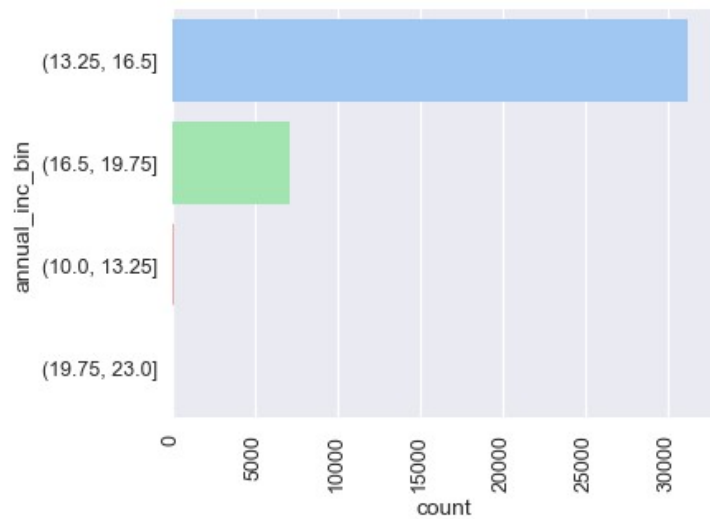


The home ownership with 'other' are charging off the loans at higher rate.

This shows that 'other' category of the home ownership should be identified as risk. Further checks are required to understand the residency status.

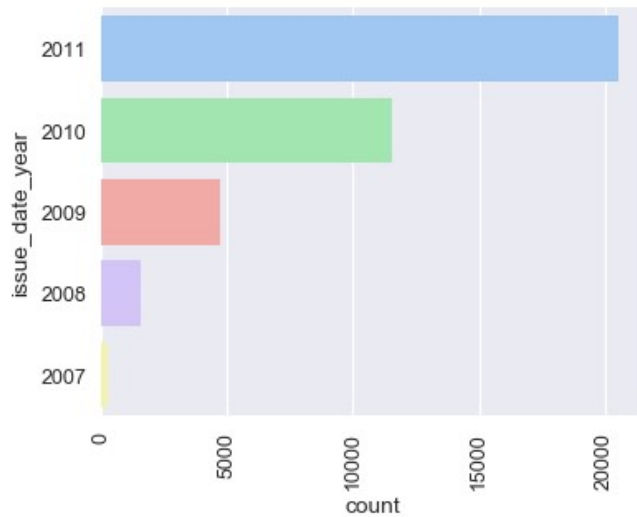
Home ownership with **Rent** has slightly higher charge off rate

Annual Income

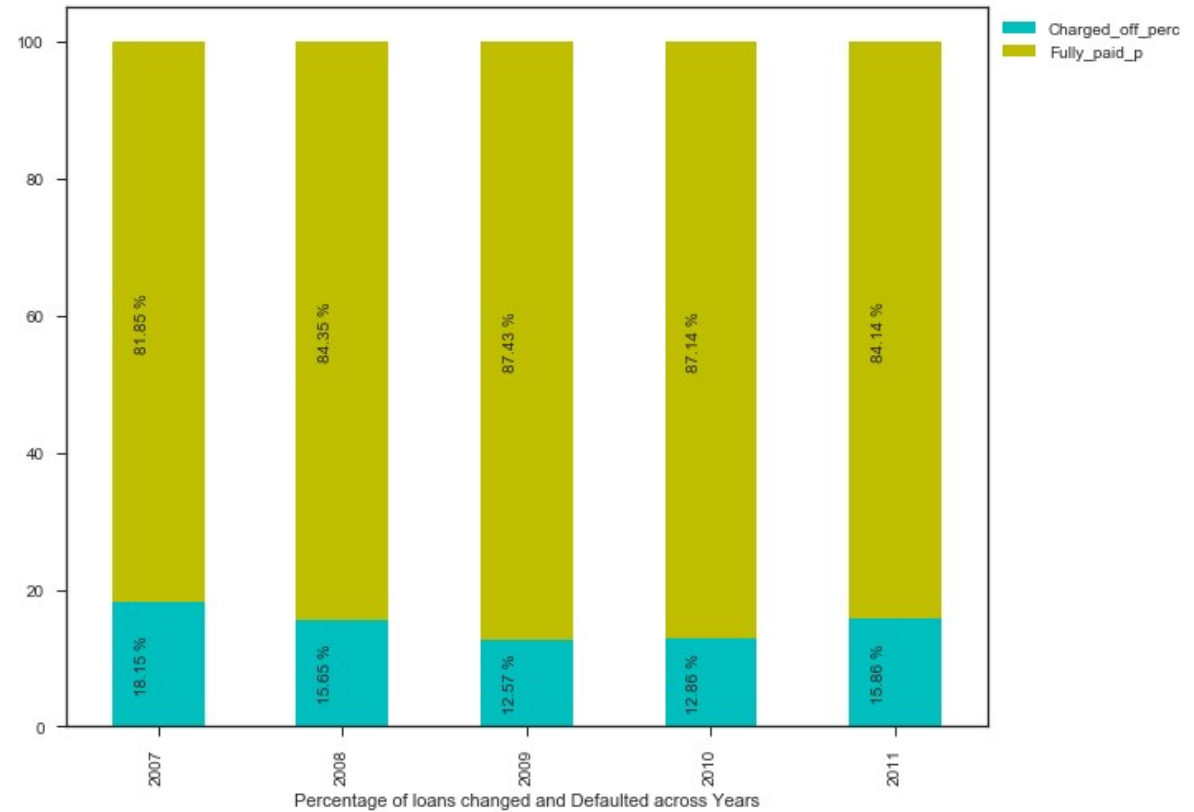


The general trend is, as the **annual income increases**, the **percentage of charged off loans decreases**. This is evident from the fact that loans in the range 16.5 - 19.75 have lower charged off percentage compared to 13.25 - 16.5 which is highly frequent.

