



Lending Club Case Study

Swapnil Kumar Johari

04-05-2022

Business Understanding

Lending Club specializes in lending various types of loans to urban customers. 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

In other words, **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.**



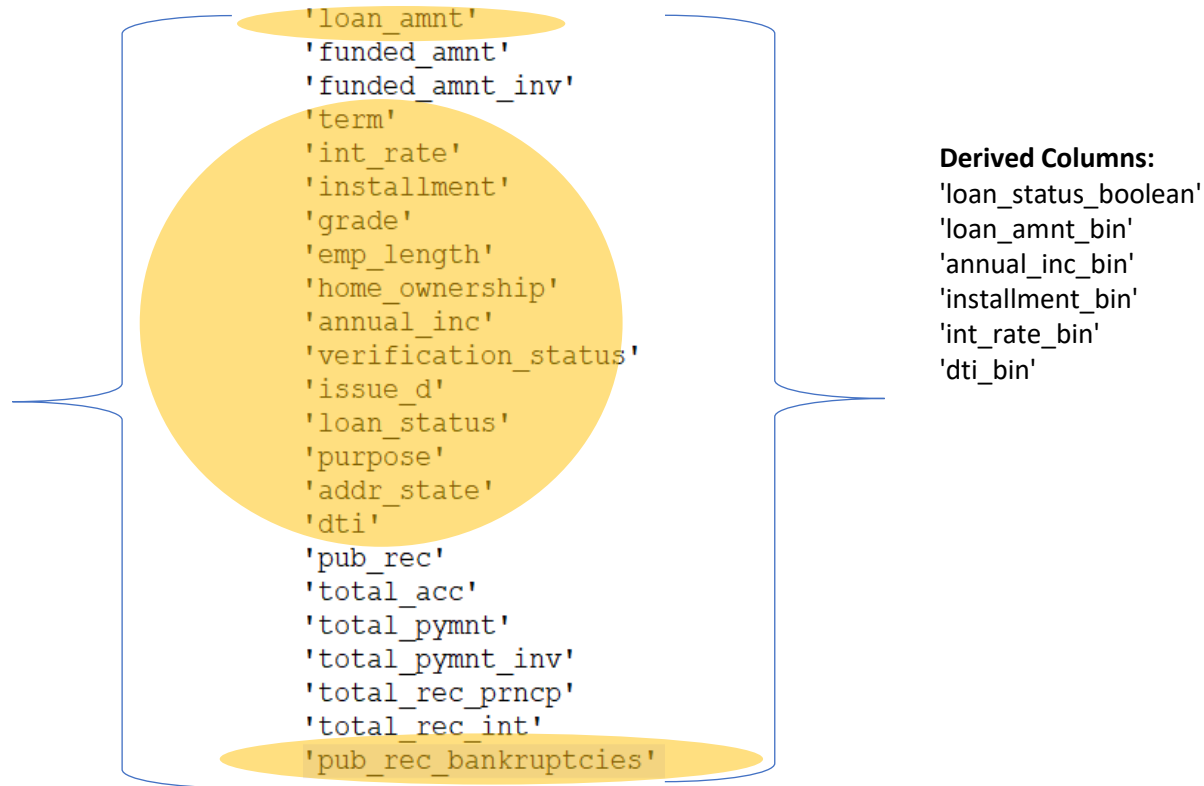
Following approaches were used while analyzing data

- Univariate Analysis
- Segmented Analysis
- Bivariate Analysis

FYI: Many Features are dropped initially based on Definition wrt Business Objectives

```
'acc_open_past_24mths'  
'collection_recovery_fee'  
'delinq_2yrs'  
'earliest_cr_line'  
'inq_last_6mths'  
'mo_sin_old_il_acct'  
'mo_sin_old_rev_tl_op'  
'mths_since_last_delinq'  
'mths_since_last_major_derog'  
'mths_since_recent_revolt_delinq'  
'num_accts_ever_120_pd'  
'open_acc'  
'tax_liens'  
'bc_open_to_buy'  
'collections_12_mths_ex_med'  
'desc'  
'emp_title'  
'id'  
'inq_fi'  
'last_credit_pull_d'  
'last_pymnt_amnt'  
'last_pymnt_d'  
'member_id'  
'mo_sin_rcnt_rev_tl_op'  
'mo_sin_rcnt_tl'  
'mths_since_last_record'  
'mths_since_rcnt_il'  
'mths_since_recent_bc'  
'mths_since_recent_bc_dlg'  
'mths_since_recent_inq'  
'next_pymnt_d'  
'total_rec_late_fee'  
'url'
```

After Cleaning/Standardizing data and Outliers treatment, Analysis to be done on

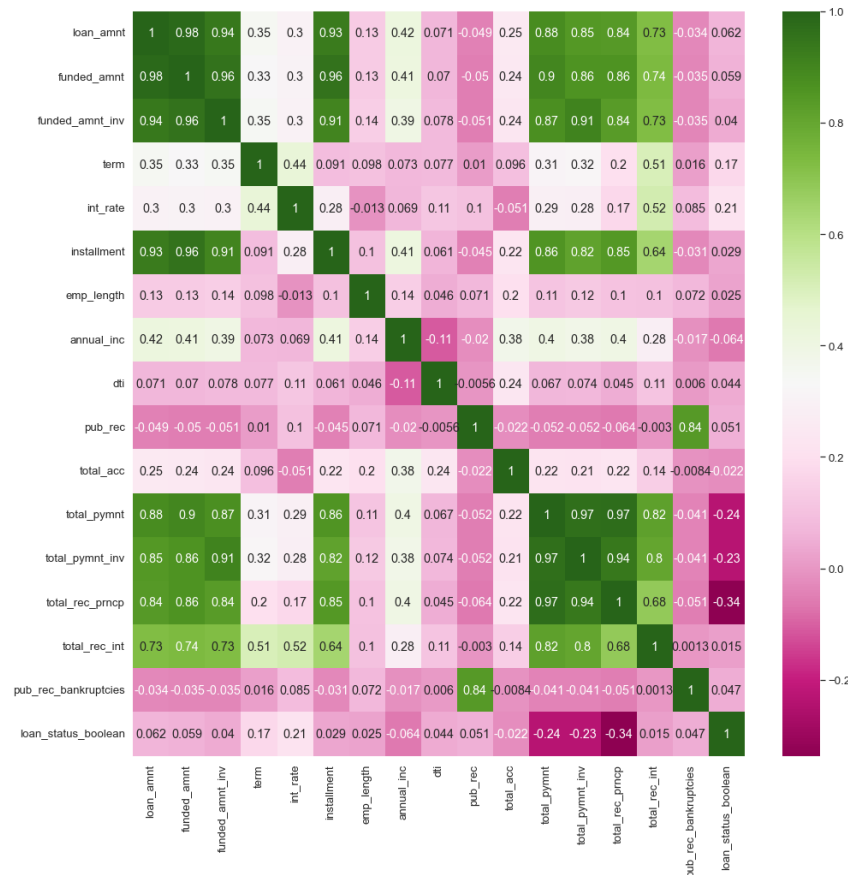


Co-relation between Variables

Loan Amount, Funded Amount, Total Payment and related fields are co-related to each other.

