Lending Club Case Study

Group Members:

Anns Rahim
Sankalp Gupta

About Lending Club

- Lending Club is a financial institution that provides loans for various purposes like car loan, home loan, education loan, etc
- Lending Club has collected historical lending data of about 40,000 loans disbursed over a period of 5 years from 2007 to 2011
- Lending Club wishes to analyse this data to have a better understanding of loan defaults so that it can make an informed decision about giving out loans in the future
- This Presentation shares the major observations made after analysing the above data, observing the trends in the above data, and noting some features that will be of certain interest to the owners of the Lending Club financial institution.
- Based on our observation of the peculiar features, we have come to some recommendations that we will suggest in the conclusion of this presentation

Hierarchy of problem solving methods

Data Analysis

Data Analysis

Data Analysis

Data Cleaning

Removing the null valued columns, unnecessary variables and checking the null value percentage and removing the respective rows.

Data Understanding

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

Univariate Analysis

Analysing each column, plotting the distributions of each column.

Segmented Univariate Analysis

Analysing the continuous data columns with respect to the categorical column

Bivariate Analysis

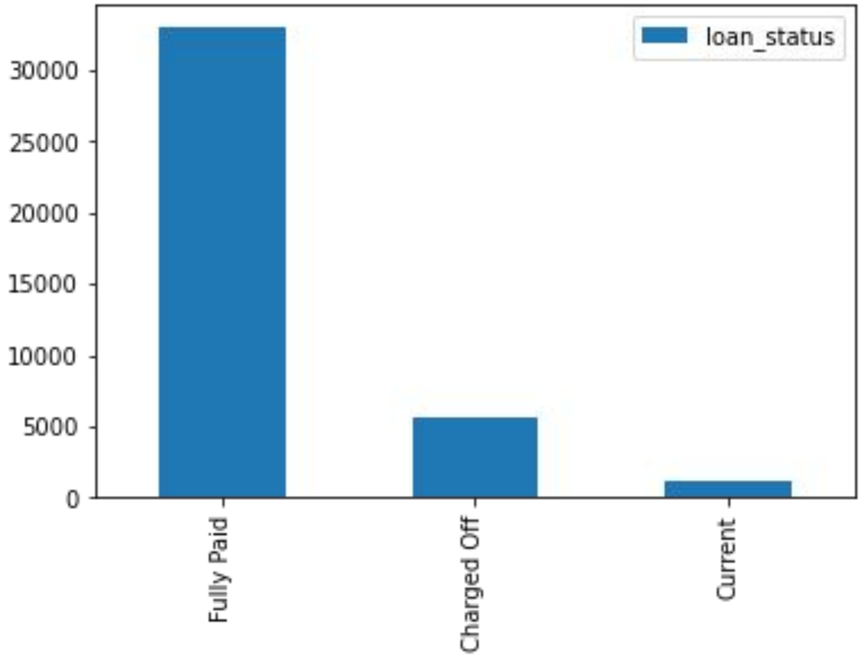
Analysing the two variable behaviour like term and loan status with respect to loan amount.

Recommendations

Analysing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.

