LENDING CLUB CASE STUDY

Prepared By,
Ankit Lohiya
Arun Vignesh R

INTRODUCTION

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:

o If the applicant is **likely to repay the loan,** then not approving the loan results in a **loss of business** to the company

o 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

PROBLEM STATEMENT

This company is the 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.

Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss

Borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as **'charged-off'** are the **'defaulters'**.

The agenda is to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.
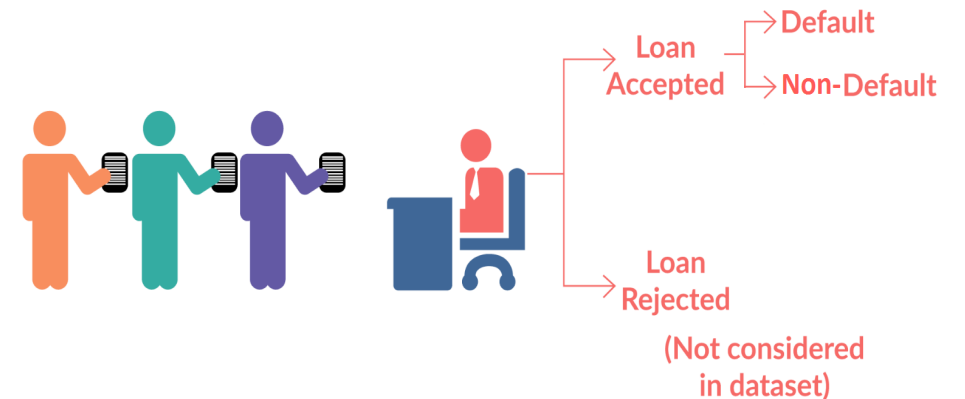
DATASET

The dataset we analyzed contains the complete loan data for all loans issued through the time period 2007 to 2011.

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

1. **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:

   - **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)

   - **Current:** Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

   - **Charged-off:** Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan

2. **Loan rejected:** The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

LOAN DATASET

QUICK LOOK INTO THE DATA SET

The data set has 111 different variable and 39717 entries (loans).

These variables can be classified to:
- Related to the borrower (employment length, home ownership, address etc)
- Loan characteristics (Loan amount, term etc)
- Customer behaviour (next payment due, months since last delinquency etc)

CLEANING THE DATASET

As part of the data analysis we have cleaned the dataset and removed those variables which doesn't have any impact on our analysis.

Out of the 111 variable for 56 variables more that 90% values are null hence we have removed them.

Apart from that some variables especially those which belongs to customer behavioral type has also been removed from the data, since these information will not be available to us when applicant requests for the loan.

We also have ignored the loan entries which are currently ongoing, since this won't give us any clarity on when it will be fully paid or defaulted

After all these we have about 26 relevant variables on which we will be working on.

INTRODUCTION TO THE VARIABLES

Following are the list of variables on which the analysis is done. You will be seeing these variables a lot in the upcoming slides.

id	A unique LC assigned ID for the loan listing.
addr_state	The state provided by the borrower in the loan application
annual_inc	The self-reported annual income provided by the borrower during registration.
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
funded_amnt	The total amount committed to that loan at that point in time.
funded_amnt_inv	The total amount committed by investors for that loan at that point in time.
grade	LC assigned loan grade
home_ownership	The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
installment	The monthly payment owed by the borrower if the loan originates.
int_rate	Interest Rate on the loan
issue_d	The month which the loan was funded
Pub_rec_bankrupties	Number of public record bankruptcies
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
loan_status	Current status of the loan
mths_since_last_delinq	The number of months since the borrower's last delinquency.
purpose	A category provided by the borrower for the loan request.
sub_grade	LC assigned loan subgrade
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified
annual_inc_binned*	Segments of annual incomes
funded_amnt_inv_binned*	Segments of funded amount by investors
Loan_amnt_binned	Segments of loan amount requested by borrower
Dti_binned	Segments of DTO
Int_rate_binned	Segments of interest rate

* These are extra variables created using the existing variables

DECISION MADE ON VARIABLE TO CONSIDER
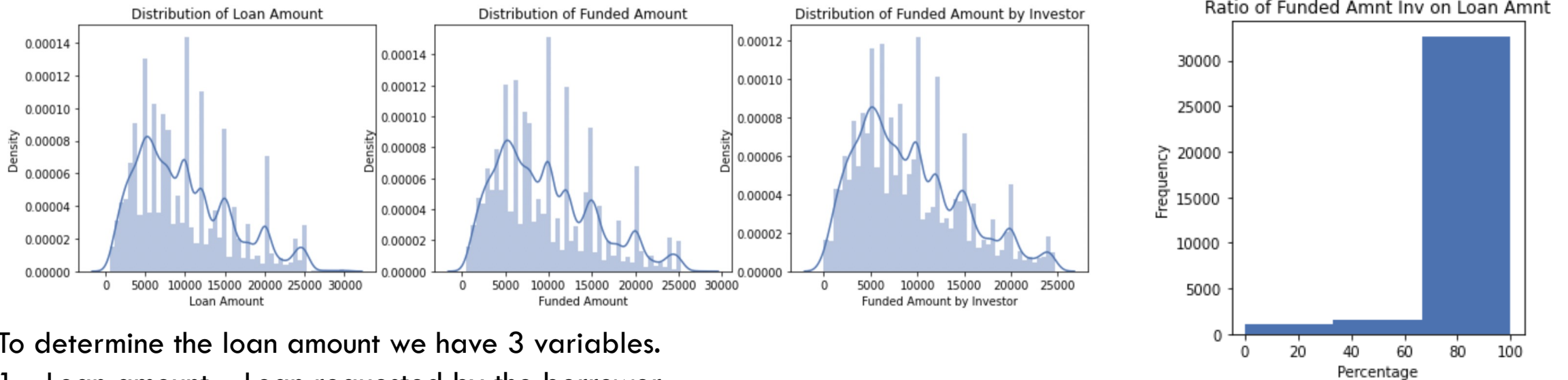
To determine the loan amount we have 3 variables.
1. Loan amount – Loan requested by the borrower
2. Fund amount – Fund approved by lending club
3. Fund amount by investor – Fund provided by the investor
- You can see that the distribution is same for all three variables i.e Loan Amount, Funded Amount and Funded Amount by Investor
- The ratio of Funded Amount Inv against Loan Amount has considerable number of entries between 0-80%. Which means that not for all requested Loan Amount, Funded amount by investor is 80 - 100%
- The final amount which a borrower gets is the Funded Amount by Investor. **Hence we have decided to perform all analysis mostly using Funded Amount by investor!**

PERCENTAGE OF FULLY PAID & CHARGED OFF LOANS

Percentage of Loan Status
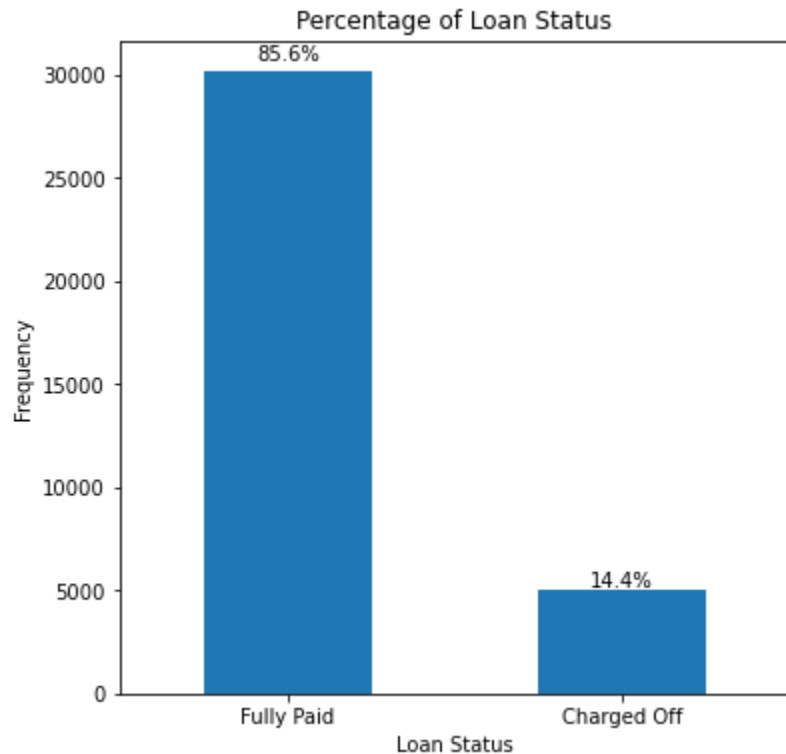
- As per the previous data records appx 85% of the loans are paid off and appx 15% are charged off
- Going forward we will be looking into how different variables behave here
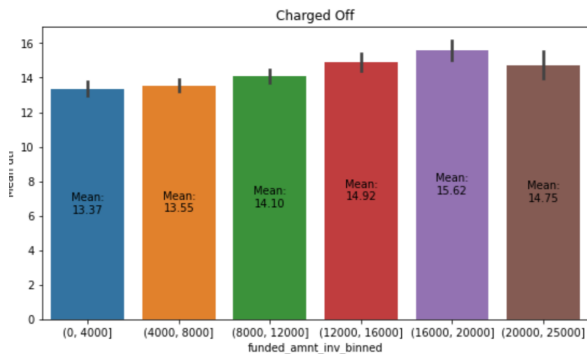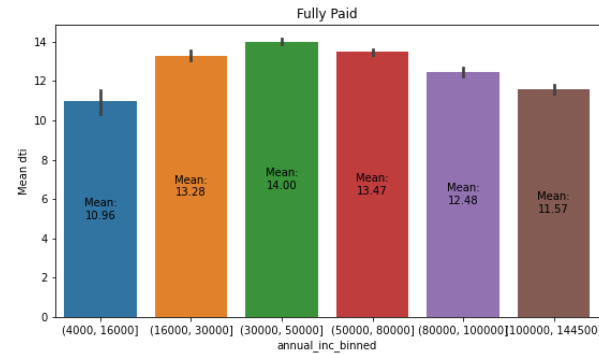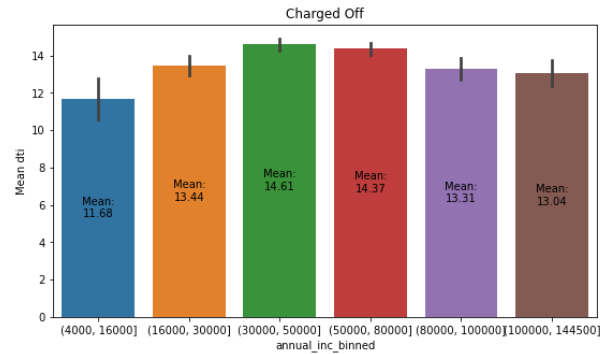
DISTRIBUTION OF INTEREST RATE & DTI
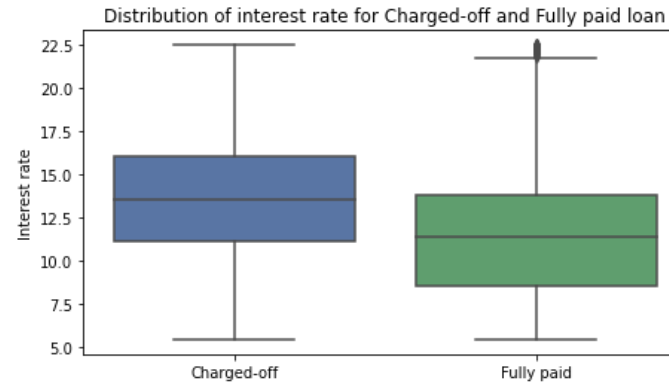
- Interest rate is higher for Charged-off compared to Fully paid.
- **Higher the interest rate, Loan becomes more risky**
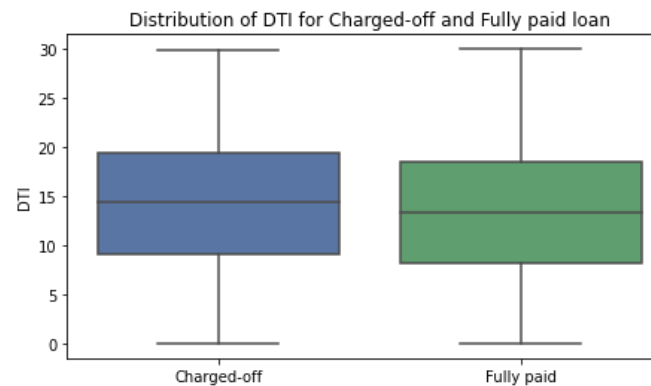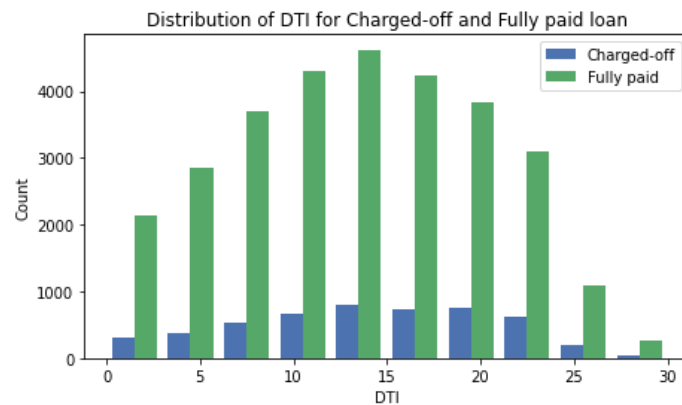
- DTI is slighter higher for Charged-off category when compared to Fully paid.