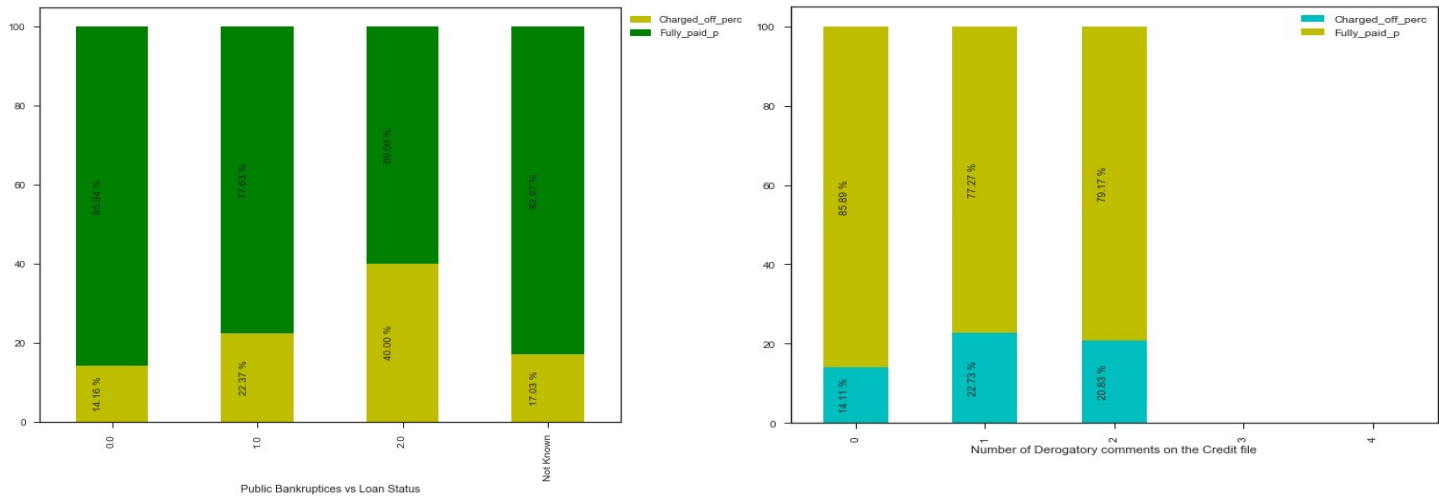
Year of Issue



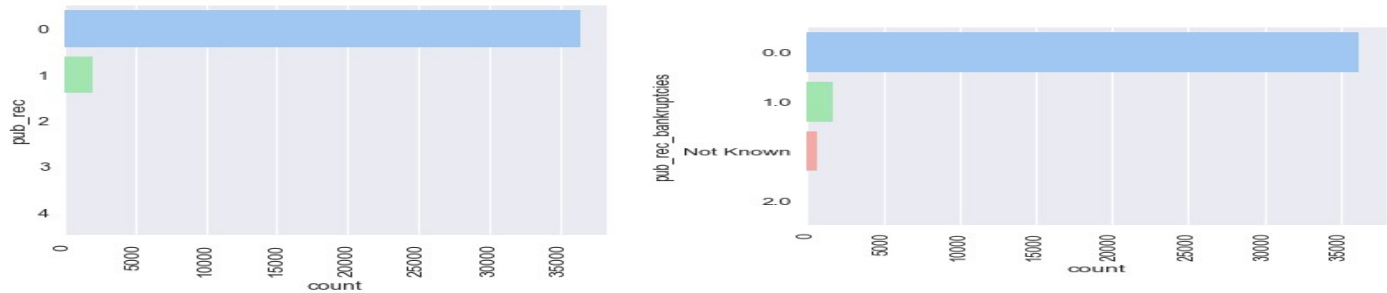
The number of loans gradually increased from 2007 till 2011. However, the percentage of the charge off loans issued have reduced from 2007 till 2010. In 2011, there is high charge off rate. This could be partly due to number of loans issued. Further analysis will be required to draw any conclusions on this behavior.



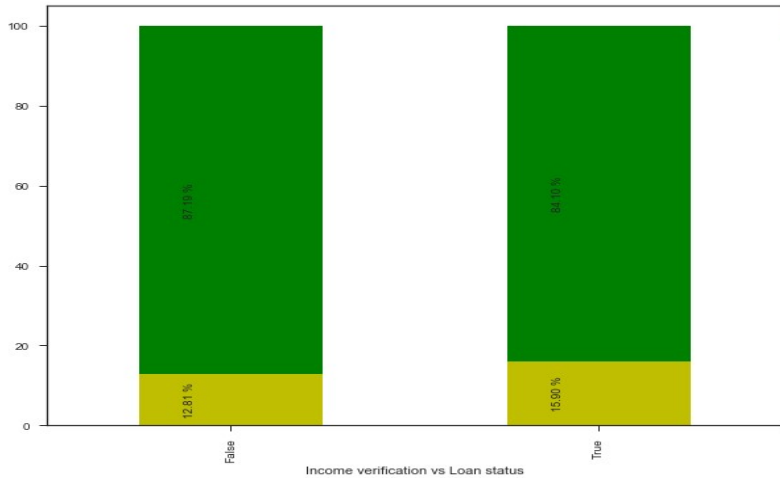
Derogatory comments and Bankruptcies



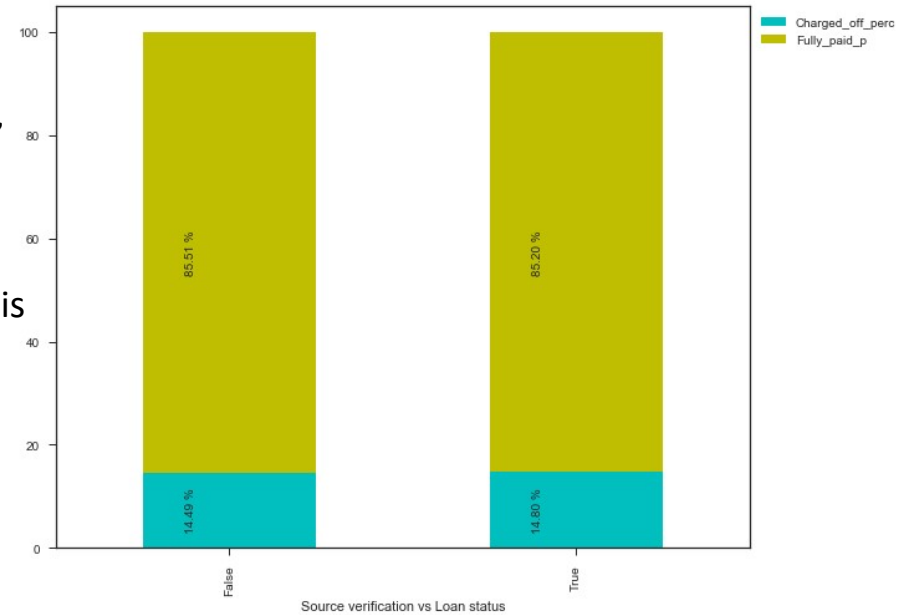
It is evident from the above charts that, **the high derogatory and bankruptcies are, the higher is the charge off rate.**