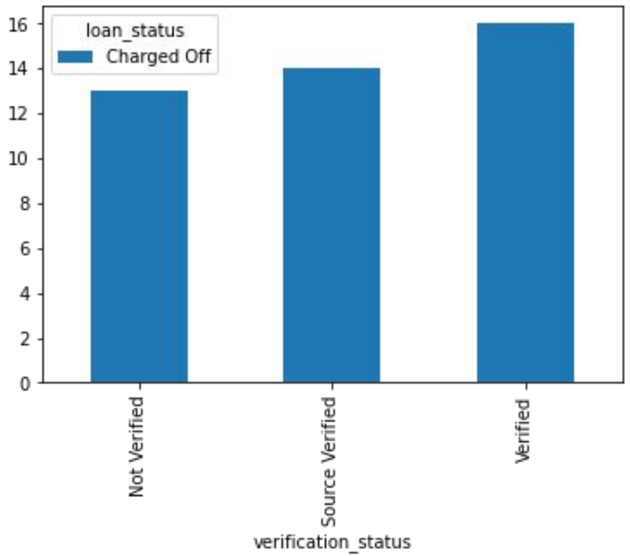
Data Analysis

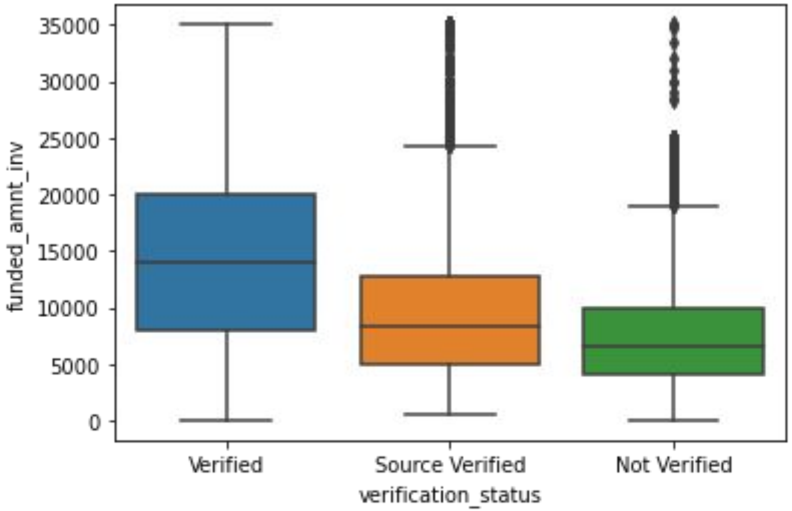
Charged Off Loans



- **We can see that about 14% loans are Charged Off**
- **And almost 82% loans are Fully paid**
- **2% loan in current**