# DTI WITH RESPECT TO ANNUAL INCOME/FUNDED AMOUNT BY INVESTOR
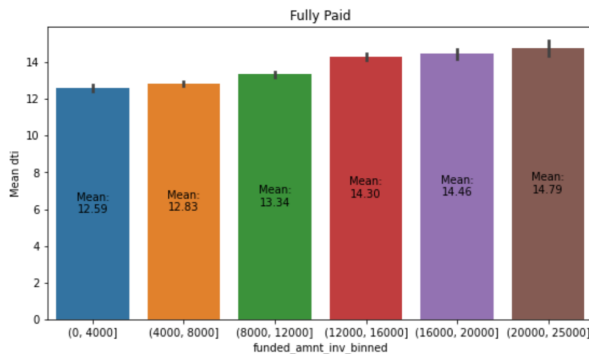
- In general when we compare b/w fully paid and charged off the dti is more by 1-2(%) in each income category.

- **With this we could provide an upper limit for DTI in each income category, more than that would be considered risky, for ex: for income class b/w 30k-50k providing loan for borrowers with dti greater that 14 is risky**

- **The same is with the funded amount as well, for each funded amount range dti ratio can be determined over which it will be considered risky**

Note: Mean here means average

# INTEREST RATE WITH RESPECT TO ANNUAL INCOME/FUNDED AMOUNT BY INVESTOR
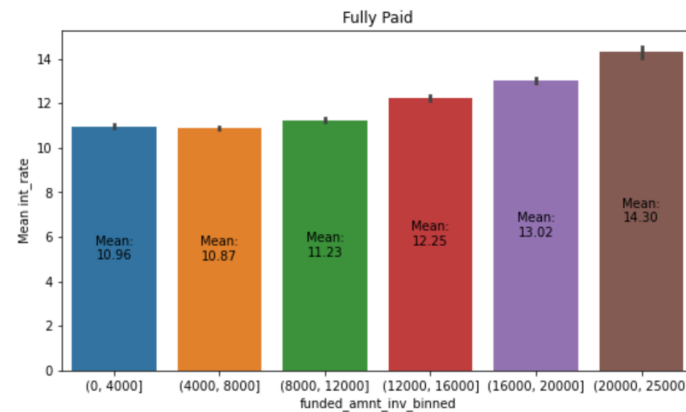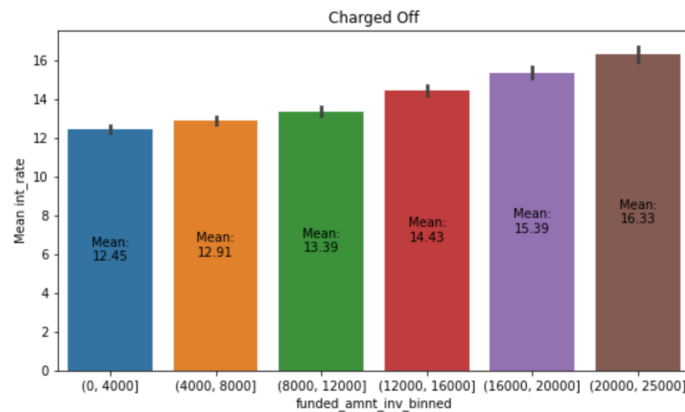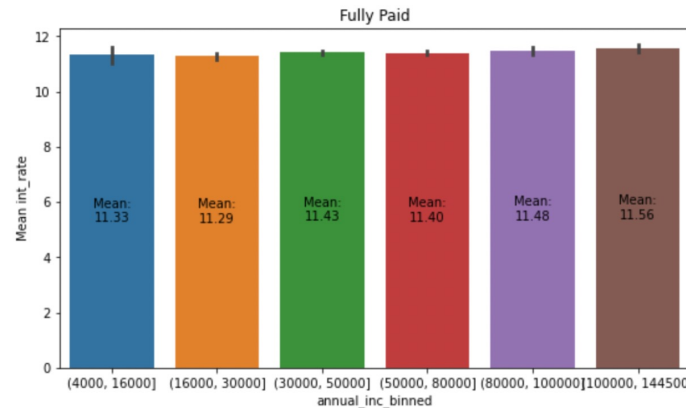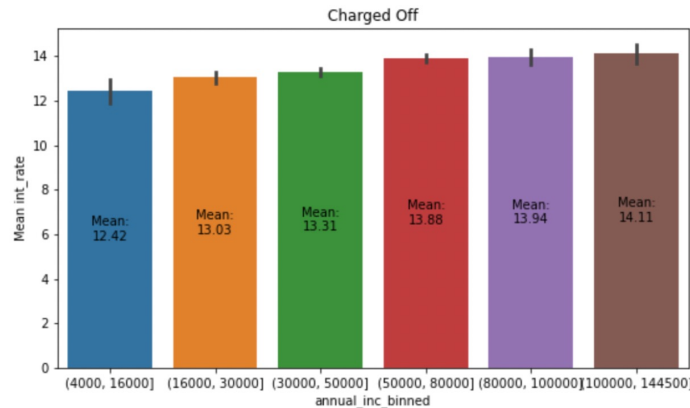
- In general when we compare b/w fully paid and charged off the interest rate is more by 1-2(%) in each funded amount invested/annual income category.
- **With this we could provide an upper limit for interest rate in each funded amount category/annual income, more than that the loan would be considered risky**