As the loan Amount increases, Total Payment/ Funded Amount increases.

More payment towards the total principal Amount is inversely proportional to Loan Status Boolean where 0 : Paid and 1: Charged Off



Categorical Value:

Unordered:

['home_ownership', 'verification Status', 'purpose',
'loan_status_boolean', 'addr_state']

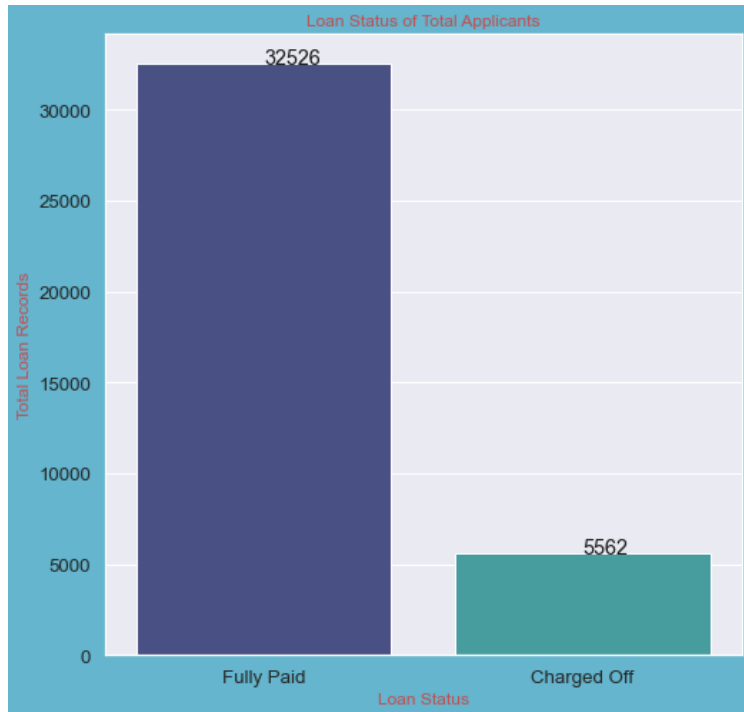
Ordered:

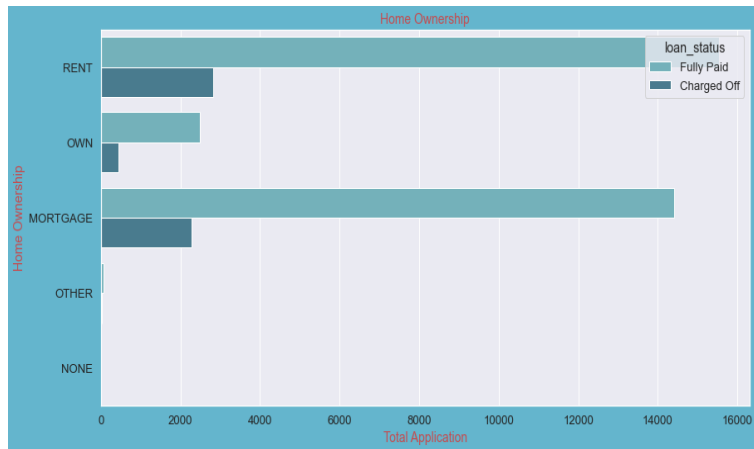
['grade', 'emp_length', 'term', 'issued_m']

Quantitative Variable:

['loan_amnt', 'funded_amnt', 'funded_amnt_inv', 'int_rate', 'installment',
'annual_inc', 'dti', 'total_pymnt', 'total_pymnt_inv']

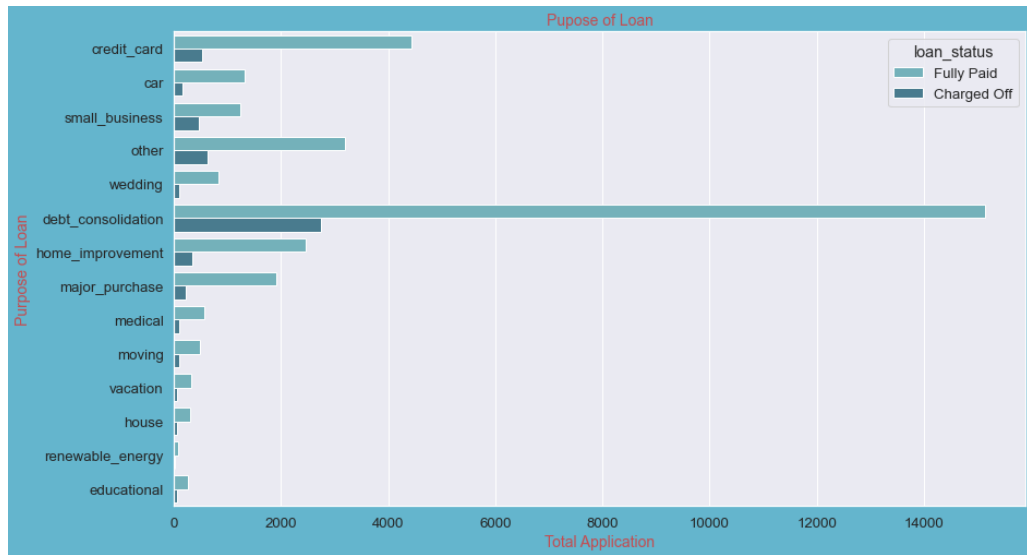
Out of Total Records, around 15% applicants are the defaulter ones





Home Ownership	Charged Off (%)
Rent	15%
Own	15%
Mortgaged	13%
Other	19%

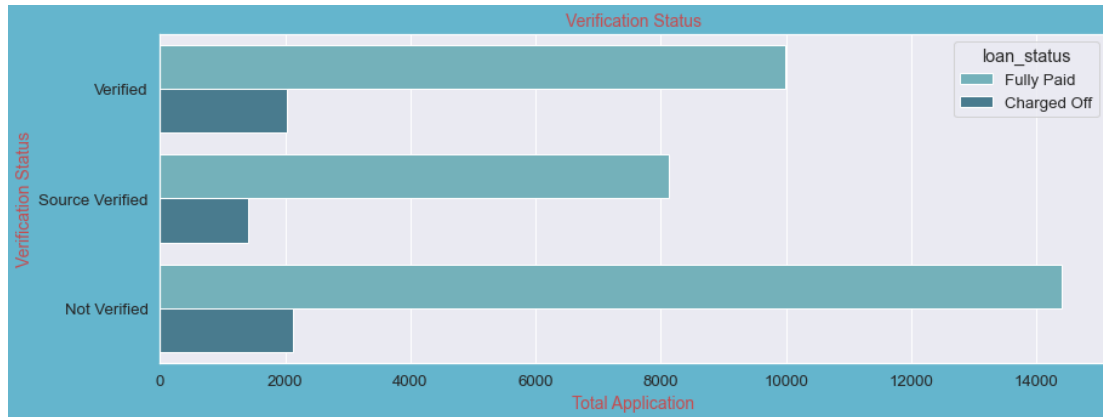
Most of the applicant lives in **Rented House** or have **mortgaged** their property but out of Overall Charged Off % in each Category, The applicants where **Other** is Mentioned are more Likely be the defaulters one.