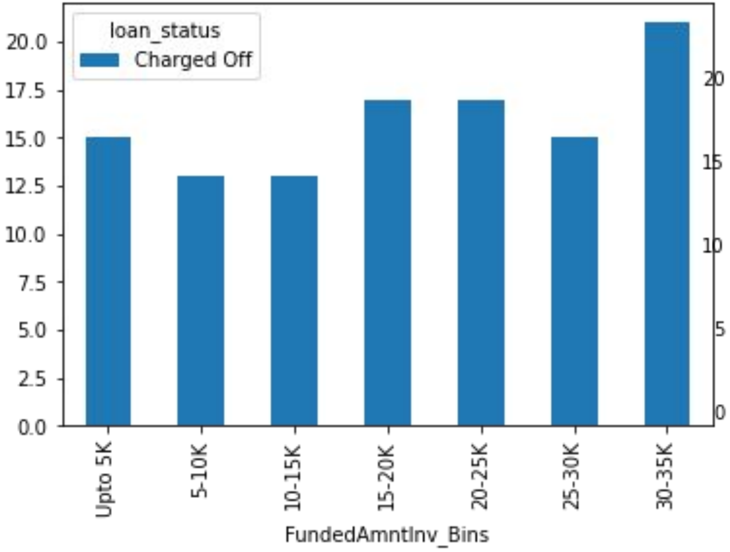
Verification on Loan

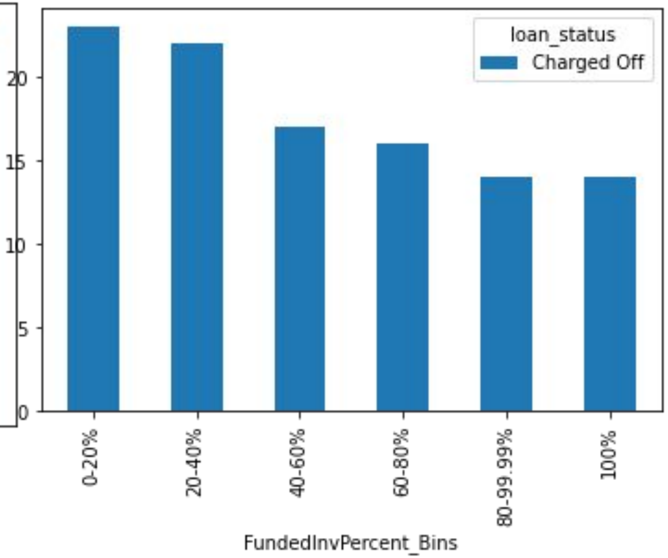




- **Nearly 43% loans disbursed are not verified. That's huge**
- **The Charge Off rate is higher in verified Loan**
- **it becomes clear when loan amount is compared with verification status on a box plot. Clearly unverified loans are those that have lower loan value, hence lesser chances of default**

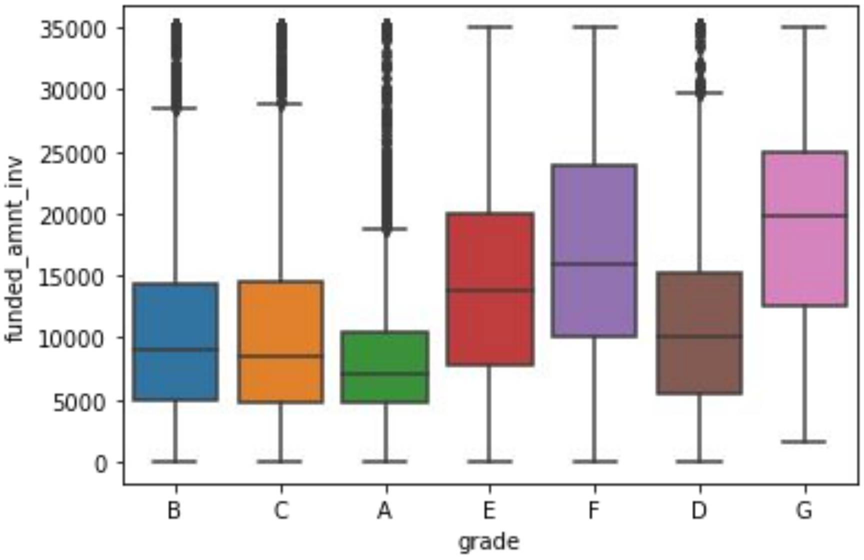
Funded Amount

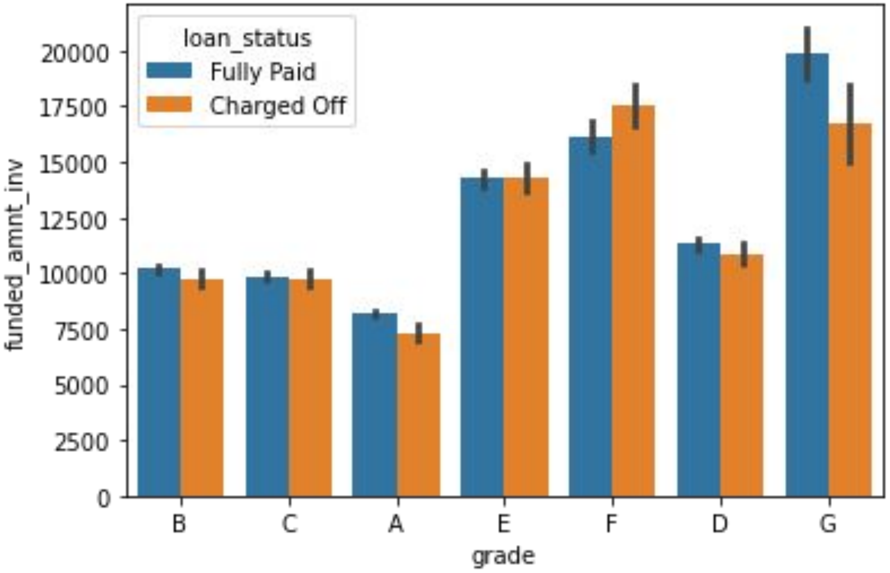




- **Charged Offs (Defaults) are happening on slightly higher value loans compared to Fully Paid**
- **Most of the current loans are of higher value than both fully paid and charged off loans**
- **There is some trend that higher the loan, higher the defaults**
- **But a bigger trend is that if loan is partially funded, then the chances of default are higher**
- **Observation: Its better to reject a loan completely, than to fund it partially**