DISTRIBUTION OF ANNUAL INCOME AND EMPLOYEE LENGTH

Distribution of Annual Income for Charged-off and Fully paid loan

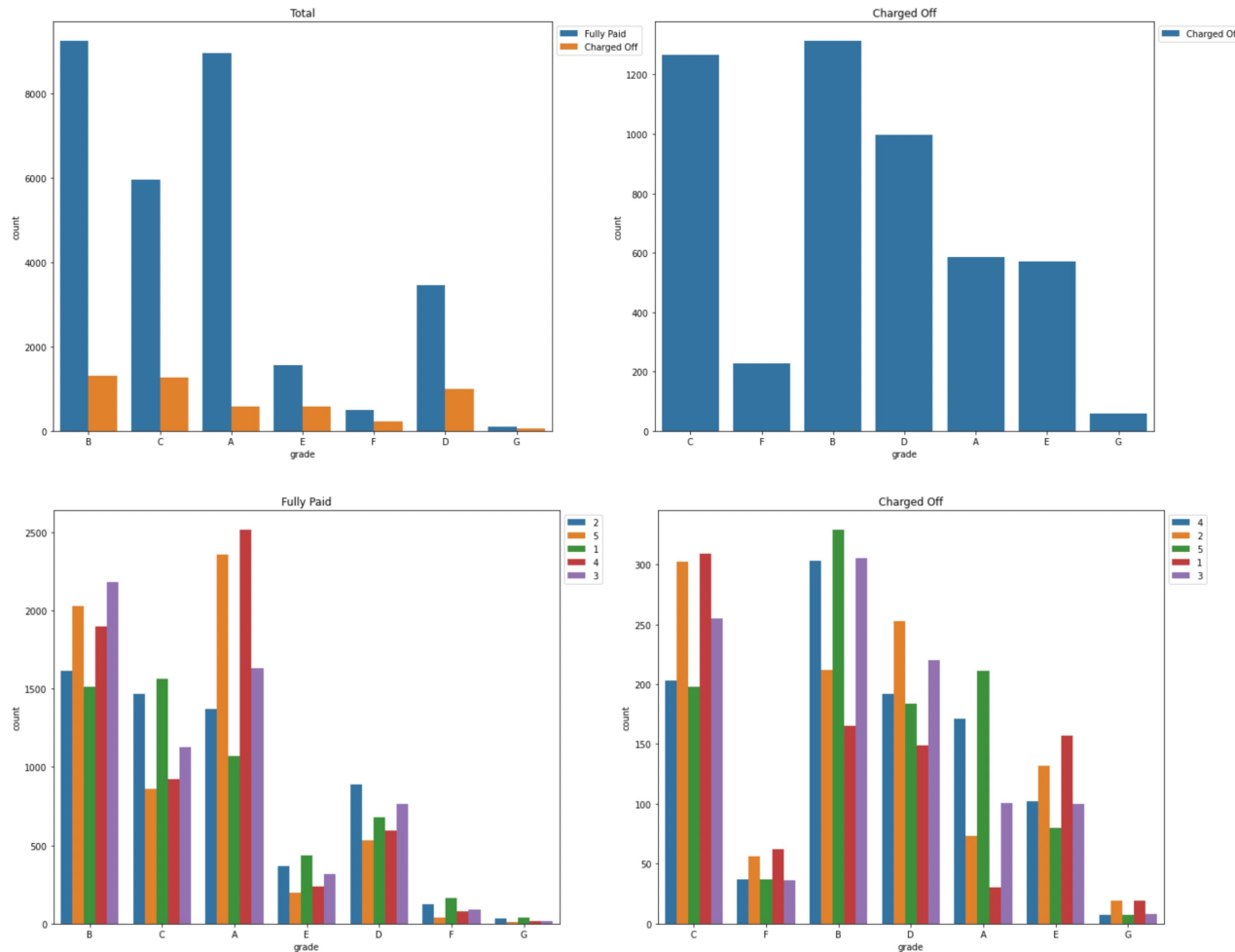Distribution of Employee length for Charged-off and Fully paid loan
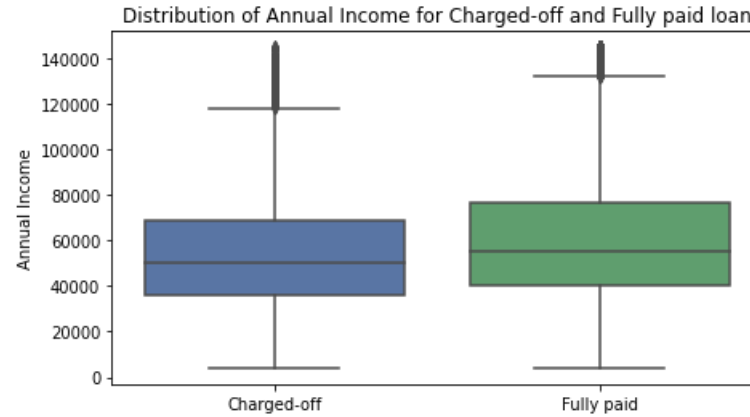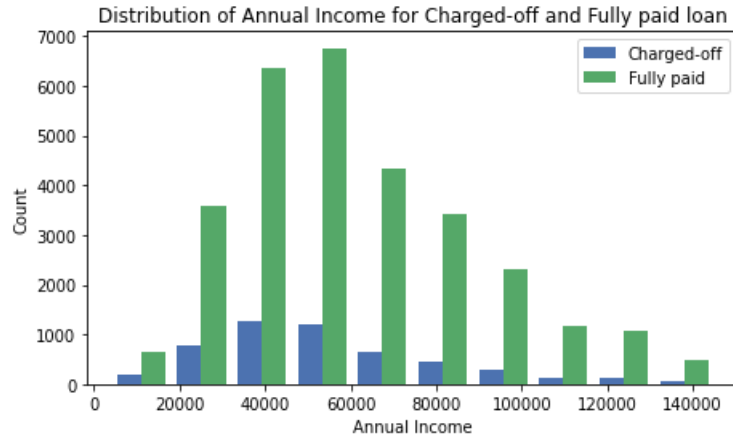
- Annual income for Charged-off is slightly lesser compared to fully paid loans
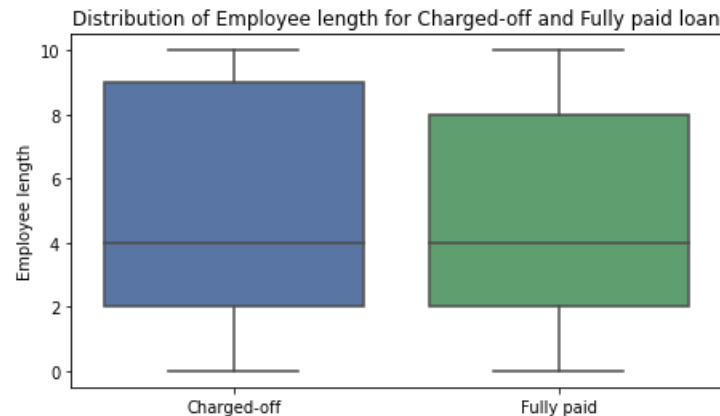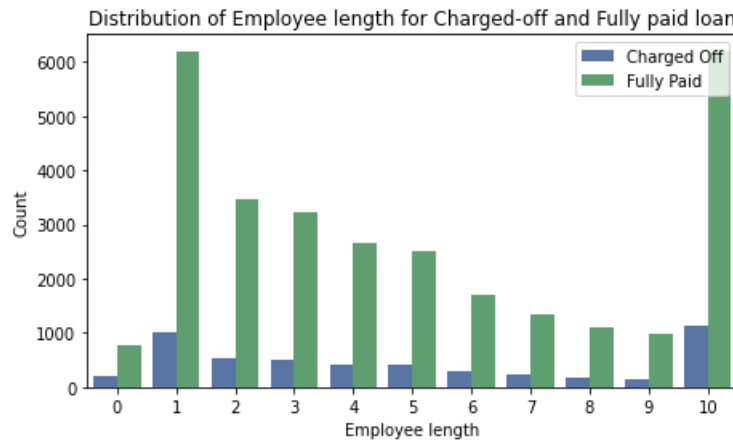
- Employee Length is slightly high for Charged-off when compared to Fully Paid

DISTRIBUTION ACROSS GRADE AND SUBGRADE

- **Grade B borrowers tends to default more**(may be most of applicants are graded as B) **but E,F,G grade borrowers have more ratio of defaulters, the risk is higher for these grades**

- **More number of defaulters are from B5 but groups D2,E1,F1,F2,G1,G2 have more ratio of defaulters, the risk is higher for these subgrades**

Disclaimer – Determination of grades by Lending club is unknown as of now

GRADE AND HOME OWNERSHIP DISTRIBUTION
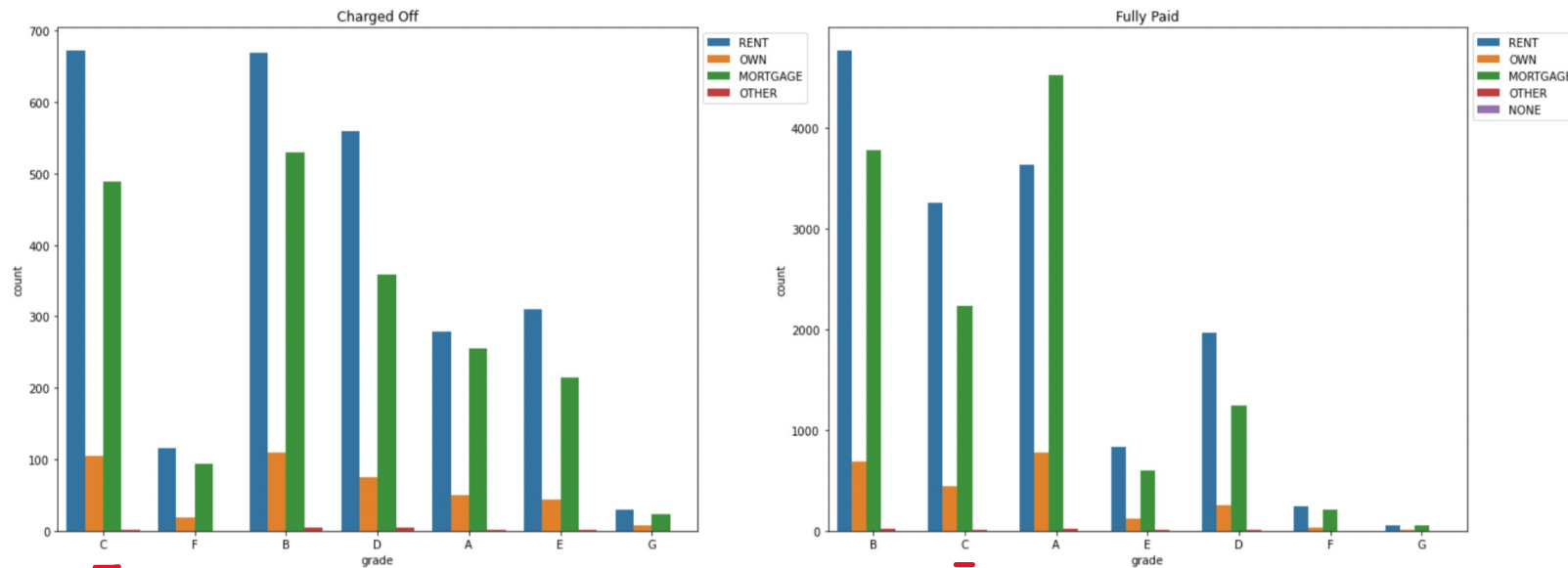
- **The C grade borrowers have high rate of defaulting across all home ownership, because C's peak is less when compared to A & B in Fully paid , but in Charged off compared to A & B, C has more defaulters**
- **In all grades Rented borrowers tend to default more**
- *Another interesting observation is that in the fully paid section only A grade borrowers have higher peak in home ownership Mortgage*

# DISTRIBUTION ACROSS ISSUE MONTH AND ISSUE YEAR
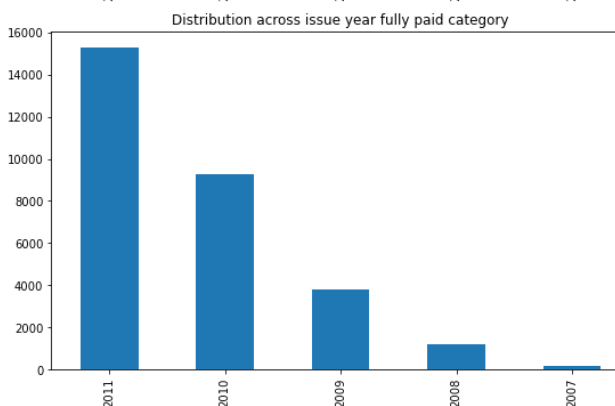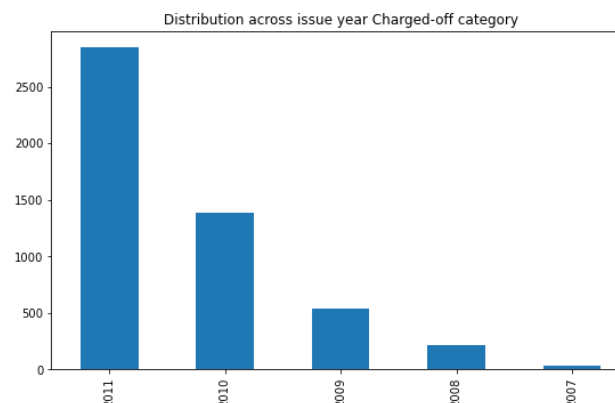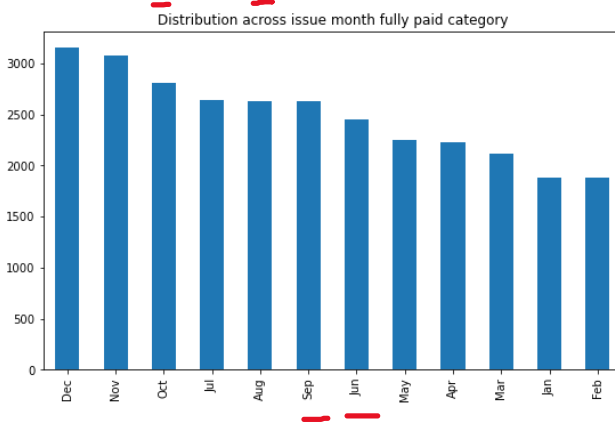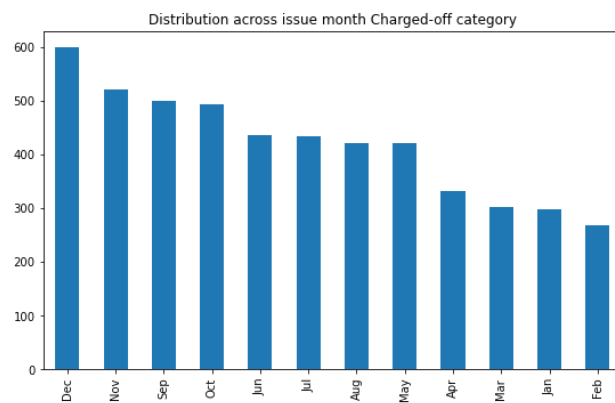
Issue month and issue year are the month and year on which the loans are issue