Purpose	Charged Off %
small_business	27%
renewable_energy	19%
educational	17%
other	16%
house	16%
moving	16%
medical	16%
debt_consolidation	15%
vacation	14%
home_improvement	12%
credit_card	11%
car	11%
major_purchase	10%
wedding	10%

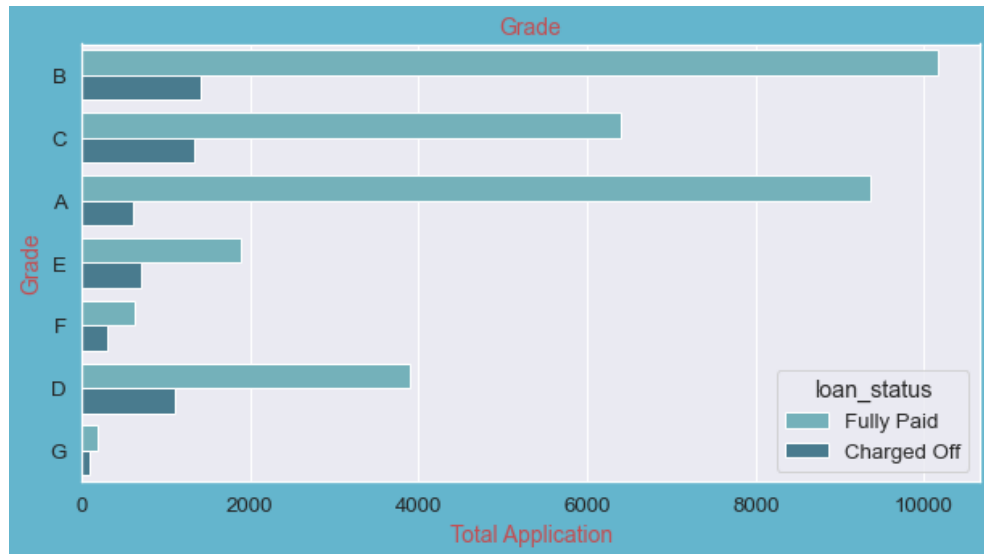
Most of the applicants have taken loan for '**debt_consolidation**', '**other**' and '**credit_card**'
Out of which '**small_business**' and '**renewable_energy**' are the one who were charged Off.

Univariate: Verification Status



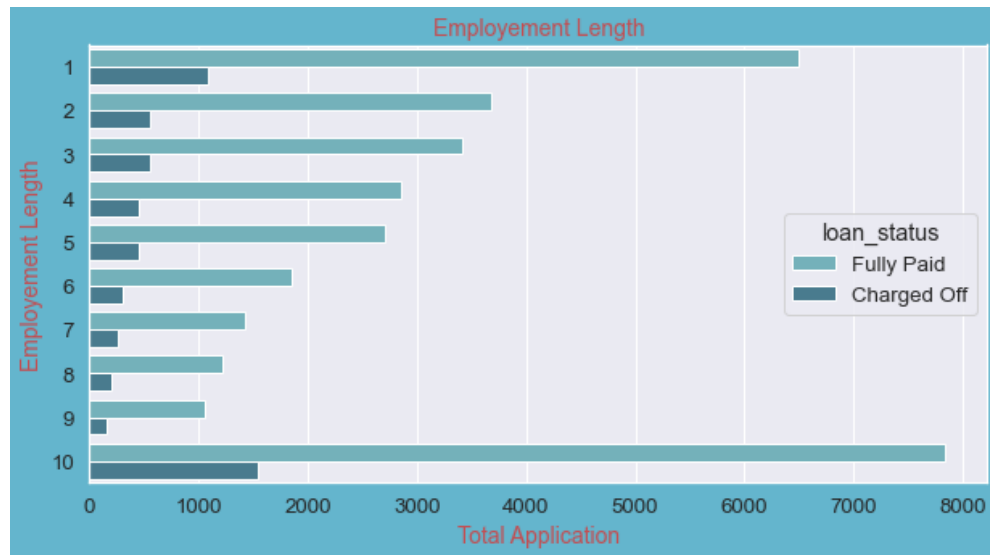
verification_status	Charged Off %
Verified	17%
Source_Verified	15%
Not Verified	13%

This data shows suspicious data where most of the Verified applicants are charged Off.



Grades	Charged Off %
G	34%
F	33%
E	27%
D	22%
C	17%
B	12%
A	6%

Most of the applicants are B/C/D graded who got charged off, out of which E/F/G graded applicants are most likely to Charged off