Loan Grades

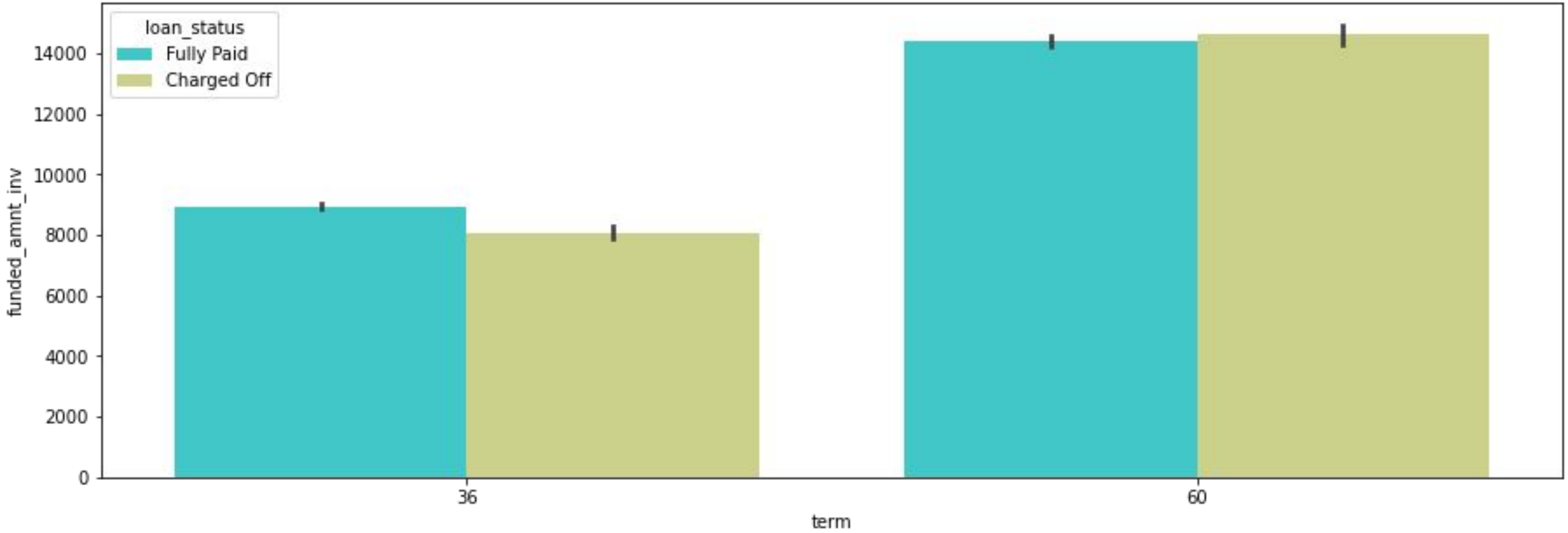




- **Grade F and G has the higher average of loan amount**
- **For Grade A the low average loan amount is disbursed**