If you look at the graphs we could see that

- In general 2011 and Dec is the year and month having more default, also the total no of loan is in general more during this year and month
- Sept and June seems to be at the higher side in charged off chart and lower side at the fully paid chart which is contrary to other months where the order the almost same b/w charged off and fully paid
- **This could mean that loans applied on Sept and June have higher chance of being defaulted than other months**

INSTALLMENTS WITH RESPECT TO HOME OWNERSHIP
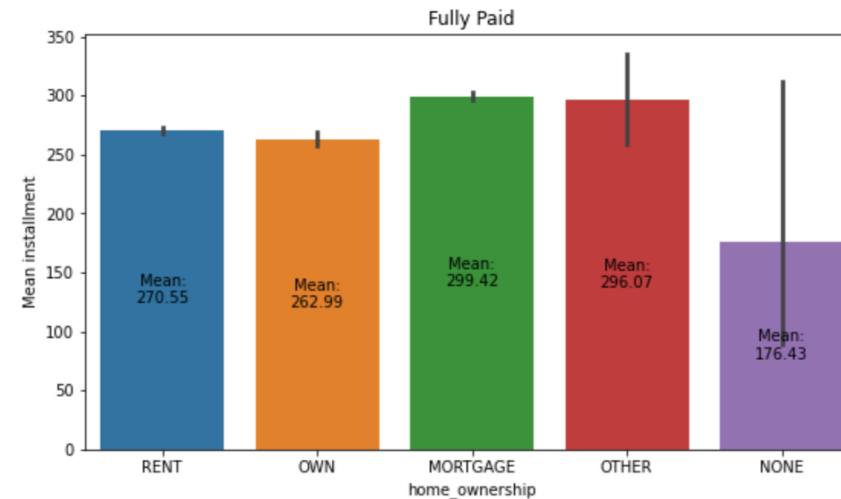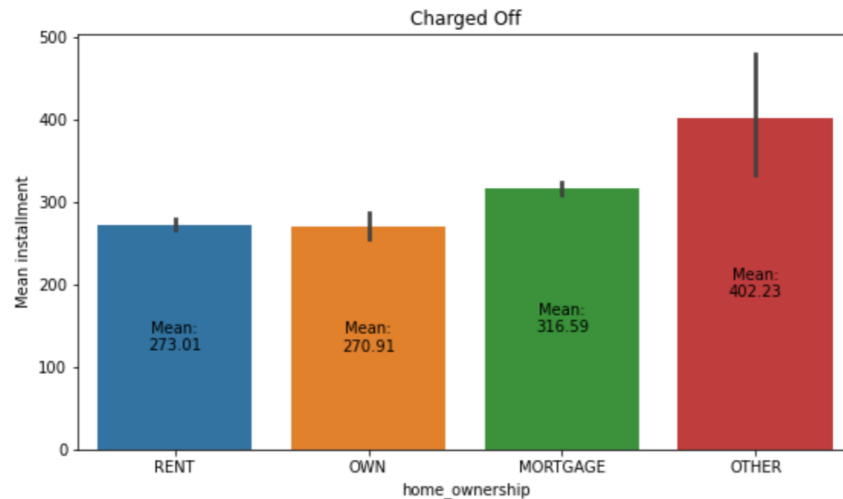
- Home ownership mentioned as OTHERS shows a huge difference in average of installment b/w charged off and fully paid
- **Higher the installment more risky the loan is for home ownership 'OTHER'**
- **Borrowers with MORTGAGE with higher installment tend to be defaulters**

DISTRIBUTION WITH ANNUAL INCOME AND PURPOSE

Count of Fully Paid in Annual Income vs Purpose

Count of Charged-off in Annual Income vs Purpose

Charged Off Rate
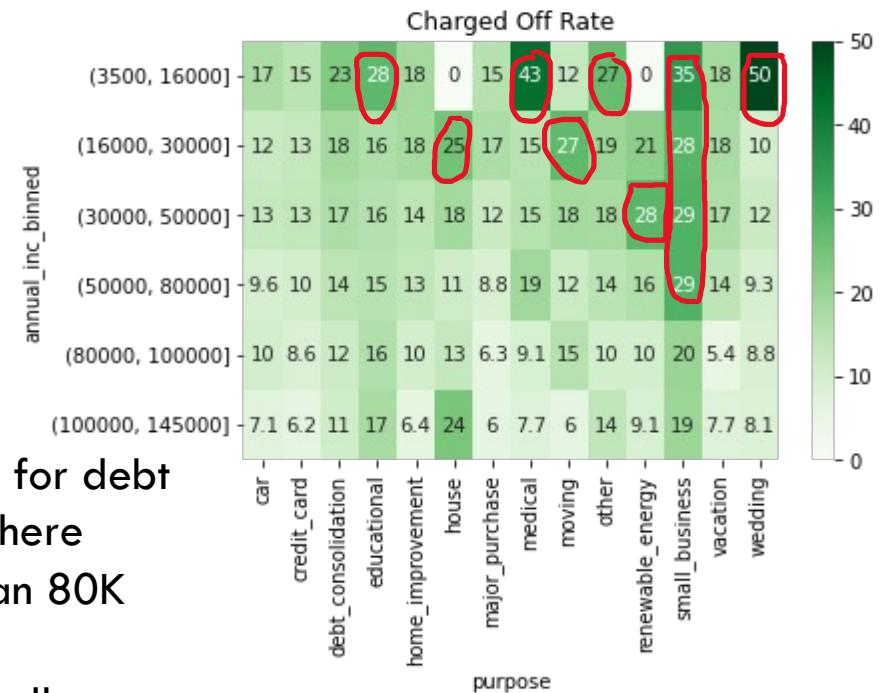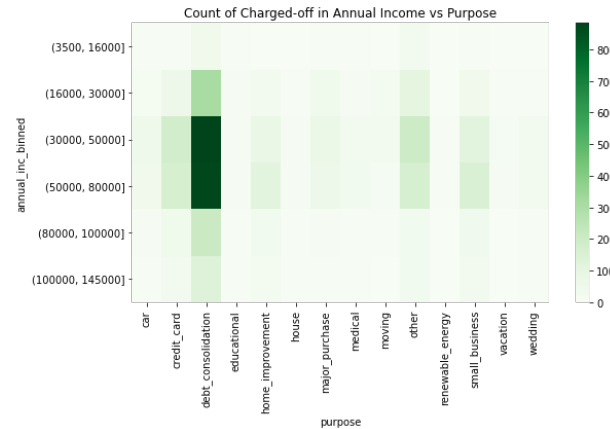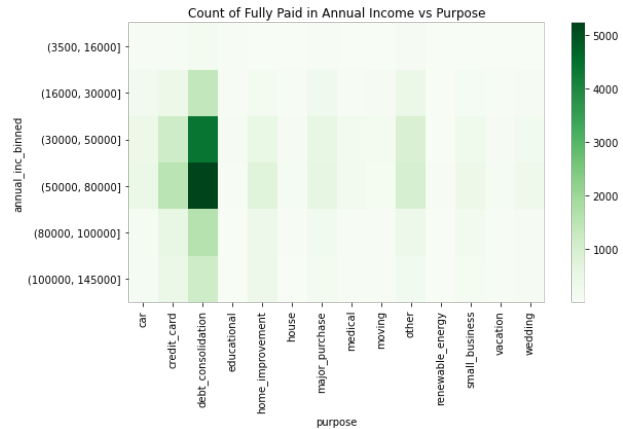
- Borrowers with an annual income of 30k-50k,50k-80k request loan for debt consolidation a lot and thereby the no of defaulters are also more here
- What we can see here is that borrowers with annual income less than 80K who request loan for small business have more 30% defaulters
- We can also observe such high rates in following combinations as well:
  - Educational - 4k-16k
  - Renewable energy - 30k-50k
- **Using this information it is possible to set a minimum limit to annual income to provide the loan based on the purpose, below which the loan would be considered risky**

Charged Off Rate graph shows
the rate of defaulters in each segment

RATE OF DEFAULTERS BASED ON ANNUAL INCOME AND HOME OWNERSHIP

Charged Off Rate
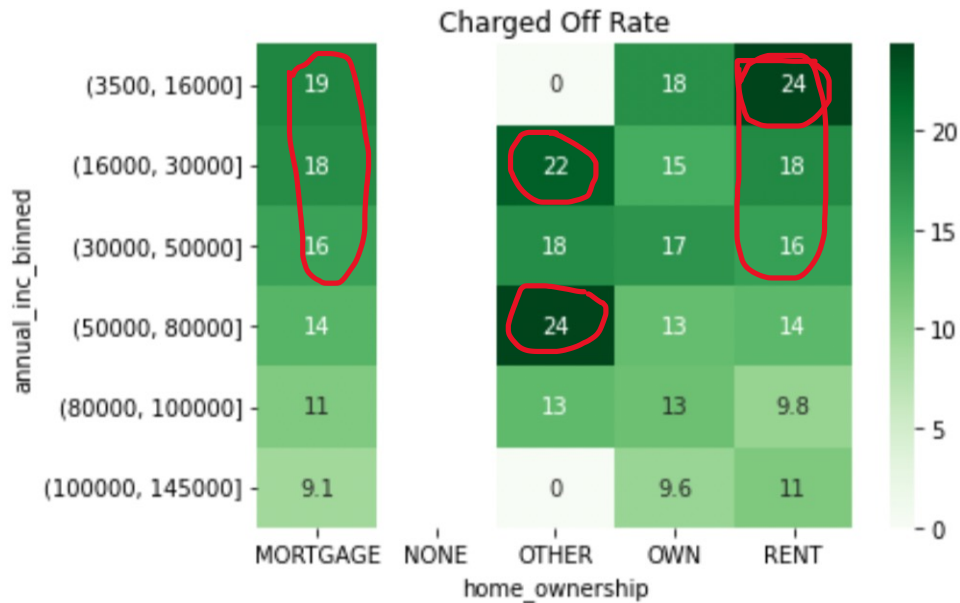
- Rate of defaulters is more (appx 24%) with home ownership rented and annual income 4K-16K
- The same is observed with home ownership 'other' and income 16K-30K & 50K-80K
- In the category rent and mortgage the chances of defaulting is more for borrowers less that 50K annual income with rates b/w 15-24