Verification Status

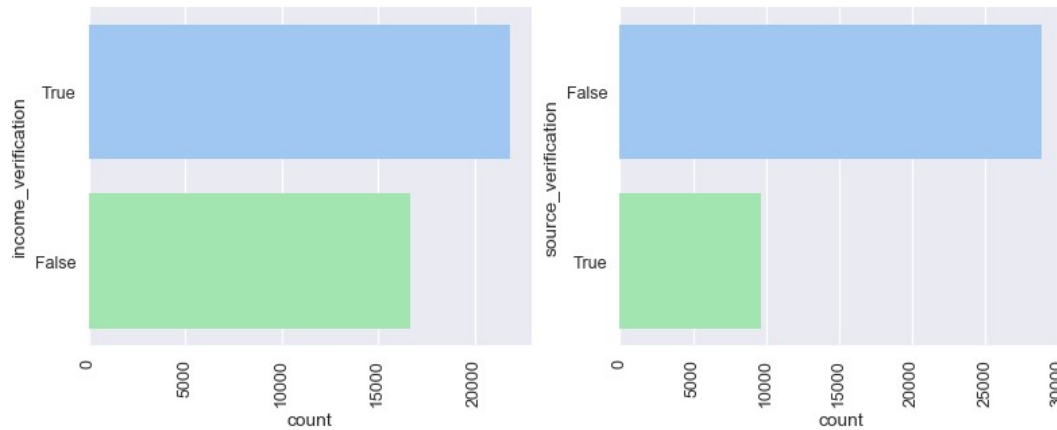


Source Verification, apparently has no effect on the charged off rate, even though there is stark difference in the counts.

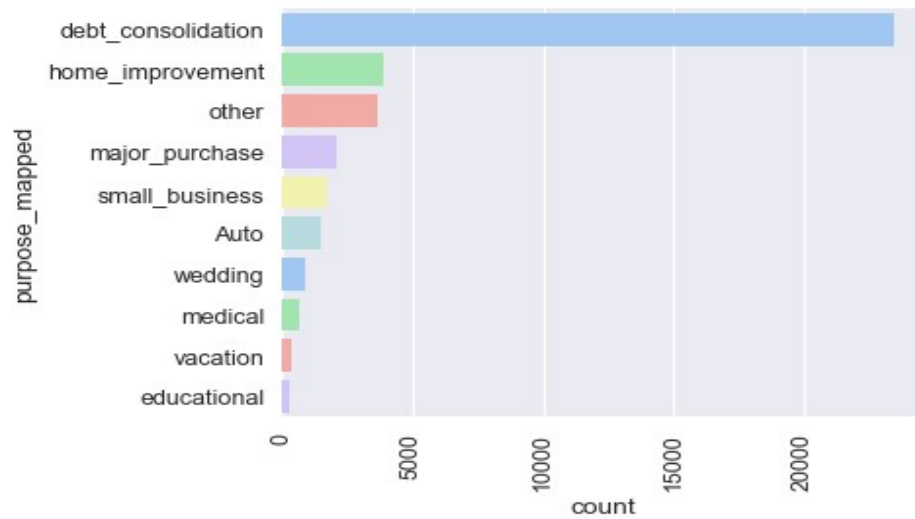


In case of **Income verification**, the charge off rate is higher ,when the income has been **verified**.

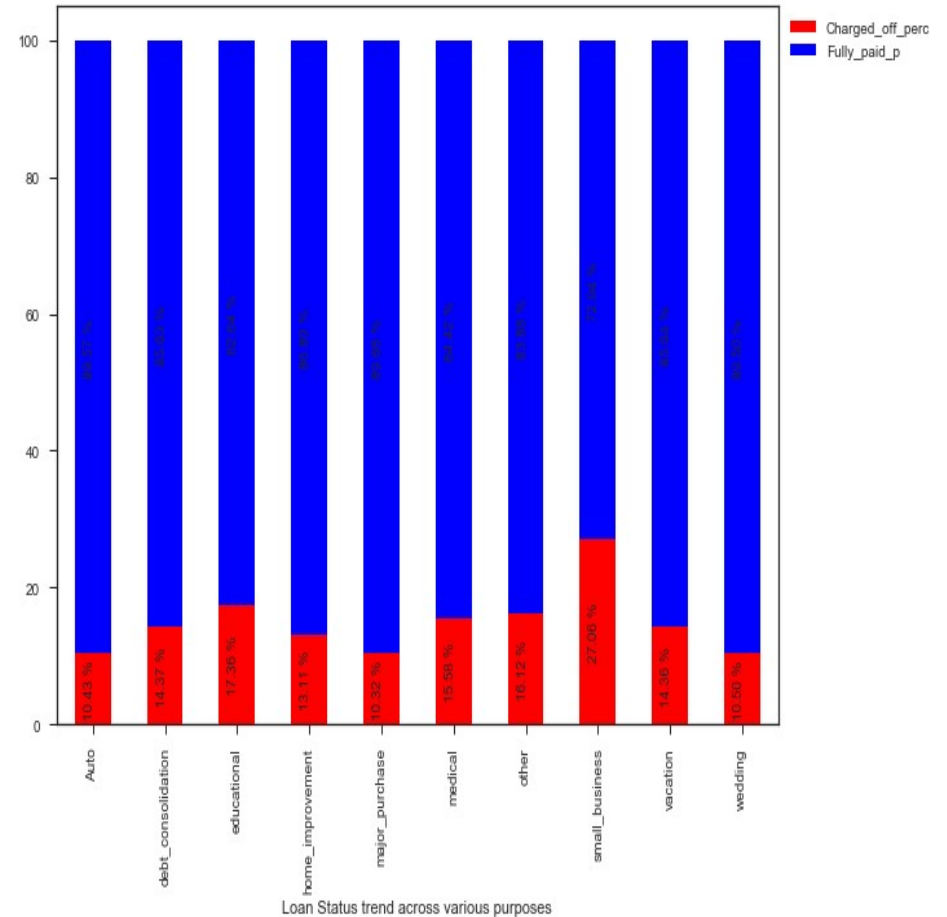
This shows, that income verification checks have no effect in case of Risky applications. The income verification process will need to be reviewed.