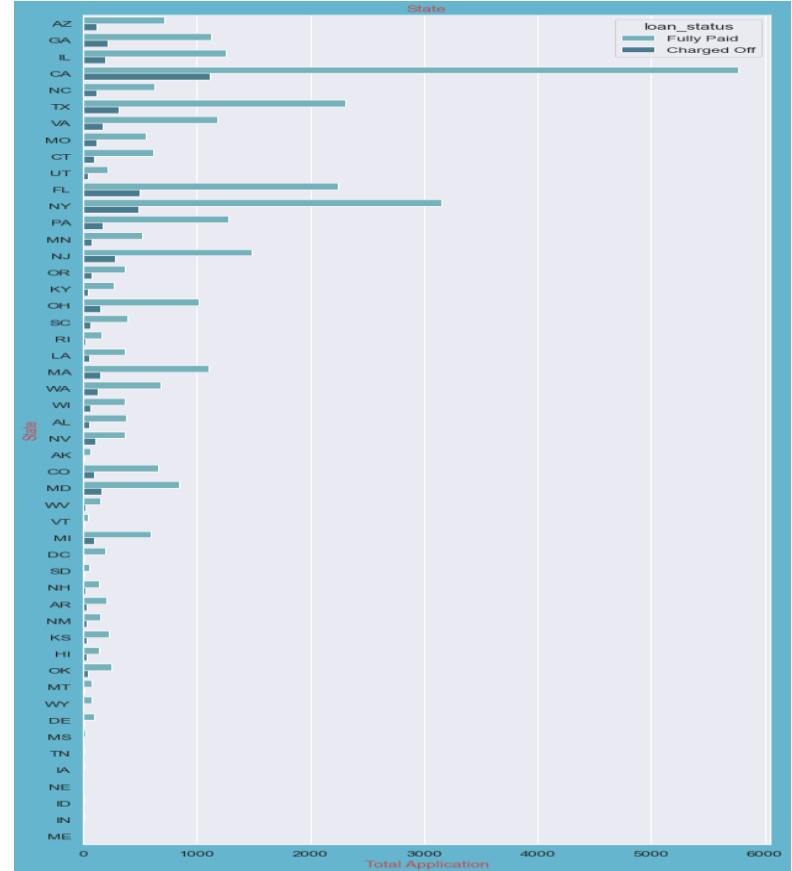


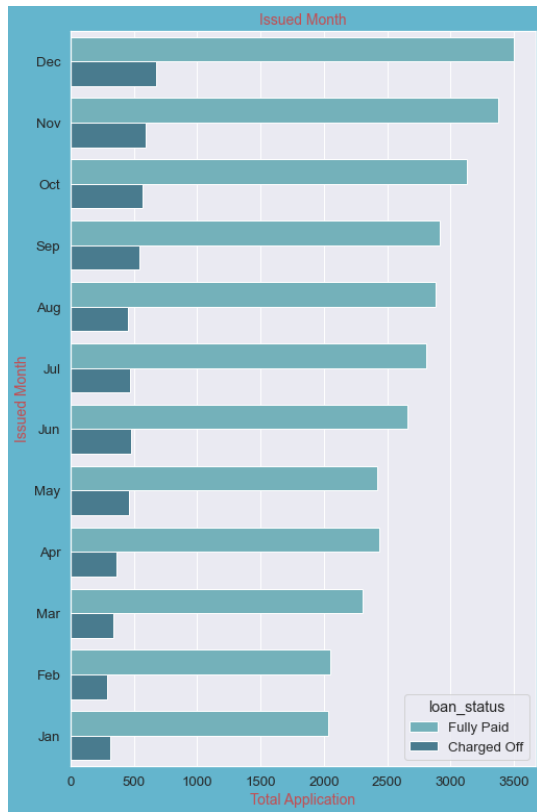
emp_length	Charged off %
10	16%
7	15%
5	14%
8	14%
1	14%
6	14%
3	14%
4	14%
2	13%
9	13%

Most of the applicants are experienced more than 10 Years and others are in Ascending order starting from Freshers, out which Experienced are the Ones who got charged off frequently

Most of the applicants are from **CA & NY States** and applicants who have taken loans from **NE/NV/SD** States are more likely to get Charged Off

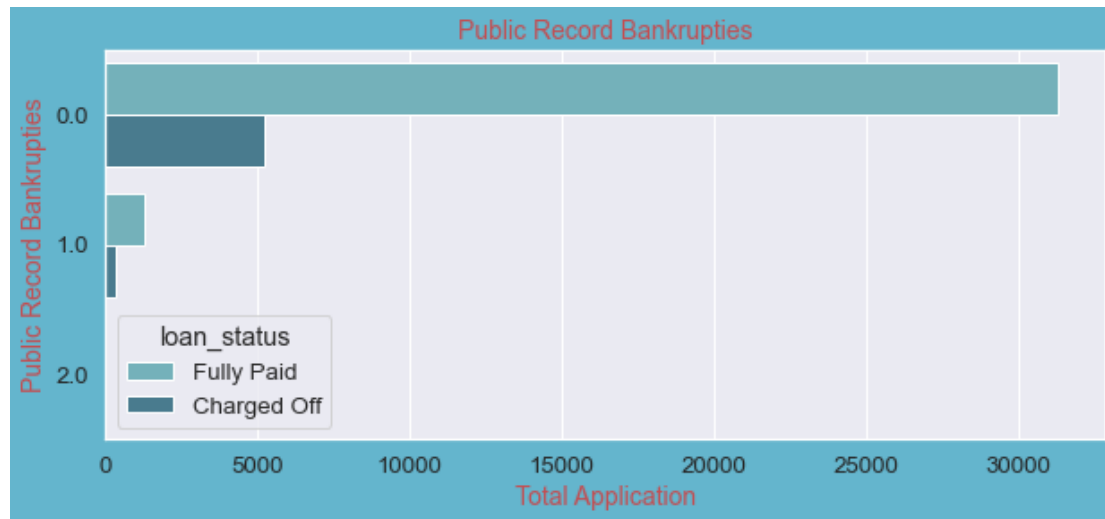
addr_state	Charged off %
NE	33%
NV	23%
SD	20%
AK	18%
FL	18%
MO	17%
HI	17%
NM	17%
ID	17%
CA	16%
OR	16%
UT	16%
GA	16%
MD	16%
NJ	16%
WA	16%
NC	15%
NH	15%
MI	15%
SC	15%
KY	14%
WI	14%
AZ	14%
OK	14%
IL	13%
NY	13%
MN	13%
MT	13%
CT	13%
RI	13%
OH	13%
CO	13%
VA	13%
LA	13%
MA	12%
AL	12%
WV	12%
KS	12%
PA	12%
TX	12%
TN	12%
AR	12%
VT	11%
MS	11%
DE	10%
DC	7%
WY	5%





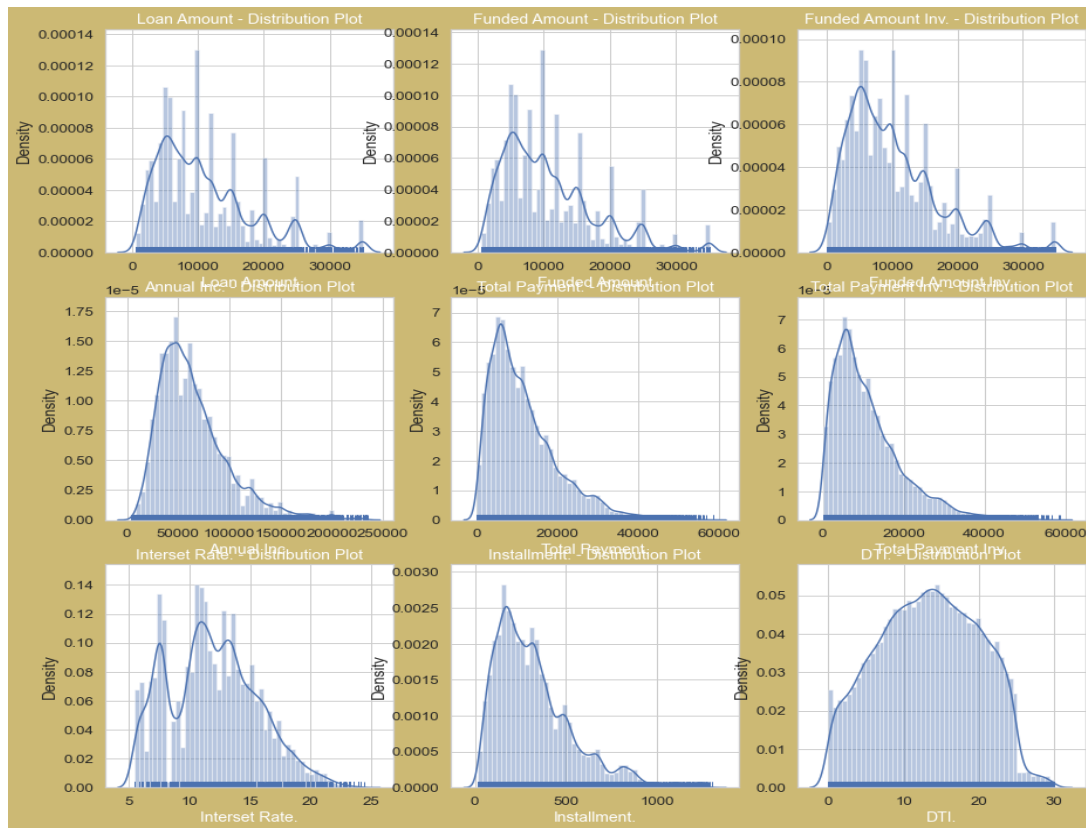
issue_d	Charged Off %
Dec	16%
May	16%
Sep	16%
Oct	15%
Jun	15%
Nov	15%
Jul	14%
Aug	14%
Jan	13%
Apr	13%
Mar	13%
Feb	12%