CO-RELATION* TABLE

Co-relation Matrix among different variables

|	id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	emp_length	annual_inc	dti	pub_rec_bankruptcies	issue_y
id	1	0.035	0.048	0.17	0.14	0.021	-0.0062	0.083	0.0033	0.093	0.006	0.85
loan_amnt	0.035	1	0.98	0.92	0.29	0.22	0.92	0.13	0.35	0.088	-0.021	0.033
funded_amnt	0.048	0.98	1	0.94	0.27	0.23	0.95	0.13	0.35	0.087	-0.02	0.046
funded_amnt_inv	0.17	0.92	0.94	1	0.3	0.22	0.88	0.14	0.33	0.098	-0.02	0.21
term	0.14	0.29	0.27	0.3	1	0.41	0.013	0.09	0.042	0.076	0.023	0.21
int_rate	0.021	0.22	0.23	0.22	0.41	1	0.21	-0.011	0.0097	0.11	0.095	0.0054
installment	-0.0062	0.92	0.95	0.88	0.013	0.21	1	0.1	0.35	0.073	-0.018	-0.025
emp_length	0.083	0.13	0.13	0.14	0.09	-0.011	0.1	1	0.2	0.055	0.054	0.088
annual_inc	0.0033	0.35	0.35	0.33	0.042	0.0097	0.35	0.2	1	-0.082	-0.0012	0.011
dti	0.093	0.088	0.087	0.098	0.076	0.11	0.073	0.055	-0.082	1	0.007	0.092
pub_rec_bankruptcies	0.006	-0.021	-0.02	-0.02	0.023	0.095	-0.018	0.054	-0.0012	0.007	1	0.016
issue_y	0.85	0.033	0.046	0.21	0.21	0.0054	-0.025	0.088	0.011	0.092	0.016	1

- We see positive co-relation of Funded amount by investor with Term, Interest Rate, Installments and Annual Income
- We see negative co-relation of Funded amount by investor with pub_rec_bankruptcies
- We also see positive co-relation between (term & interest rate), (Annual Inc & Installment, employee length)
- We also see negative co-relation between (employee length & interest rate), (Annual Income & dti, pub_rec_bankruptcies)
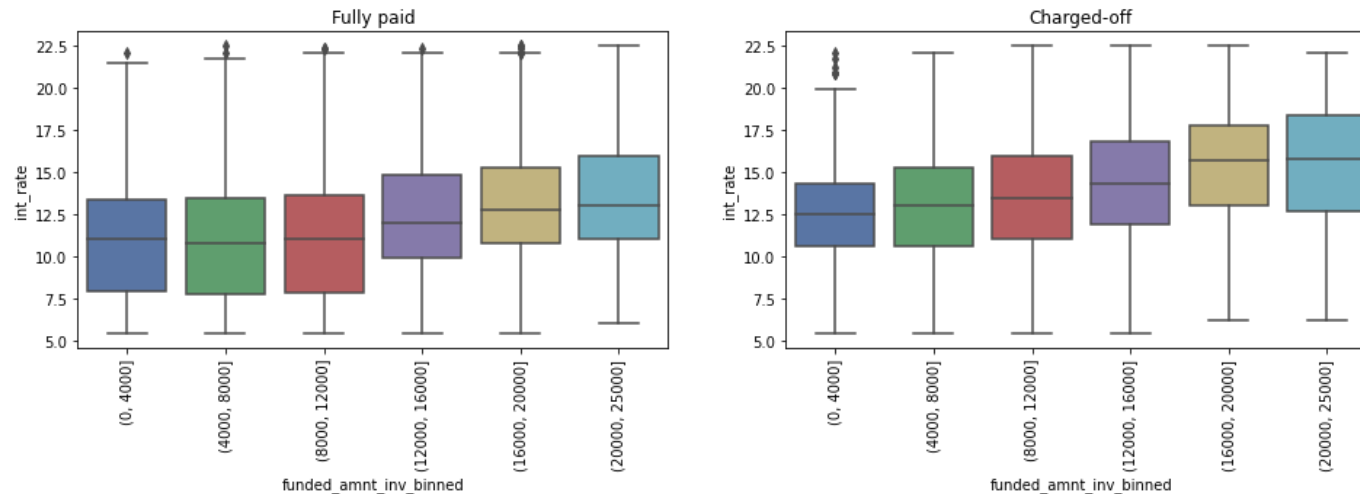
* a mutual relationship or connection between two or more things

BEHAVIOR OF FUNDED AMOUNT BY INVESTOR WITH OTHER KEY VARIABLES

TERM
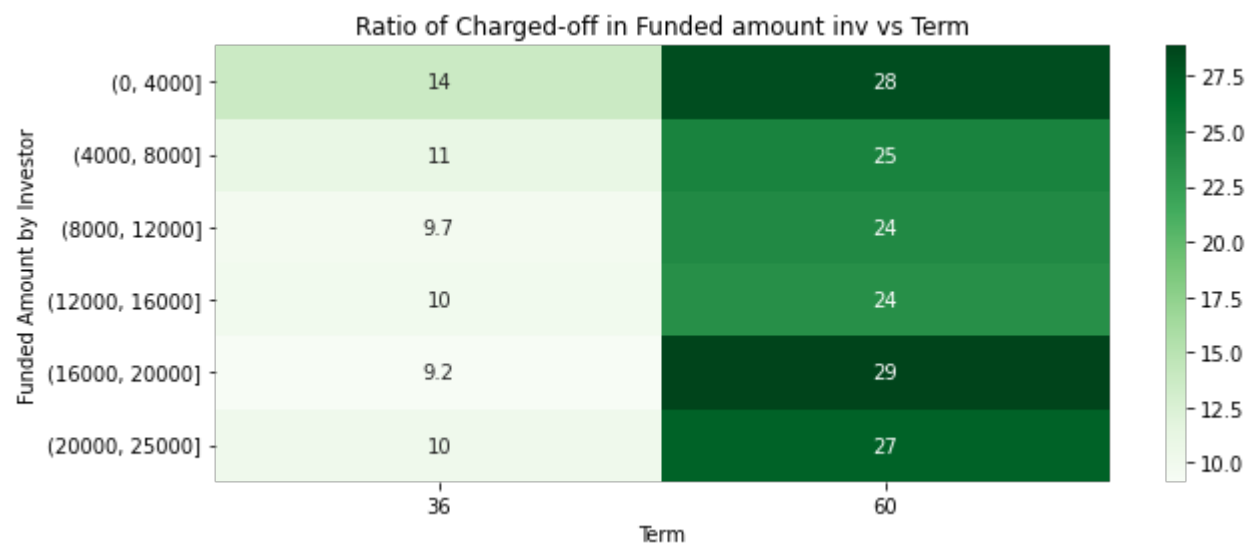
Ratio of Charged-off in Funded amount inv vs Term

- **Borrowers with 60 months Term tends to default more especially with Funded amount by investor in range of 0-4K and 16K-20K**

INTEREST RATE

- From the first graph we can infer that as the loan amount increases, the interest rate also increases and the interest rate for Charged-off is higher than paid off

- **From the second graph it is more evident that borrowers with higher interest rate tends to default more especially with Funded amount by investor in 16K to 25K and Interest rate as 16-20%**
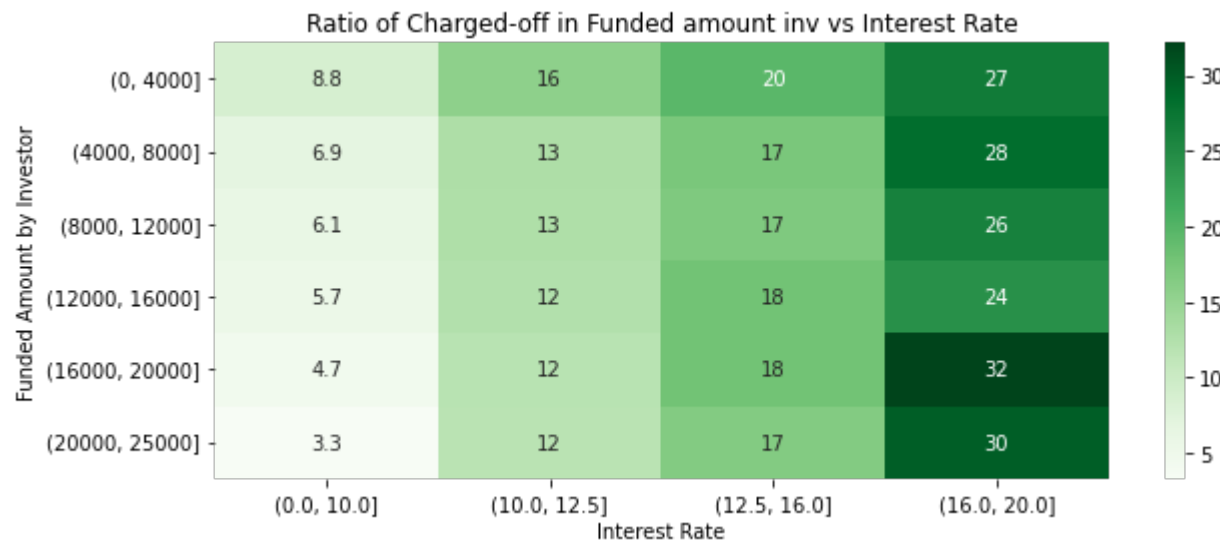
GRADE
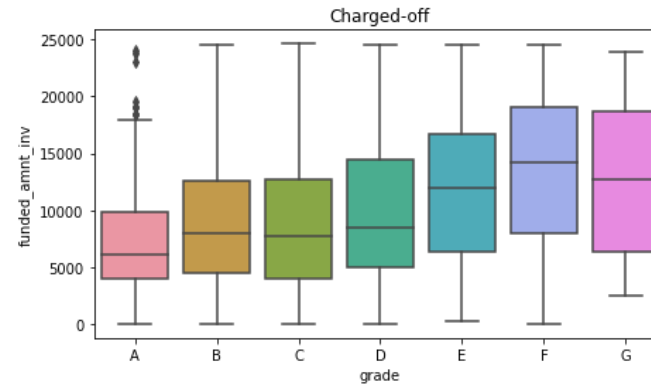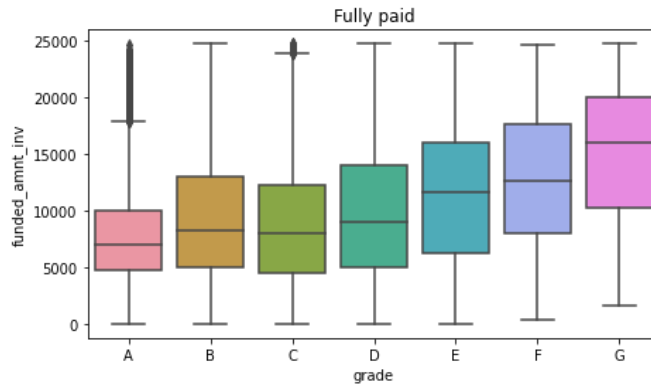
- From the first graph we can infer that for grade F – Defaulters are more towards higher amount and for grade G – Defaulters are higher towards lower amount