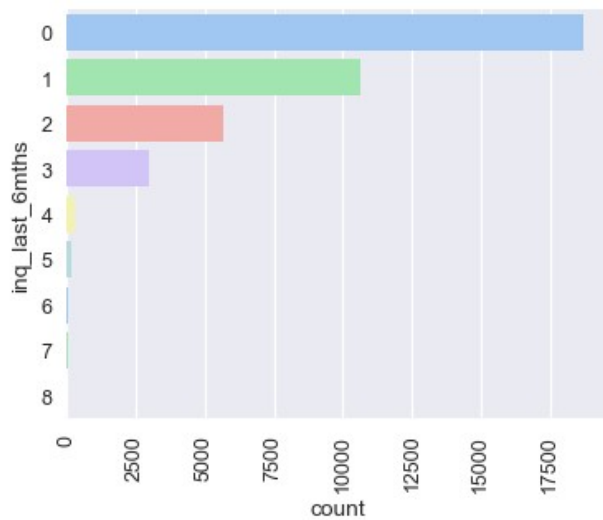
Purpose



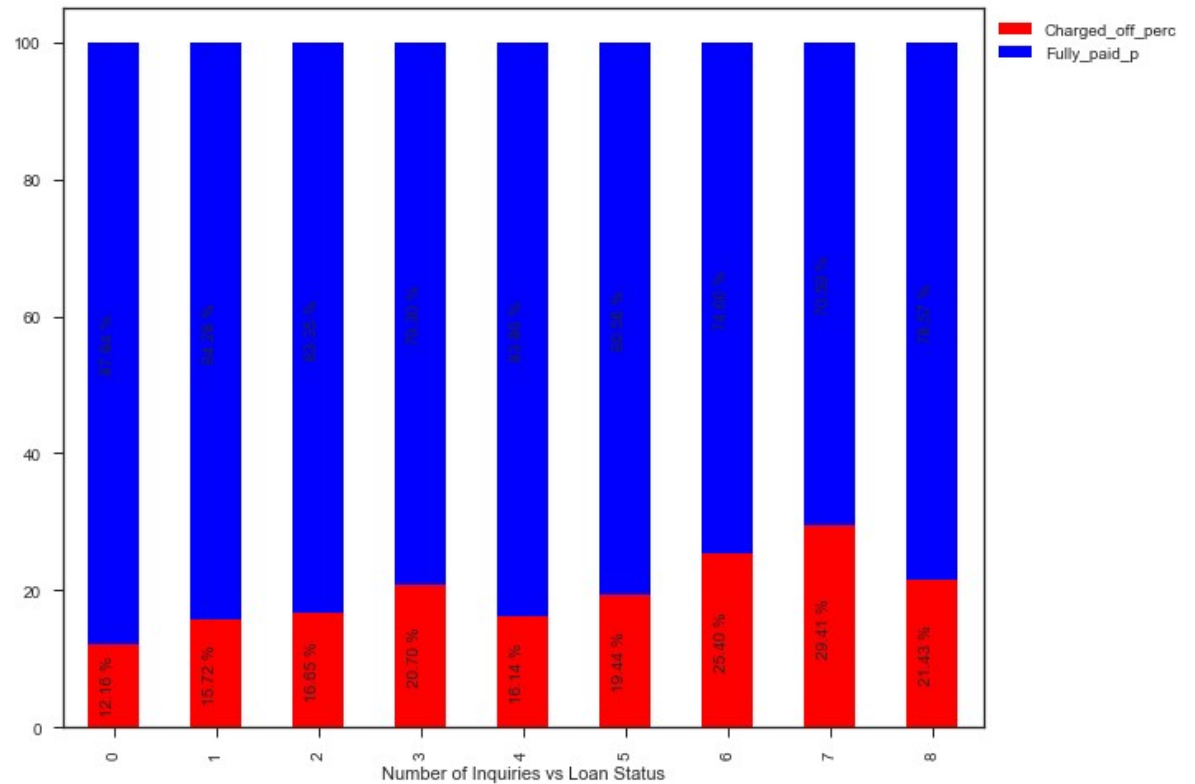
In general, the unsecured loans - like **the small business loans**, followed by **educational loans** are highly charged off. **Debit Consolidation** is the **most common purpose**



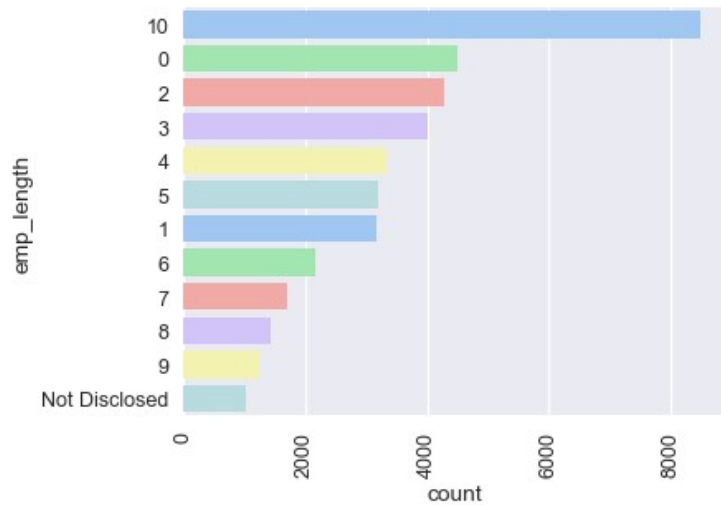
Number of Inquiries



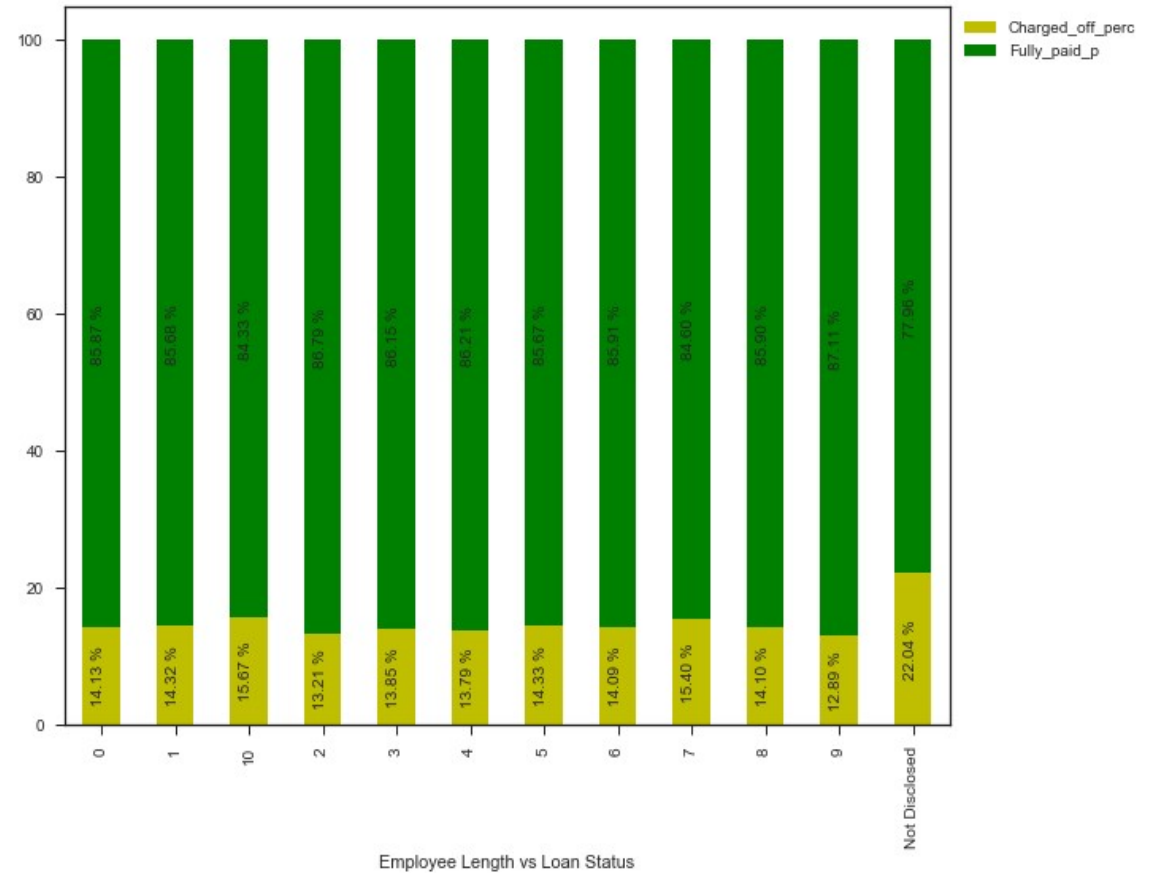
Considering both the count plot and stacked percentage plot, it is evident that, **the higher the number of inquiries, the higher is the rate of charge off.** This parameter shows the desperation of the customer for a credit line.