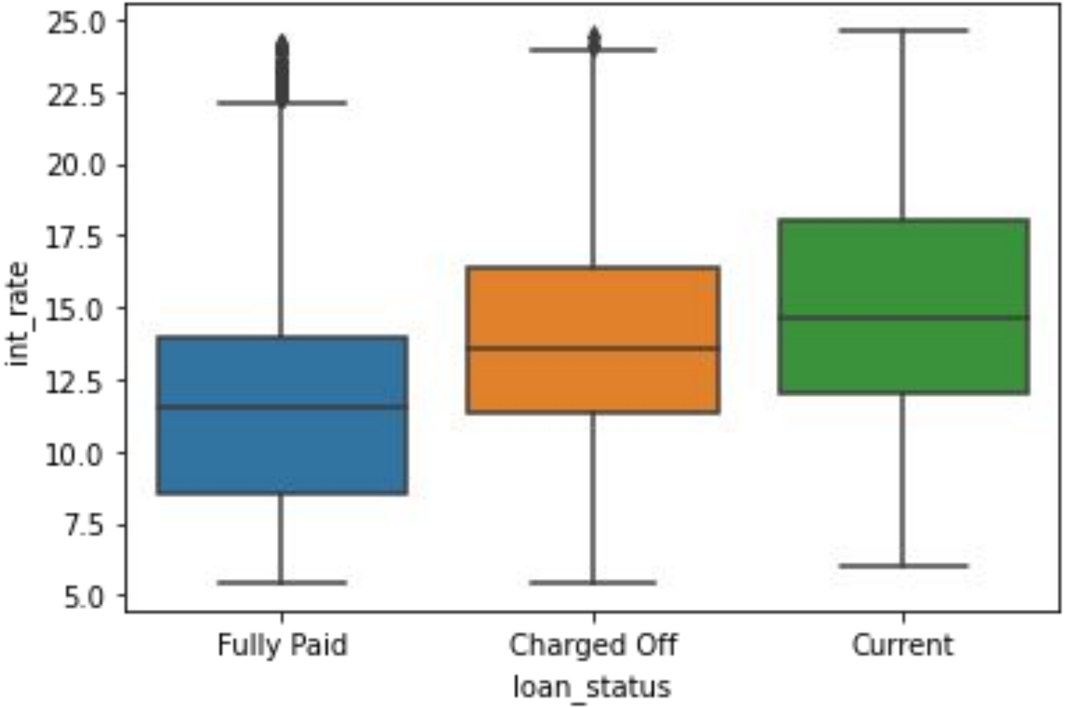
- **Grade D,E,F,G has high risk of Charge Off**

Term of loan



- **The term is of for 3 years and 5 years**
- **The Loan amount is higher for 5 years term**
- **As well as the charge off is higher for the 5 year term**
- **So the term increases the chances of charge off risk is higher**
- **LC team should consider giving loans for a shorter tenure**

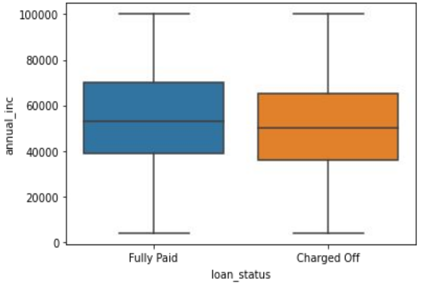
Interest rate



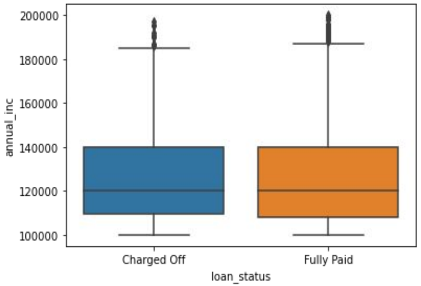
- **Most current loans are also higher value loans, hence attracting a higher rate of interest. Looking at the trend of defaults, these could turn out to be potentially high risk loans**
- **Clearly, higher the interest rate, higher chances of default. But since interest is a effect of higher risk rather than a cause of higher risk, it is clear that its a Catch-22 situation and has to be handled by LC team on a careful basis**