- From second graph it is more evident that **40% of borrowers with grade F in amount range of 20K-25K and 61% of borrowers in Grade G in amount in 4K-8K have defaulted**

EMPLOYEE LENGTH

Fully paid

Charged-off
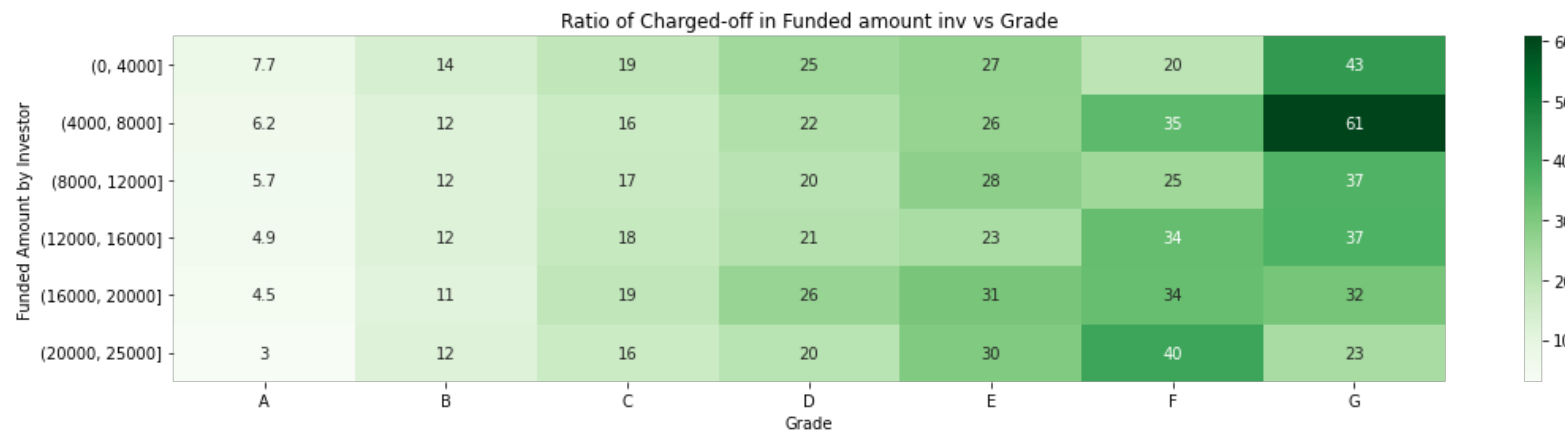
Ratio of Charged-off in Funded amount inv vs Employee Length
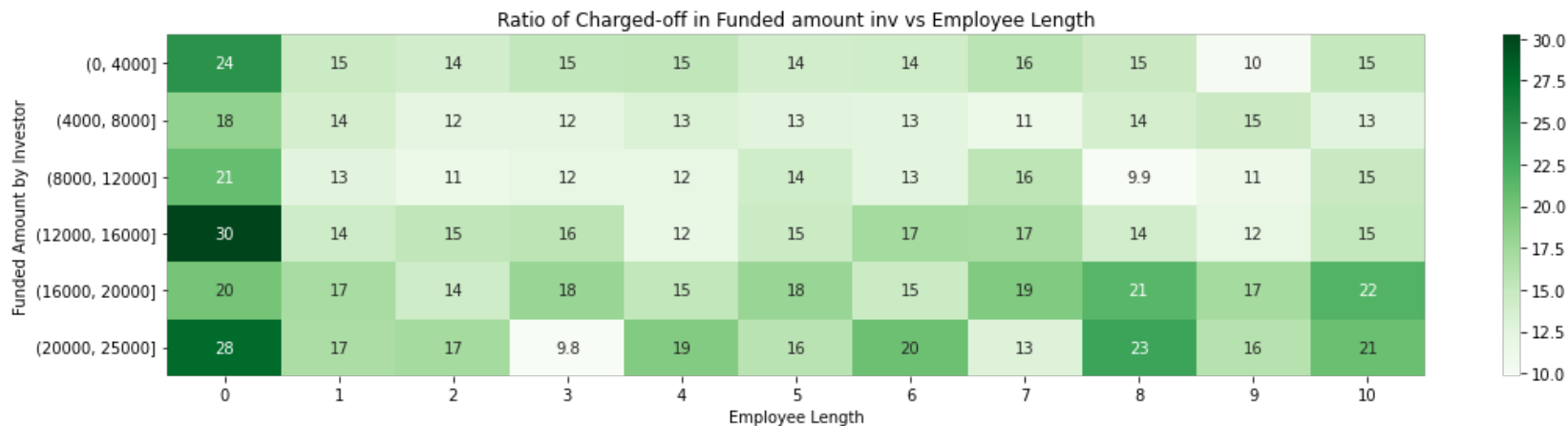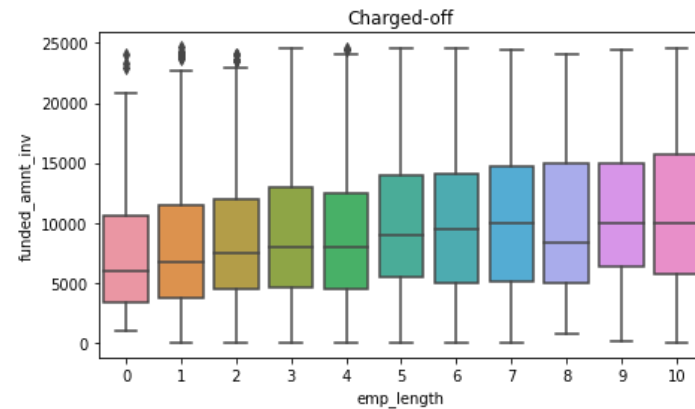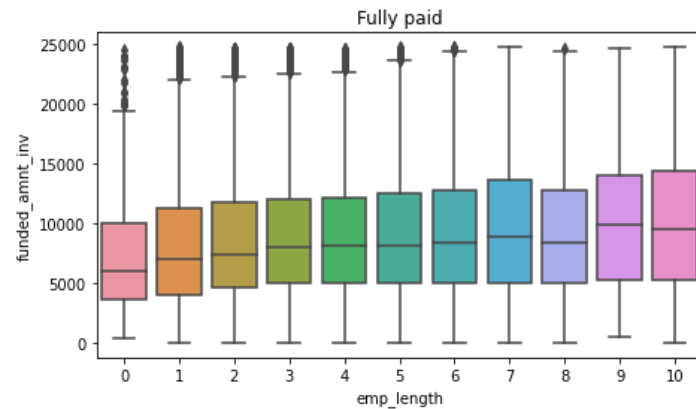
- From the first graph we can infer that the borrowers with employee length 8 tends to default more compared to employee length 9.

- From second graph it is more evident by looking at the ration percentage for 8 and 9 years.

- Also in addition we can infer that the **borrower with Employee Length of 0 years tend to default more for all amount range and borrower with Employee Length of 6,8,and 10 years tend to default with Funded amount by investor in 16K – 25K range**

HOME OWNERSHIP
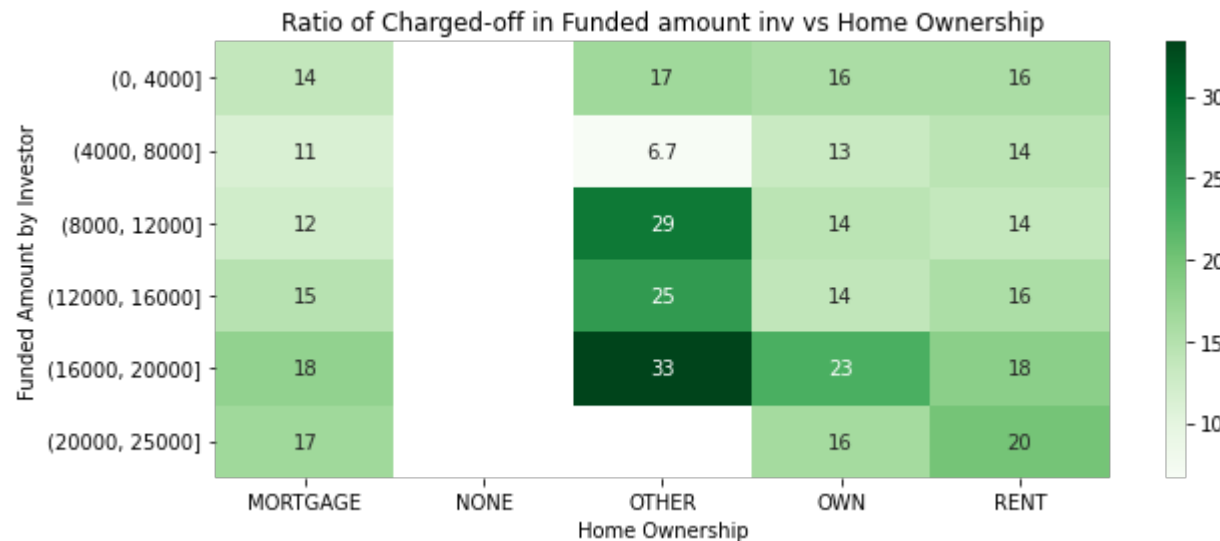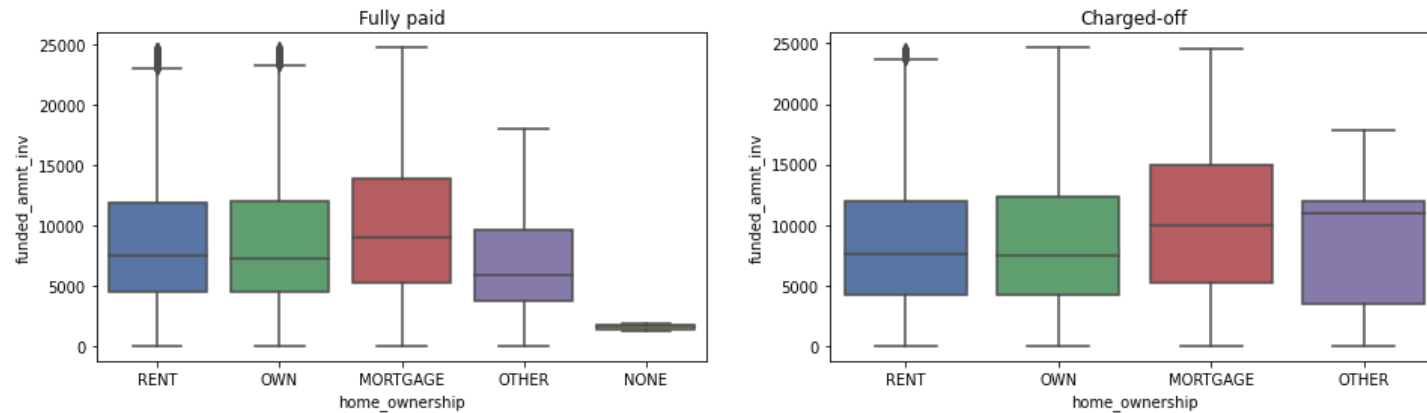
- From the first graph we can infer that the borrowers with Home ownership as Other tend to default more.