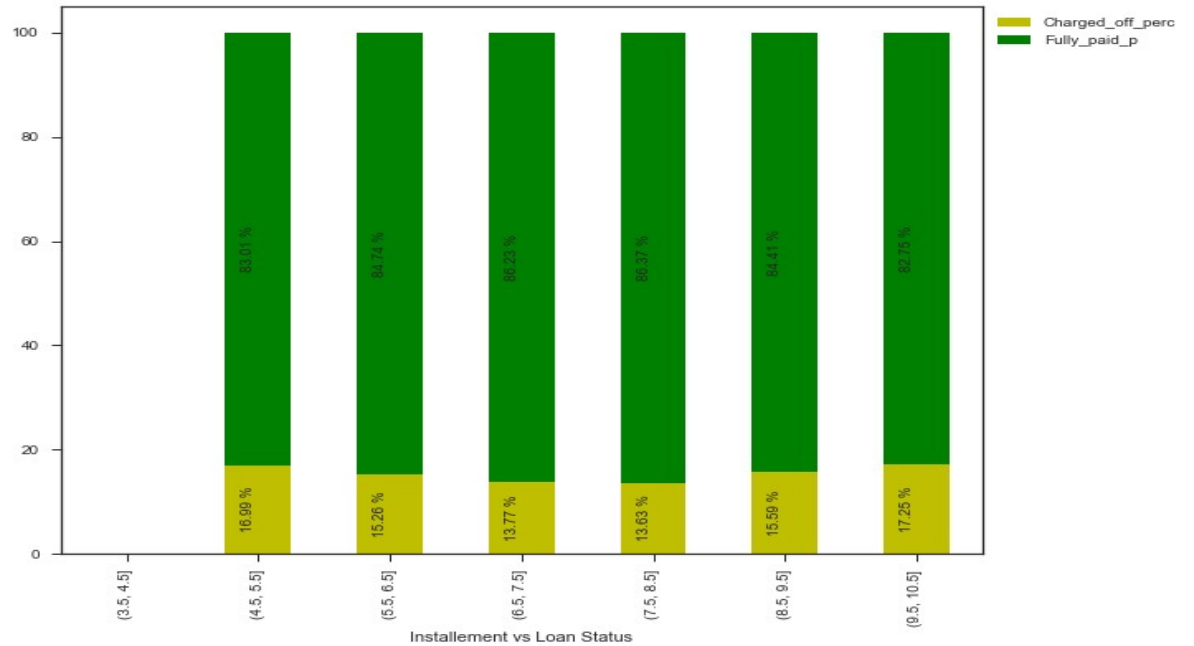
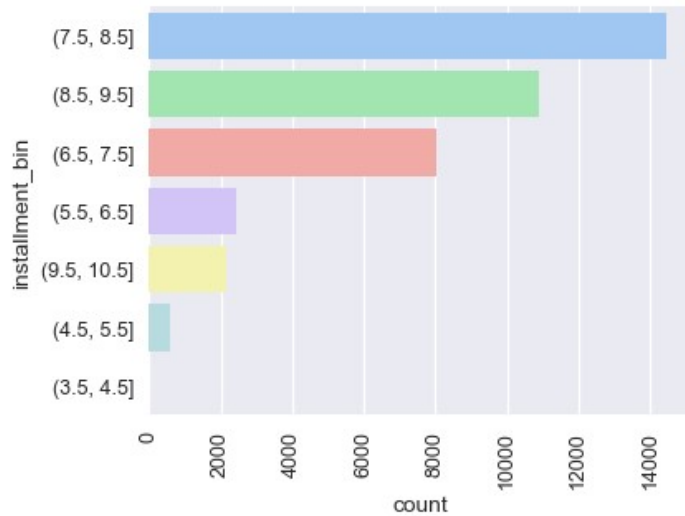
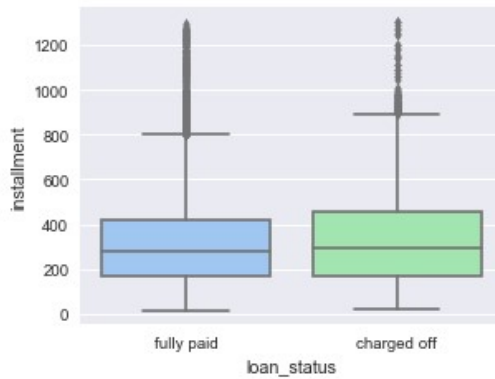
Employment Length



There is no clear trend of **employment length impact on the charged off rate**. Further investigation will be required to study this variable



Installment

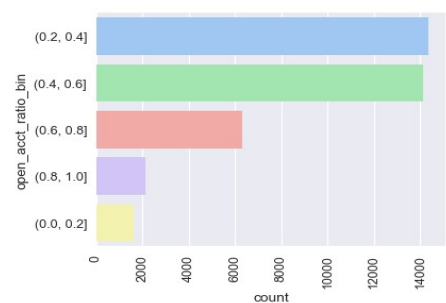


From both the count plot and stacked plot it is evident that **as the installment increases, the rate of charge off loans increases.**

Loans with installments in range 7.5 - 8.5 are common followed by 8.5 - 9.5 and 6.5 - 7.5.