Maximum applicants have applied in **December** and same people are most likely to get **charged Off**

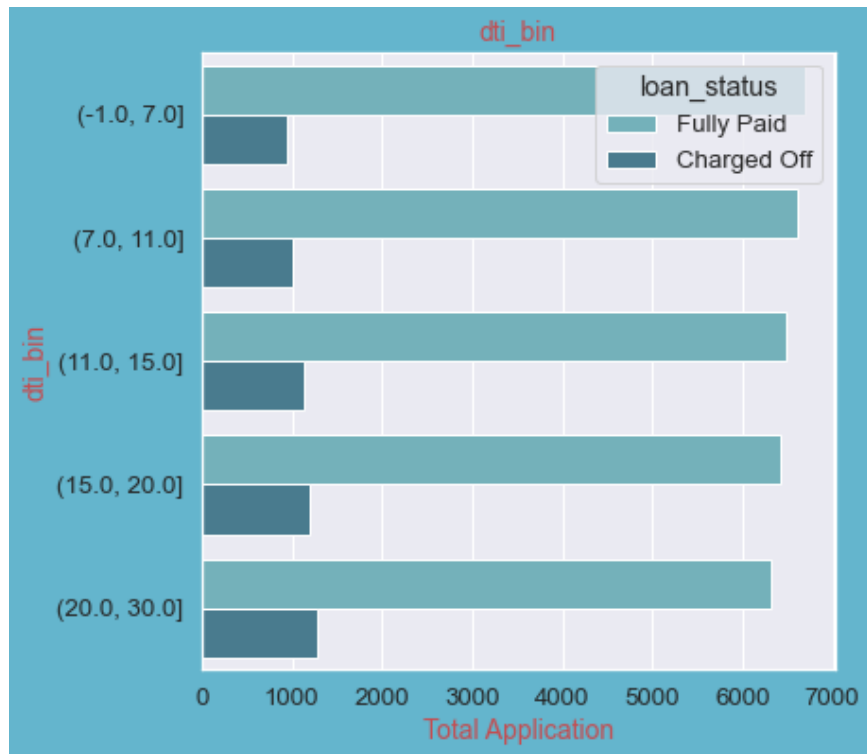


loan_status	Charged Off %
2	40%
1	22%
0	14%

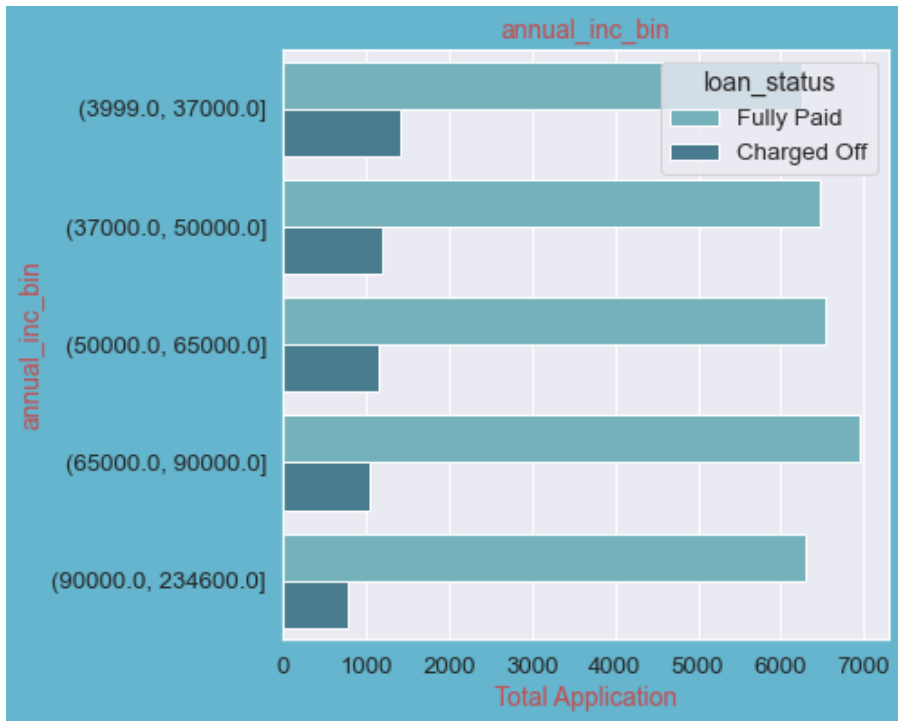
Most of the applicants who have **more than 1 bankruptcies records** are more likely to get charged off



- **Loan Amount, Funded amount and Funded Amount Inv** are having **similar** type of **curves** and tightly **corelated**, so Loan Amount can be look in to Independently.
- **Annual Inc, Total Payment and Total Payment Inv** are having **Identical plot** so Annual Inc can be look in to in more detail.
- Interest Rate/Installment/DTI needs to be analyzed independently.