- From second graph it is more evident that **borrower with House Ownership as 'Others' tend to default more for Funded amount by investor in 8K-20K range**

- **Borrower with House Ownership as 'Own' tend to default more for Funded amount by investor in 16K-20K range**

- **Borrower with House Ownership as 'Rent' tend to default more for Funded amount by investor in 16K-25K range**
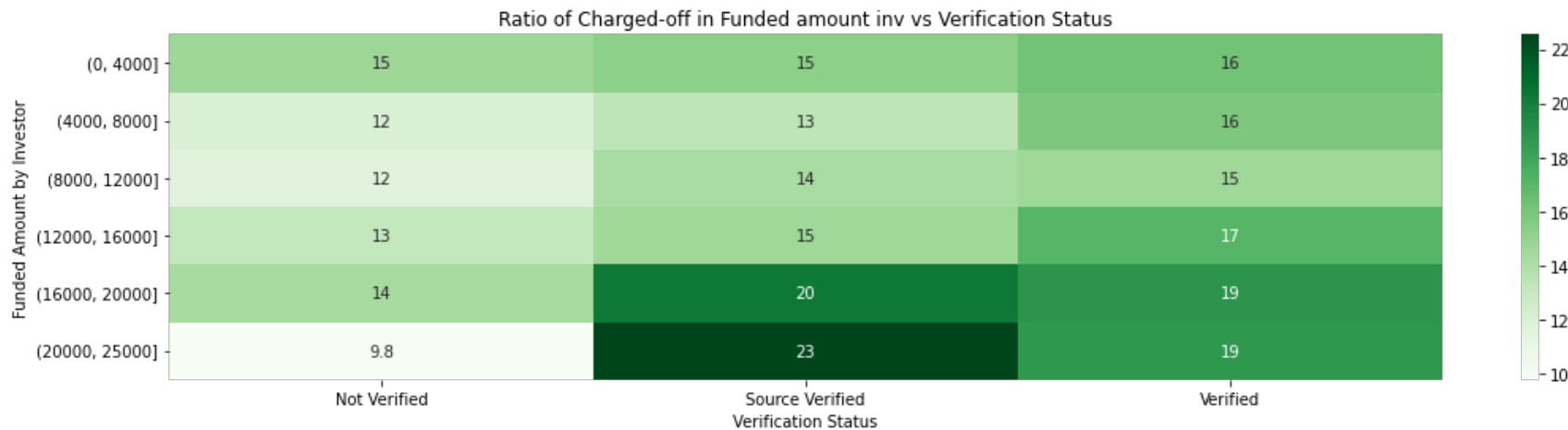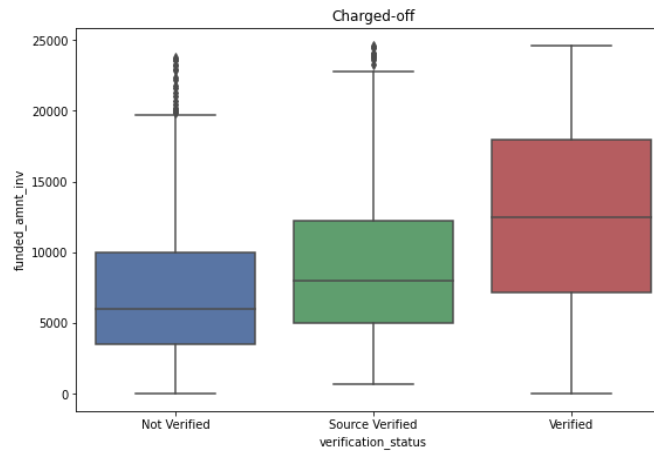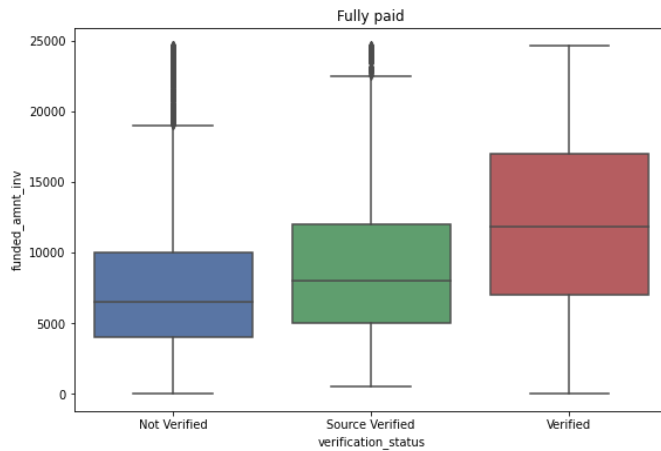
ANNUAL INCOME
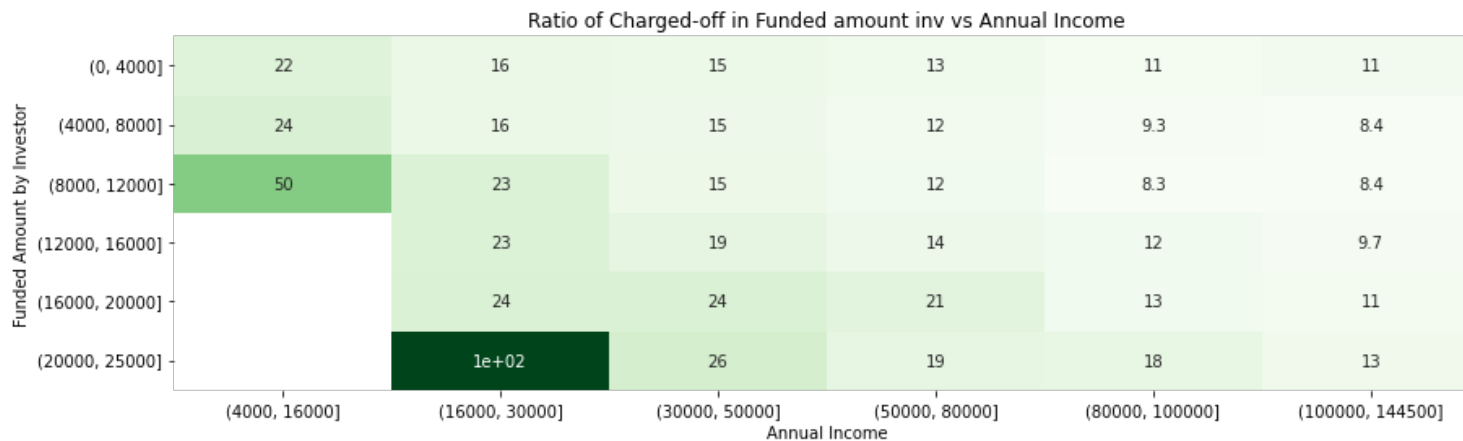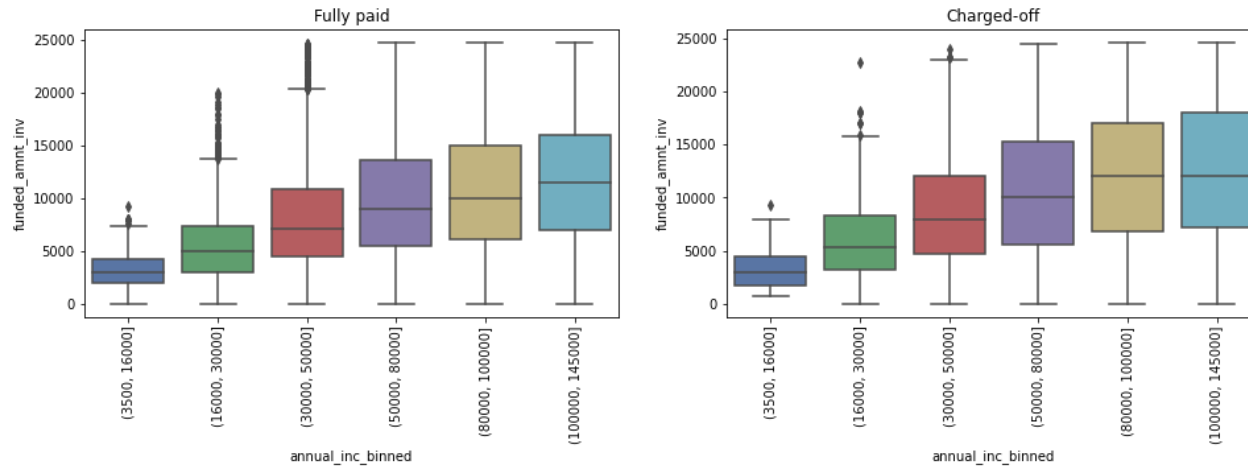
- From the first graph we can infer that as the annual income increase, the amount also increases and so does the percentage of defaulter