Annual income

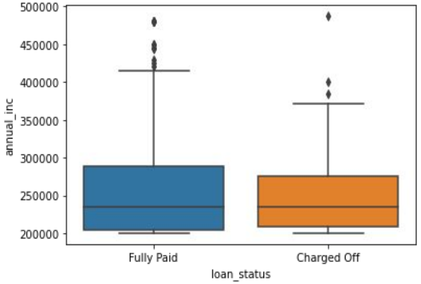
Below 100k



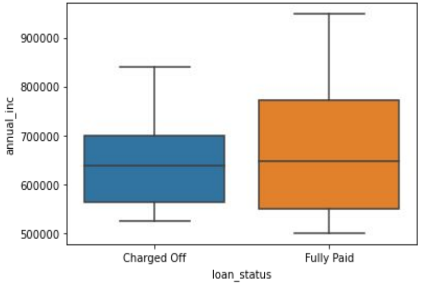
100k to 200k



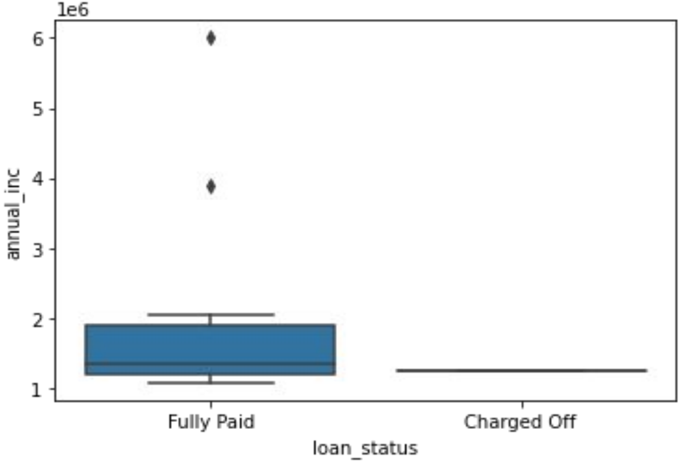
200k to 500k



500k to 1M



Above 1M

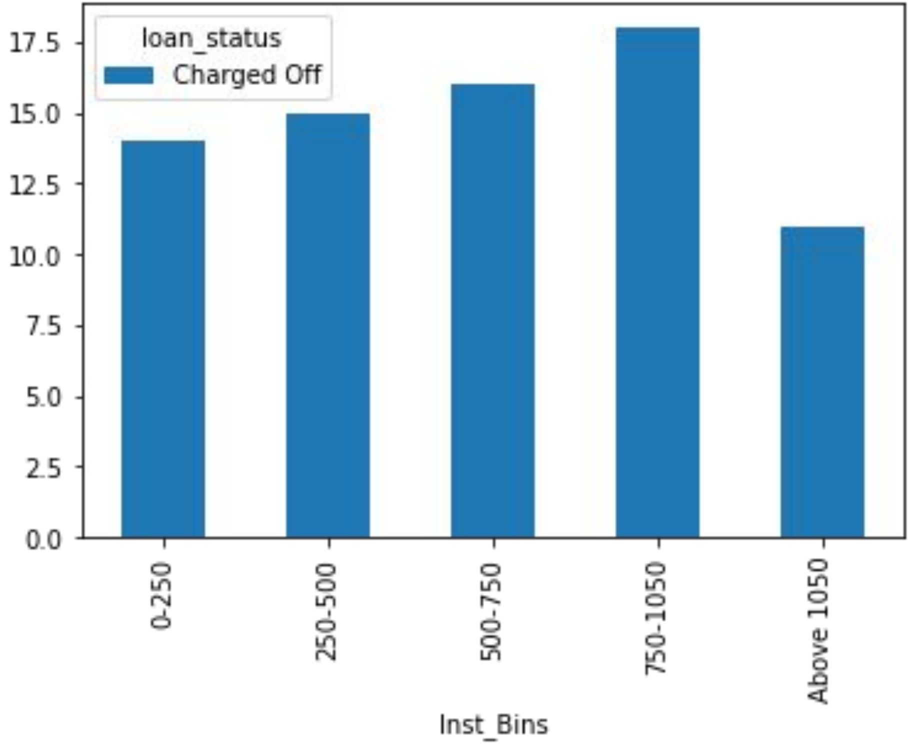


- **Income Below 100K: Defaults are happening at lower income levels**
- **100K -500K: Defaults are similar to Fully Paid**
- **500K - 1M: Defaults happening at higher income levels**
- **Above 1M : No defaults happening**