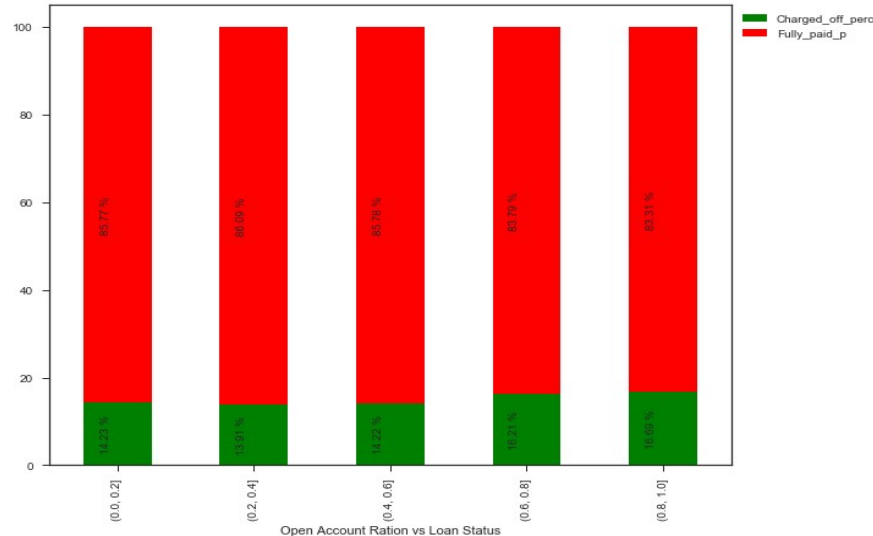
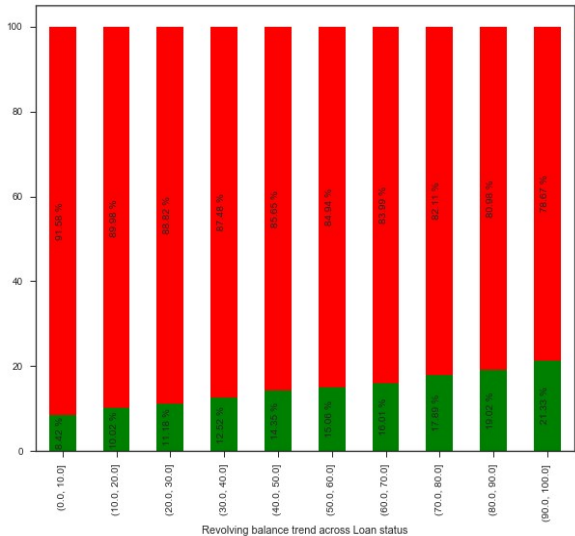
Rate of charge off is marginally higher in 8.5 - 9.5

Open Accounts and Revolving Balance

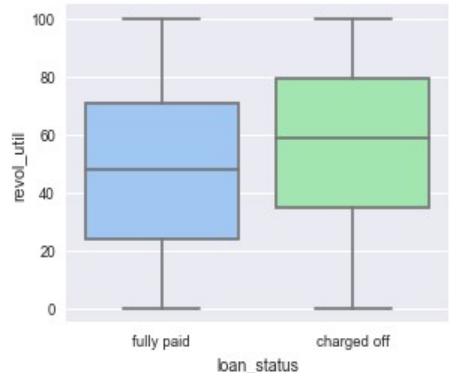
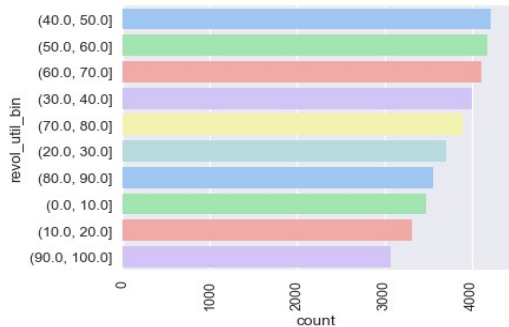


From the box plot and count plot, it is evident that, **revolving balance utilization is higher in charged off loans.**

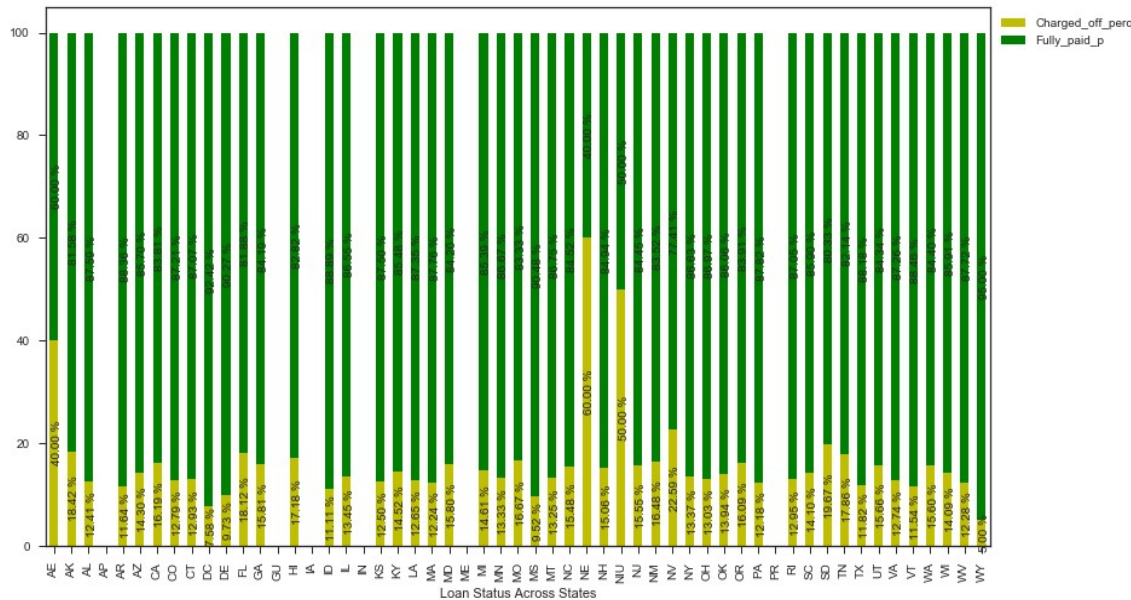
Customers with considerable revolving balance (40 - 50) are top applicants. They are also having the high percentage of charge off loans compared with applicants with utilization rate less than 40. This general trend is - the higher the revolving balance utilization rate, the higher is the chance of charging off loan.



Bank is not issuing the loans to applicants who have higher **open account ratio**. The general trend is, **the higher the open account ratio is, the higher the chance of charging off loan**