- From second graph it is more evident that **borrower with Lower Annual income and higher Funded amount by investor tend to default**
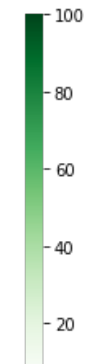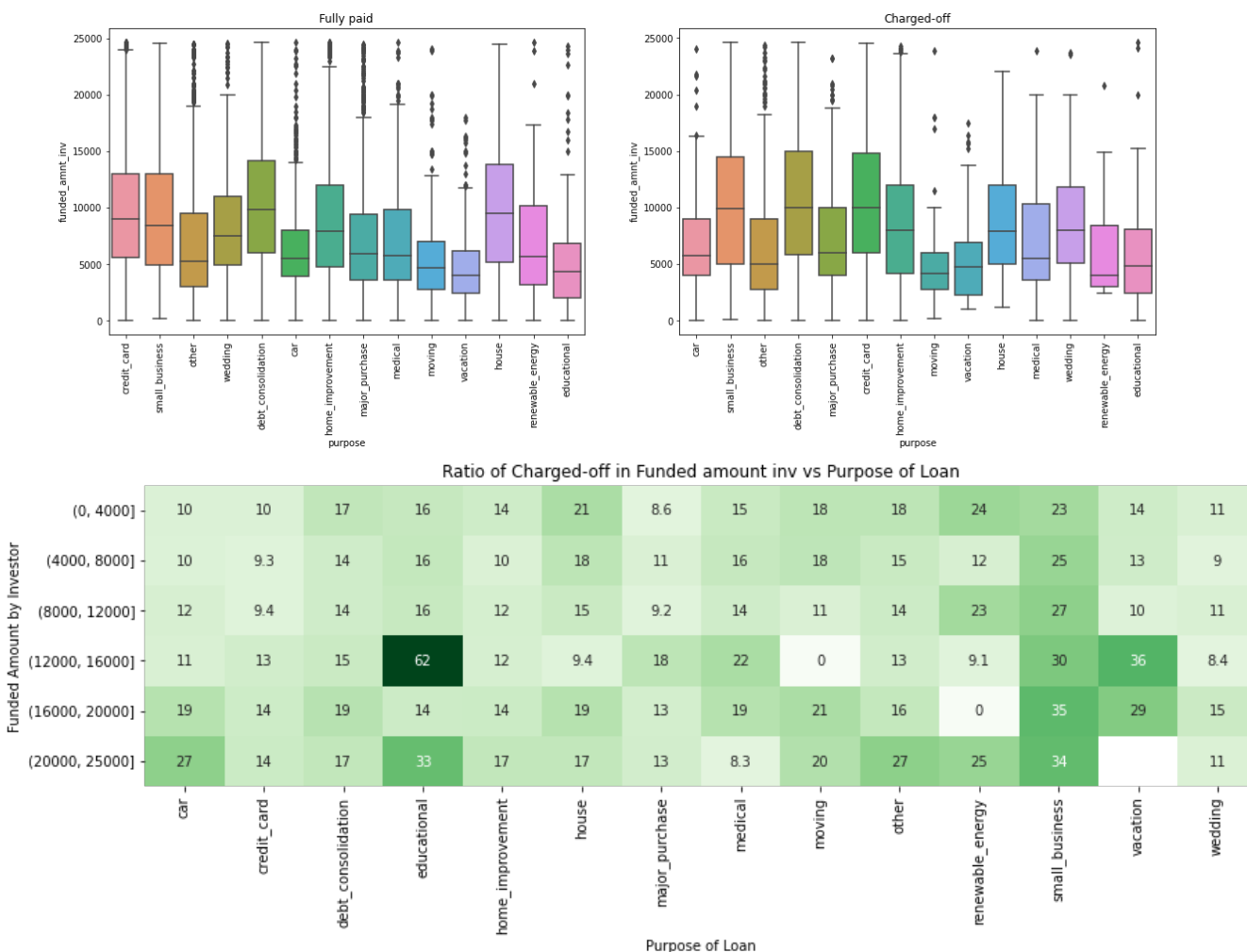
VERIFICATION STATUS

- From the first graph we can infer that the number of defaulters are more in the order of "Not Verified", "Source Verified", "Verified"

- From second graph it is more evident that **borrower with Verification status as 'Source Verified' and 'Verified' tends to default more with Funded amount by investor in 16K-25K**

PURPOSE OF LOAN

- From the first graph we can infer that the average Funded amount by investor is higher for small business, Vacation and Education

- From second graph it is more evident that **borrower with Education loan tends to default more with Funded amount by investor in 12K-16K and 20K-25K range**

- **Borrower with Vacation loan tends to default more with Funded amount by investor in 12K-20K range**

- **Borrower with Small Business and Car loan tends to default more with Funded amount by investor in 16K-25K range**

DTI

Fully paid / Charged-off

Ratio of Charged-off in Funded amount inv vs DTI
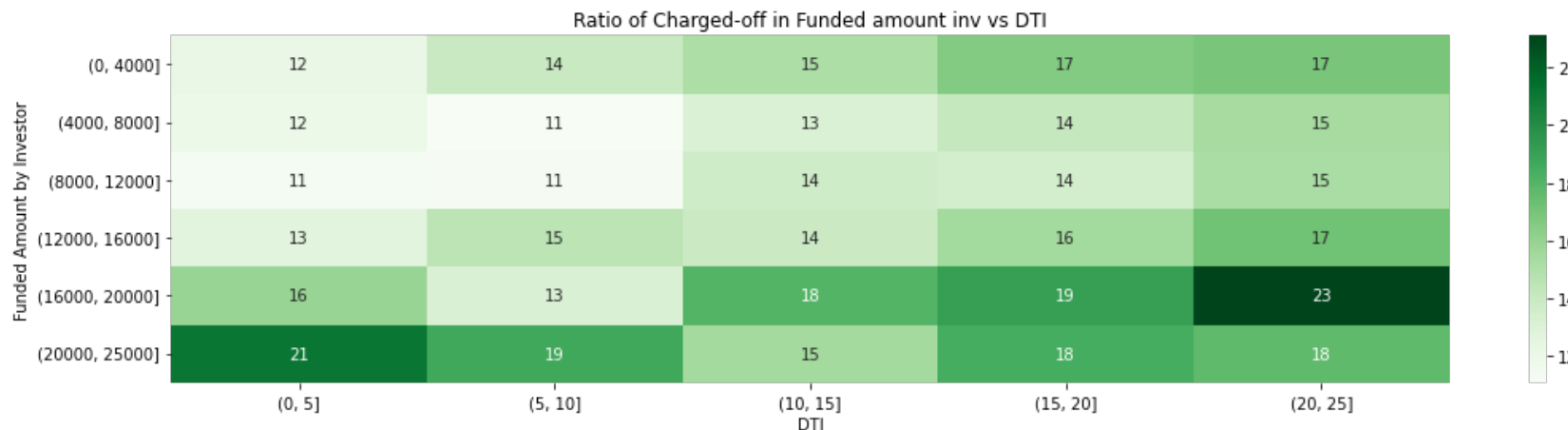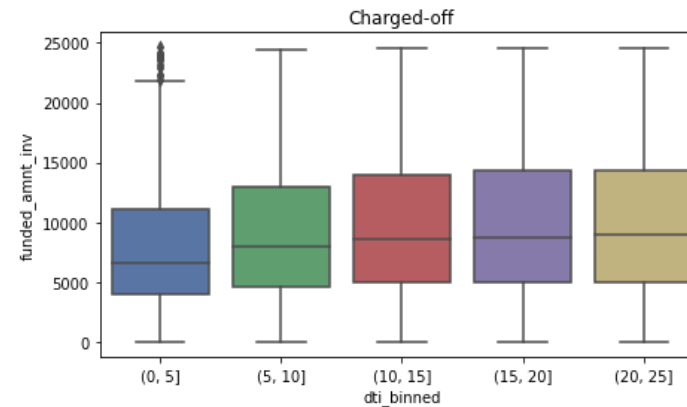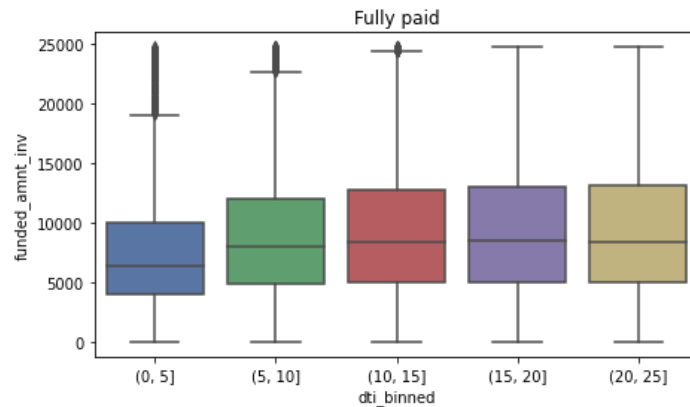
- From the first graph we can infer that the DTI is slighter higher across all amount range in defaulters

- From second graph it is more evident that **borrower with DTI in 20-25 tends to default more with Funded amount by investor in 16K-20K range**

- **Borrower with DTI in 0-10 tends to default more with Funded amount by investor in 20K-25K range**

CONCLUSION

Based on our analysis following conclusions can be derived:

Grades & Sub Grades

- The C grade borrowers have high rate of defaulting across all home ownership
- In all grades Rented borrowers tend to default more
- Subgrades D2,E1,F1,F2,G1,G2 have more ratio of defaulters, the risk is higher for these subgrades

Interest rate and DTI against funded amount by investor and annual income

- Upper limit for DTI in each income category/funded amount by investor can be determined, more than that would be considered risky
  (*check 'dti with respect to annual income/funded amount by investor' slide for more details*)
- Upper limit for interest rate in each income category/funded amount by investor can be determined, more than that would be considered risky
  (*check 'dti with respect to annual income/funded amount by investor' slide for more details*)

Installments & Home Ownership

- Higher the installment more risky the loan is for home ownership 'OTHER'
- Borrowers with MORTGAGE with higher installment tend to be defaulters

Annual Income & Purpose

- Borrowers with annual income less than 80K who request loan for small business have more 30% defaulters, hence considered to be risky
- We can also observe such high defaulters rates in following combinations as well:
  - Educational - 4k-16k
  - Renewable energy - 30k-50k

CONCLUSION

Funded amount by investors and other key variables

- Borrowers with 60 months Term tends to default more especially with Funded amount by investor in range of 0-4K and 16K-20K
- Borrowers with higher interest rate tends to default more especially with Funded amount by investor in 16K to 25K and Interest rate as 16-20%
- Borrowers with grade F in amount range of 20K-25K and borrowers with Grade G in amount in 4K-8K have high default rate
- Borrowers with Employee Length of 0 years tend to default more for all amount range and borrowers with Employee Length of 6,8,and 10 years tend to default with Funded amount by investor in 16K - 25K range
- Borrower with House Ownership as 'Others', 'Own' & 'Rent' tend to default more for Funded amount by investor in 16K-25K range
- Borrowers with Lower Annual income and higher Funded amount by investor tend to default
- Borrower with Verification status as 'Source Verified' and 'Verified' tends to default more with Funded amount by investor in 16K-25K
- Borrowers with Education loan tends to default more with Funded amount by investor in 12K-16K and 20K-25K range
- Borrowers with Vacation loan, Small Business and Car loan tends to default more with Funded amount by investor in 12K-25K range
- Borrower with DTI in all ranges tends to default more with Funded amount by investor in 16K-25K range