CONCLUSIONS:

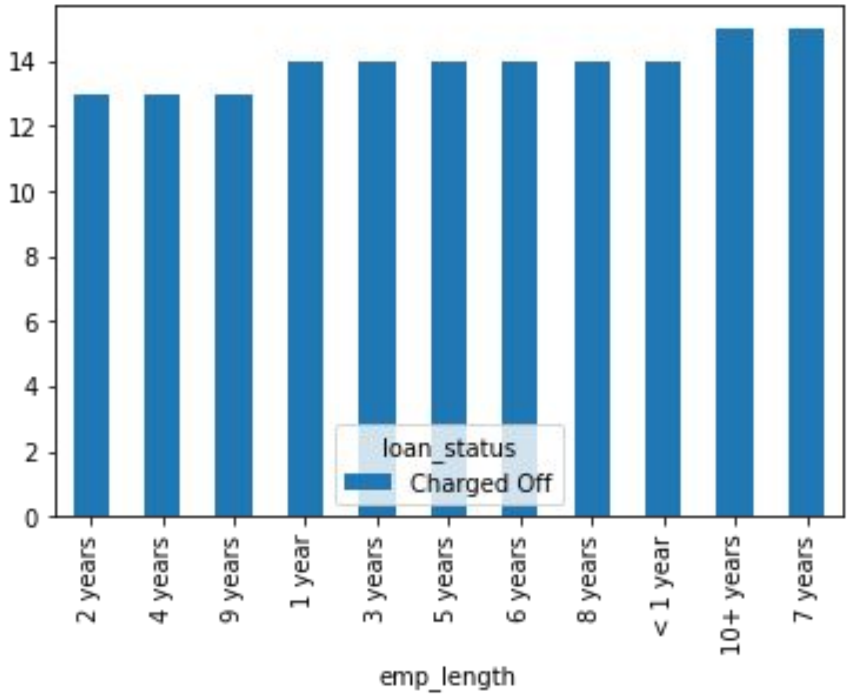
- **LC team should consider incomes between 100-500K and above 1M as safer compared to incomes below 100K and between 500K to 1 Million**

Installment



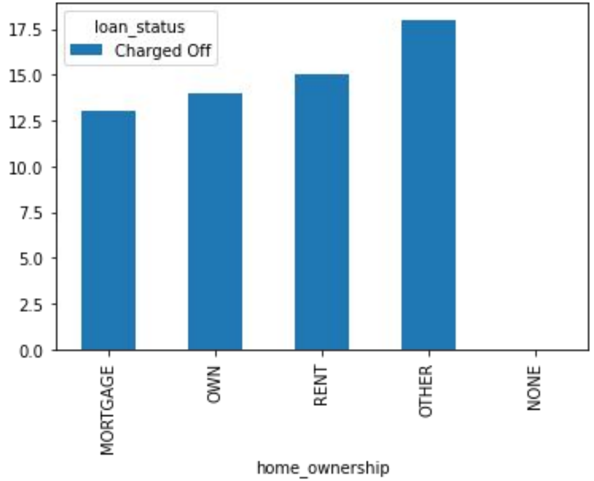
- **More people are defaulting as installment increases, BUT for installments above 1050, defaults are the lowest**
- **Higher Grade (High Risk) Loans are also the ones with higher installments.**
- **Installments are based on interest rate and tenure. Lower tenure OR higher interest rate means higher installment, which could also mean higher risk.**
- **We know from analysis of Term and Interest Rate, that lower term and Interest rate are less risky. But lower interest rate means lower profits too**
- **Its best to ignore this, as interest rate and tenure are better attributes for decision making**