State and Zip Code Analysis



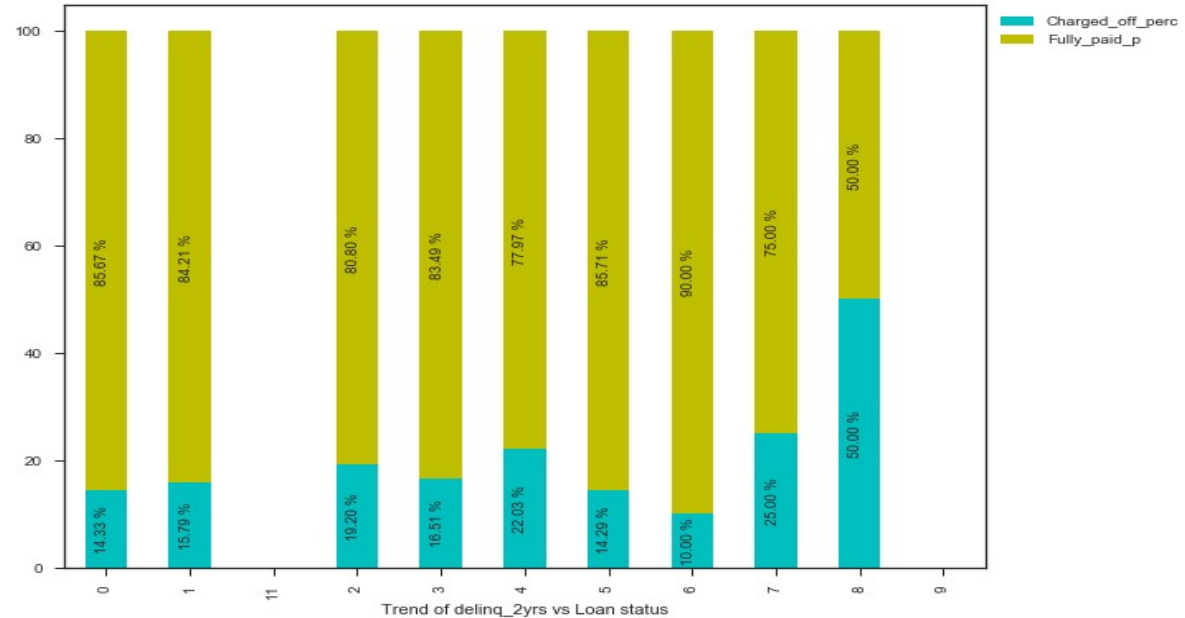
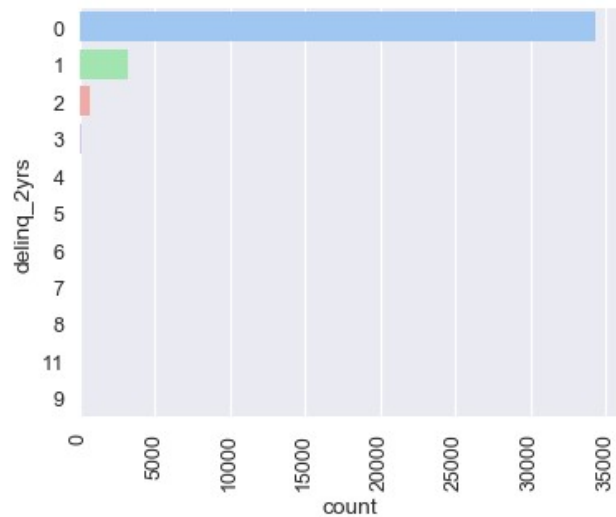
As per the above state 'CA' has highest number of charged off applications. In terms of percentages, the loans are highly charged off in the state NE, NIU and AE. Observing the number of loans applied from each of these states, it is clear, that the number of applications are less and hence this behavior. Excluding the above outlier behavior, **the state with the highest charged off percentage is NV - Nevada.**

loan_status	Counts	state_code	zip_code_str	charged off	fully paid	charged off perc
0	476	NV	891	64.00	177.00	26.56
1	71	CA	917	71.00	269.00	20.88
2	156	FL	331	70.00	267.00	20.77
3	155	FL	330	51.00	198.00	20.48
4	79	CA	925	47.00	185.00	20.26
5	455	NJ	080	45.00	190.00	19.15
6	67	CA	913	50.00	215.00	18.87
7	80	CA	926	67.00	299.00	18.31

With in a state , based on the urban area, the percentage of the defaulted loans is varying.

For example, with in 'FL' the zip code 331 has highest percentage of the defaulters.

Trend of delinq_2yrs vs Loan status



In general, LC is not accepting the customers with higher count of 'delinq_2years'.

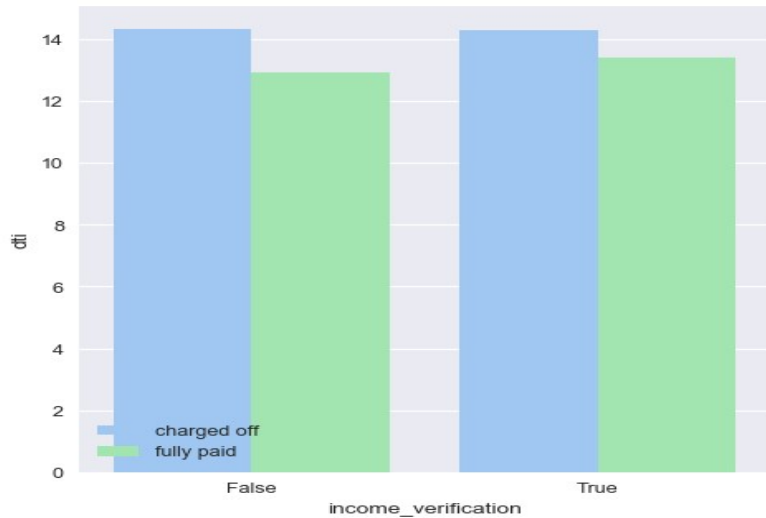
('delinq_2years' - number of 30+ days past-due incidences of delinquency)

The general trend is, **as the count of variable increases, the percentage of charged off loans increases.**

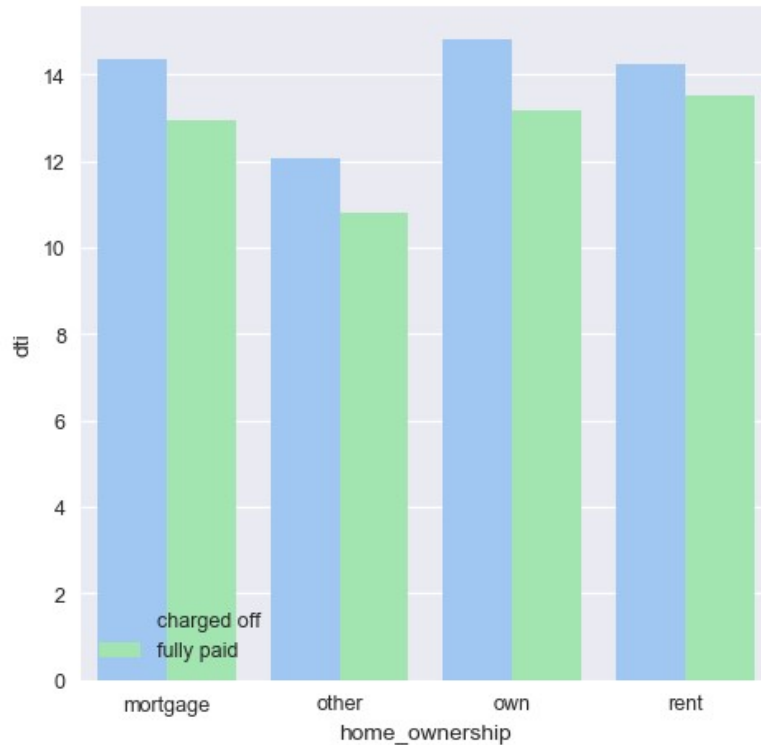
There is a **definite chance of charging off loan when the delinq_2years is high.**

Additional checks are required to understand the financial profile of the applicant in such cases.