Most of the applicants having **Higher DTI Ratio**, which can be considered High Risk for the Organization

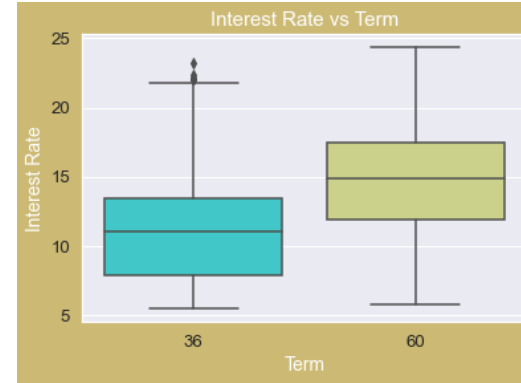
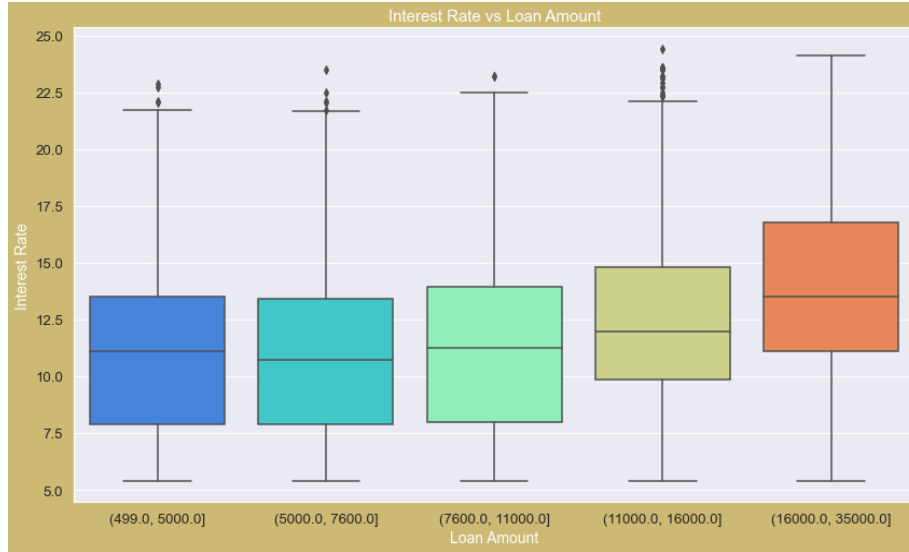


There is **high risk** for the loan issued to **lower annual Income band** of applicants

Analysis to be done on Following Features relation:

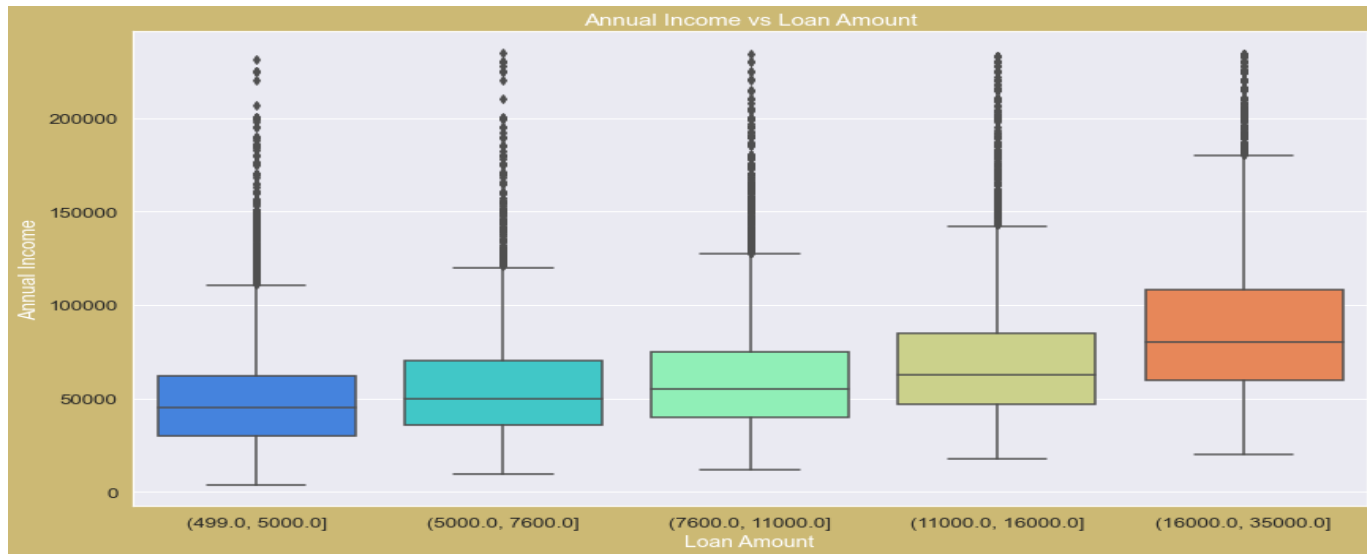
- ❖ int_rate vs Term
- ❖ loan_amnt vs int_rate
- ❖ loan_amnt vs annual_inc_bin
- ❖ int_rate_bin vs dti
- ❖ installment vs annual_inc_bin vs loan_status
- ❖ annual_inc vs total_acc vs loan_status
- ❖ dti vs annual_inc_bin
- ❖ grade vs int_rate_bin
- ❖ emp_length vs pub_rec_bankruptcies vs loan_status
- ❖ home_ownership vs purpose vs loan_status
- ❖ verification_status vs pub_rec_bankruptcies
- ❖ addr_state vs purpose vs loan_status

Loan Amount vs Interest Rate vs Term



Interest Rate increases with increase in Loan Amount and Terms

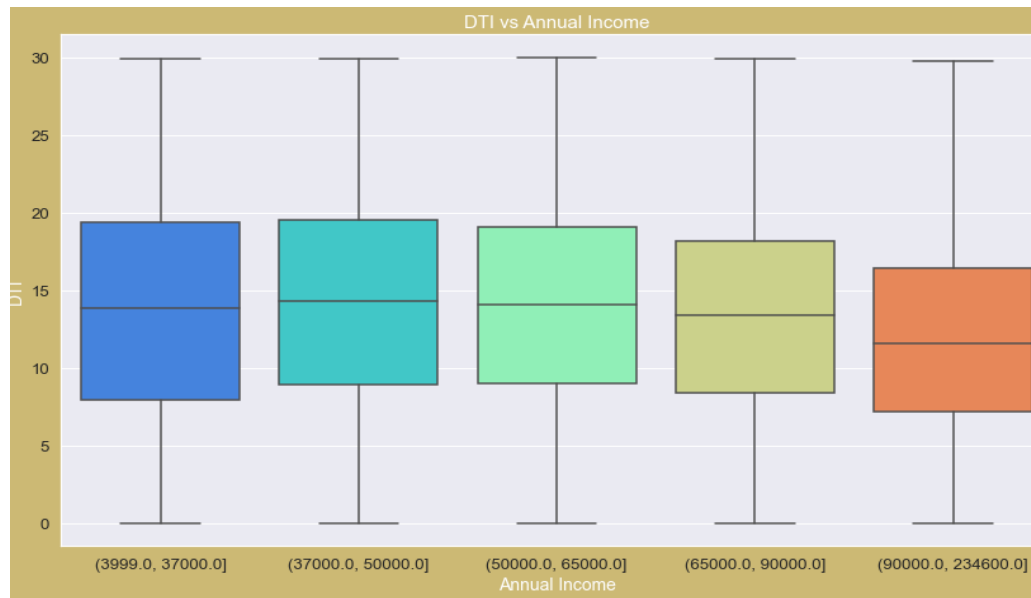
Annual Income vs Loan Amount



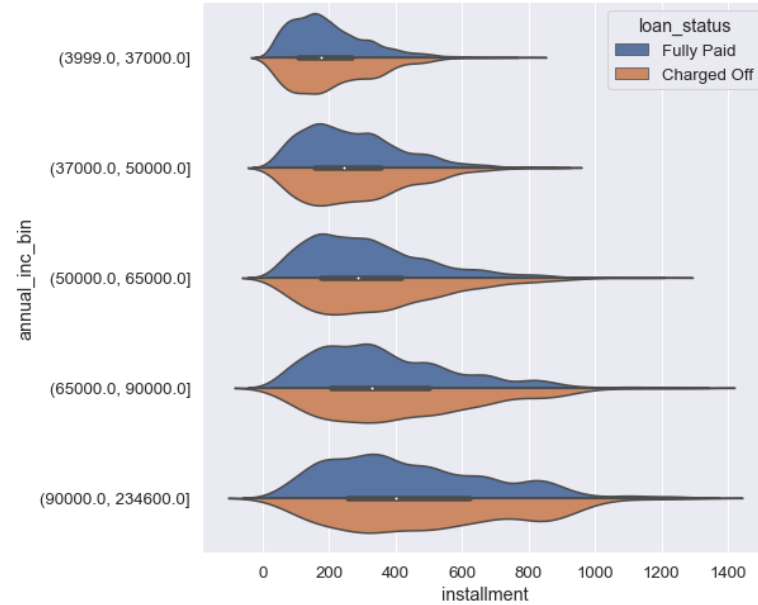
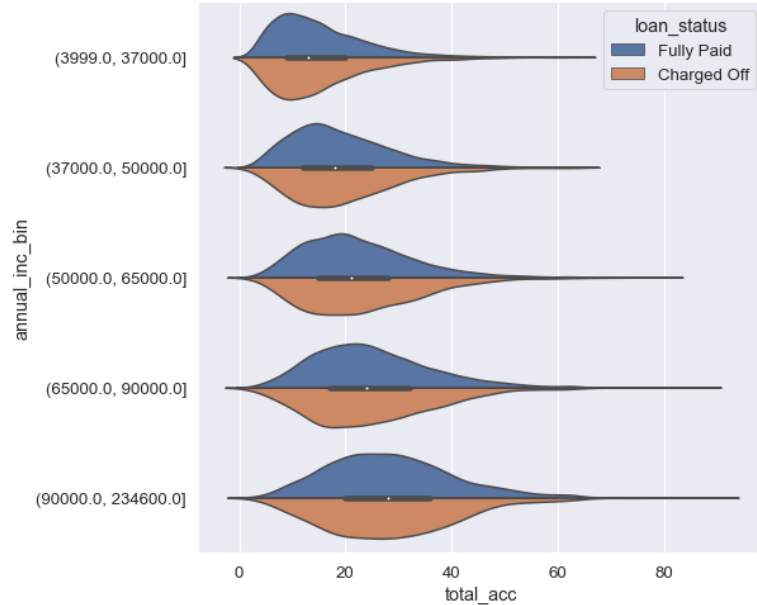
Higher the Annual income, higher the Loan Amount

> Applicants having Higher Income pays Higher Interest and preferred High Loan amount to be paid on longer Term

DTI is preferably better for Applicants having higher Income

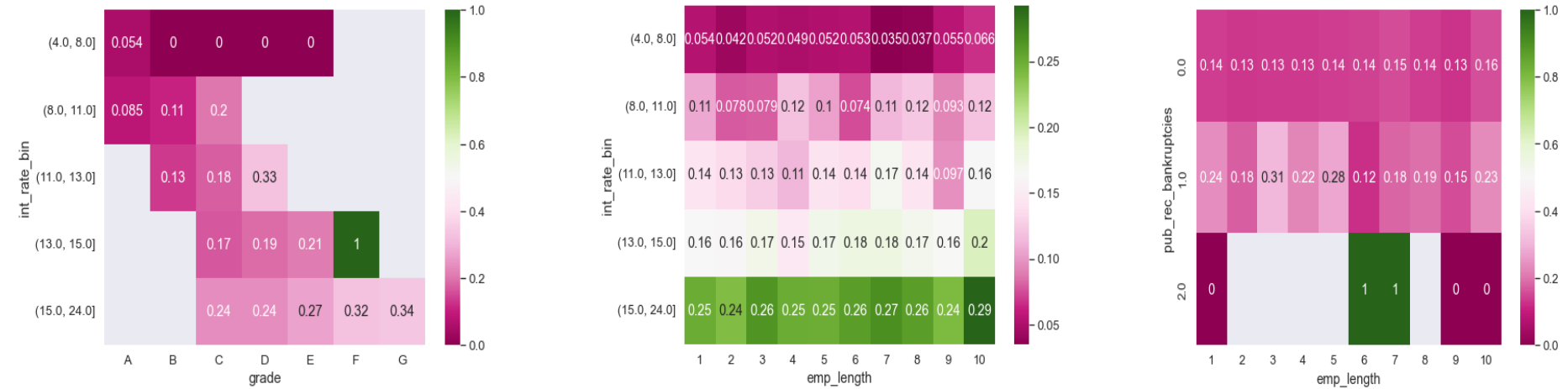


Annual Income vs Total Account vs Installment vs Loan Status



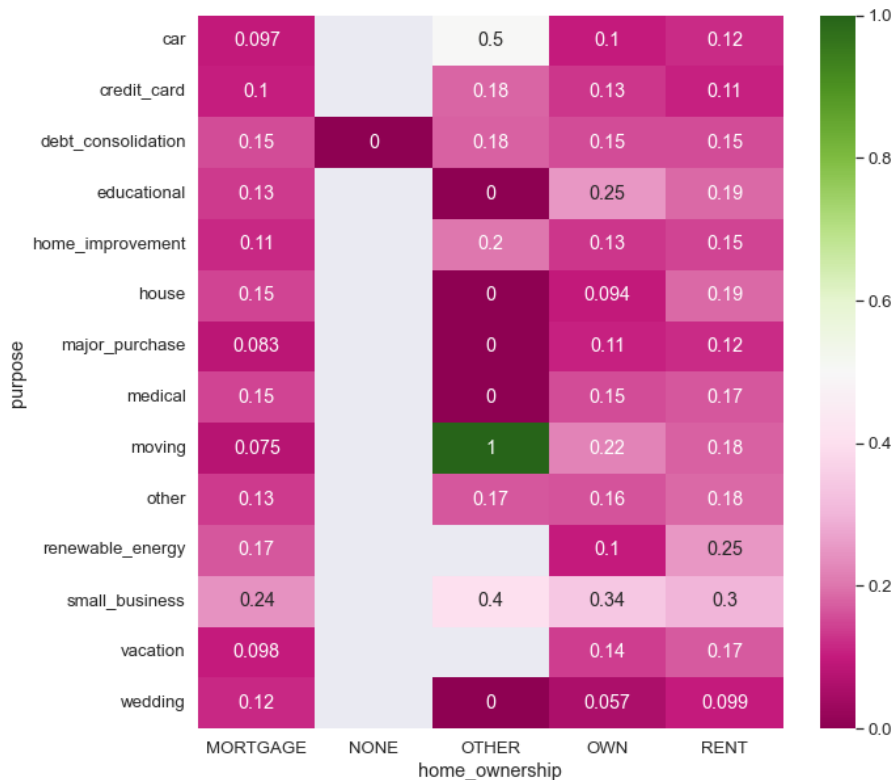
As the Annual Income Increases, Installment Increases along with Total Charged Off in each Bins