DTI ratio vs Verification Status and Home ownership

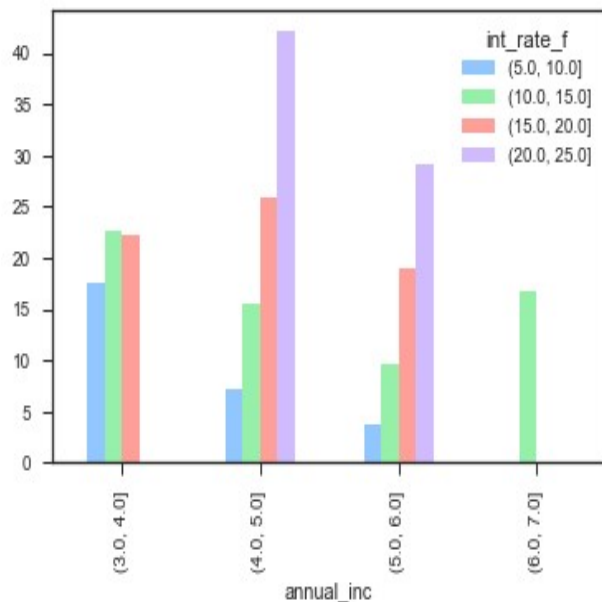


From the above it is clear, that in the cases, where **Debt to Income ratio is high**, the **income verification is not playing any role**. In such cases, verification process should be improved to assess the application.



If the DTI is higher and the home ownership is 'own', then the percentage of defaulting loan is higher.

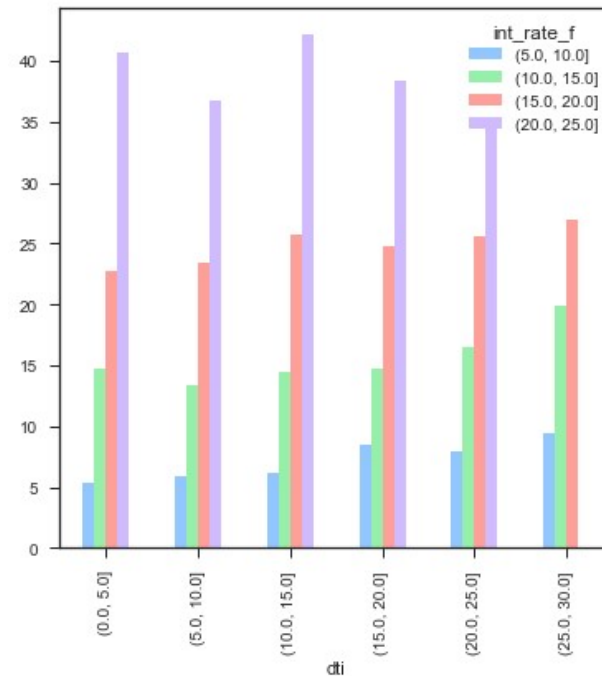
Annual Income and Interest Rate Relation ship



It is evident that -

- For same interest rate, **charge off rate decreases with increase in annual income**
- For same income, **charge off increases with increase in interest rate**

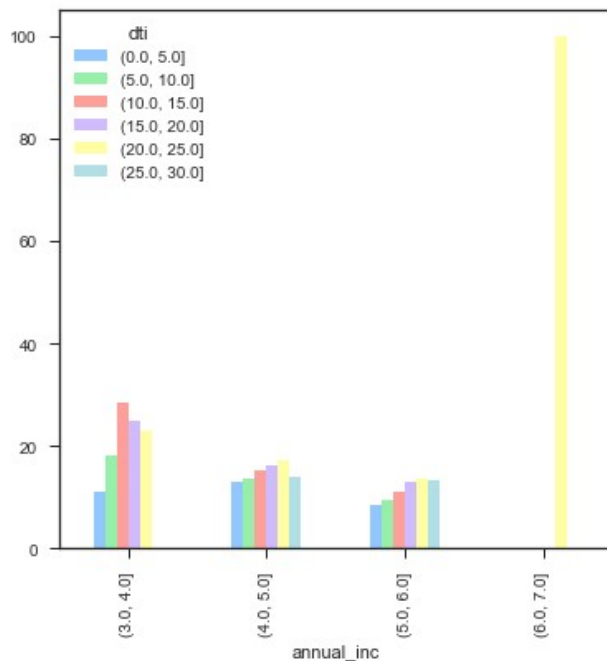
Effect of DTI and Interest Rates vs Loan status



Apparently, for the same dti, **higher interest rates result in more charge-offs**

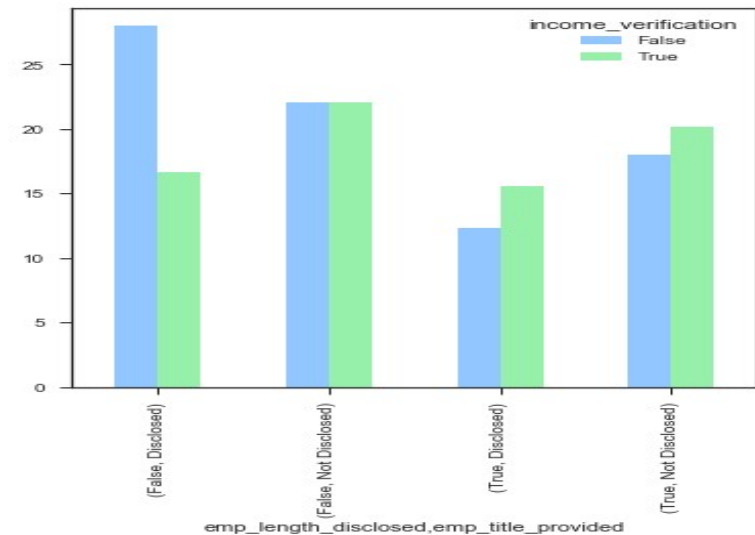
Higher interest rates have significant impact on charge offs

Annual income vs DTI



- Charge off increases when DTI increases for same annual income
- Charge off decrease when salary increase for same annual income

Relationship of employment length/title Disclosure and Income Verification vs loan status



- When **employment length is disclosed and employment title is provided, charge offs are less**
- When **employment title is provided and income is verified, charge offs are less**
- All other combination have a relatively high charge-off

It is good to have employee title for income verification (self employed and own business need more scrutiny). Additionally, when employment title is provided employment length should be expected. This will allow a higher scrutiny.