Bivariate Categorical: Interest Rate, Grade, Employee Length, Public Bankruptcies



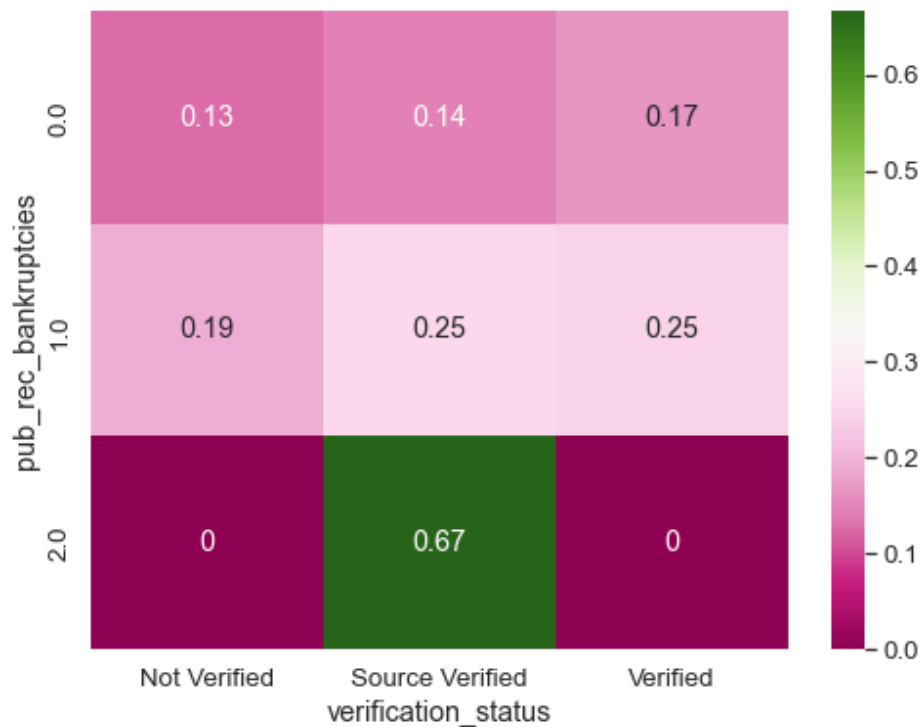
Applicants having grade 'F' and Interest rate 13 - 15 are more likely to charged Off irrespective of employee length until there is Public Bankruptcies record for employee having length 6-7.

Purpose vs Home Ownership



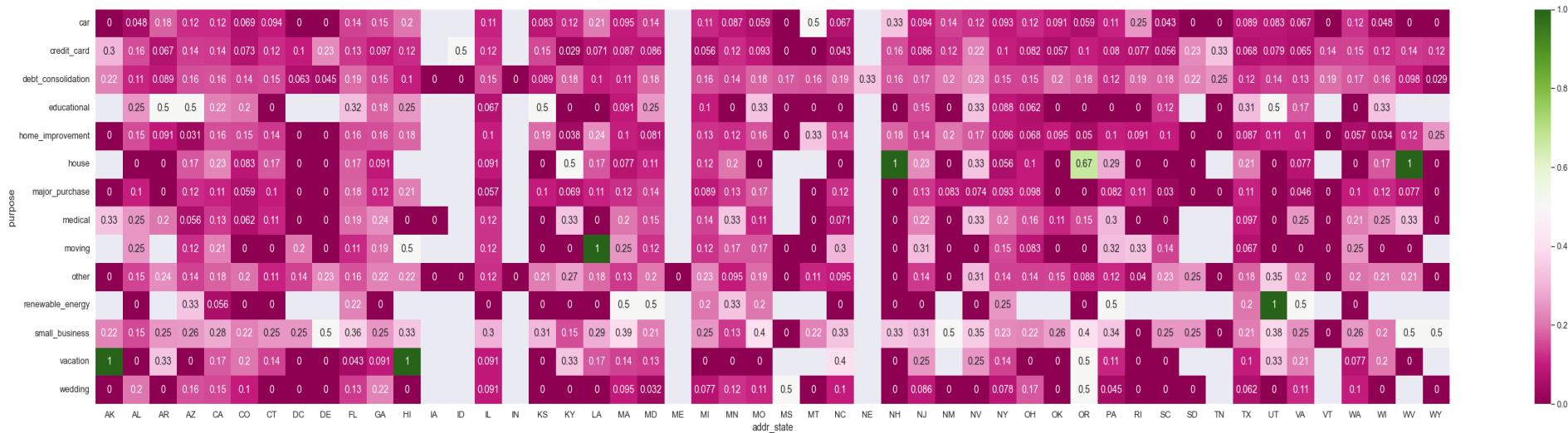
Applicants have mention home Ownership as **Other** and taken loan for **Moving** Purpose are more likely to have Charged Off

Bankruptcies vs Verification Status



This data shows that there is some issue with Verification Process as the Applicant's having large number of bankruptcies reported are more like to have Charged Off

Purpose vs Address State



Applicants applied from States **AK/HI/LA/NH/OR/UT/WV** having purpose of **Vacation/Moving/House/Renewable Energy** respectively are more likely to get Charged Off.

- ❑ There are some ambiguity with Verification Process because the applicant who got verified are Riskier to Organization, The LC Should change/update process of Verification.
- ❑ The applicants where **Other** is Mentioned are more Likely be the defaulters one who are having purpose of *small_business* and *renewable_energy*. To get more insight on this , replace Others Category with some precise Variables.
- ❑ Need to add some regulations who are taking loan for *Vacation/Moving/House/Renewable Energy* specifically from states *AK/HI/LA/NH/OR/UT/WV*.
- ❑ *E/F/G* Graded Applicants are riskier one and if higher rate of interest i.e., more than 13% is offered to them and Bankruptcies record more than 1 then it will highly increase the probability of Charged Off.
- ❑ Loan Issued on *May/September/December* are more likely to get charged Off.
- ❑ Bankruptcies records more than 1 are more likely gets charged Off
- ❑ Higher Interest Rate i.e., more than 13% effect applicants in repayment of loan
- ❑ Hight Interest Rate and Longer Term Increase the Probability of charged off.