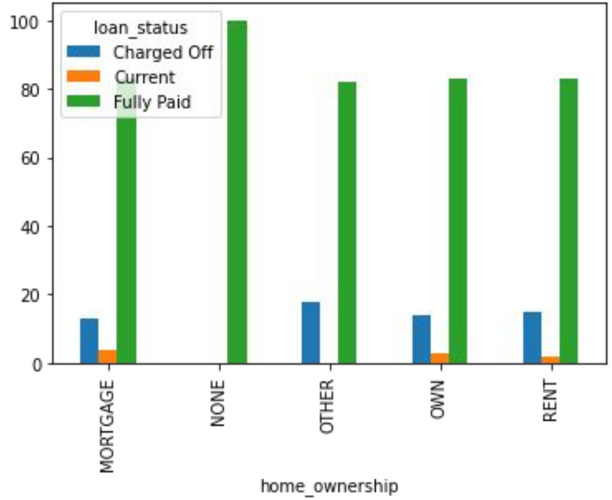
Employee length



Maximum defaults are in categories where employment lenght is 7 yrs and 10+ years, whereas minimum is 2yrs, 4yrs and 9 yrs

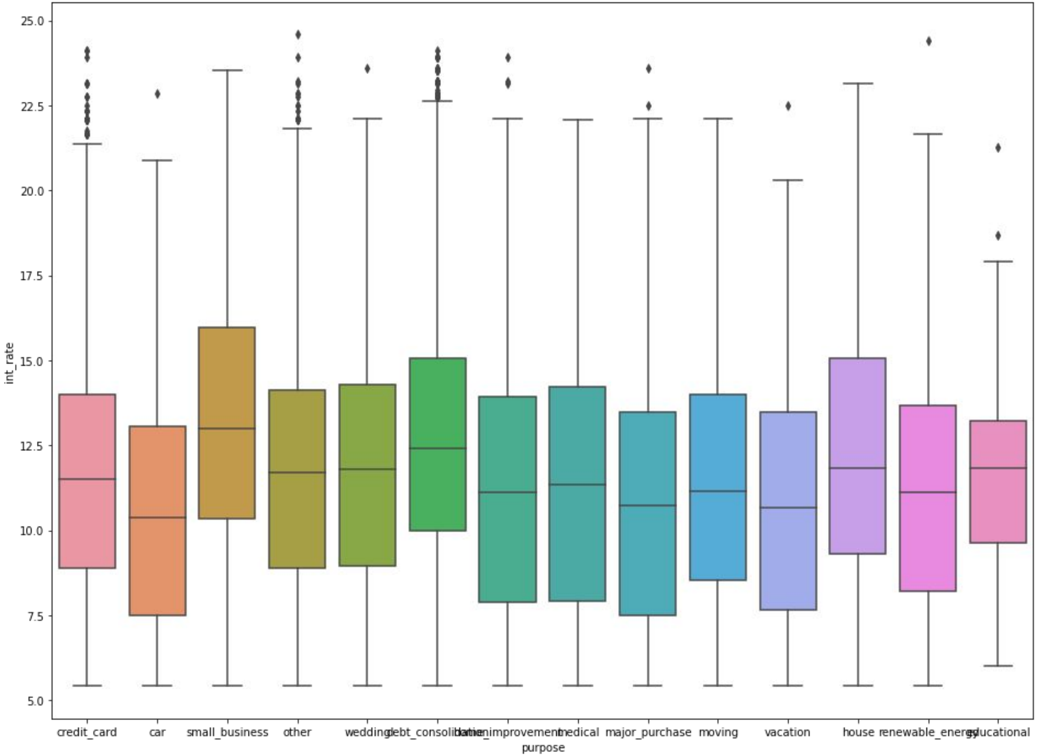
Home ownership





- **Maximum defaults are in categories where Home Ownership is mentioned as OTHER**
- **Home Ownership of NONE has no defaults. After analysis of such cases compared to Loan Amount and Annual Income, it turns out that these are people who have middle income levels and have taken very low amount of loan, hence their defaults are literally ZERO**
- **LC team should be careful giving loans to people who Home Ownership as OTHER**

Purpose of loan

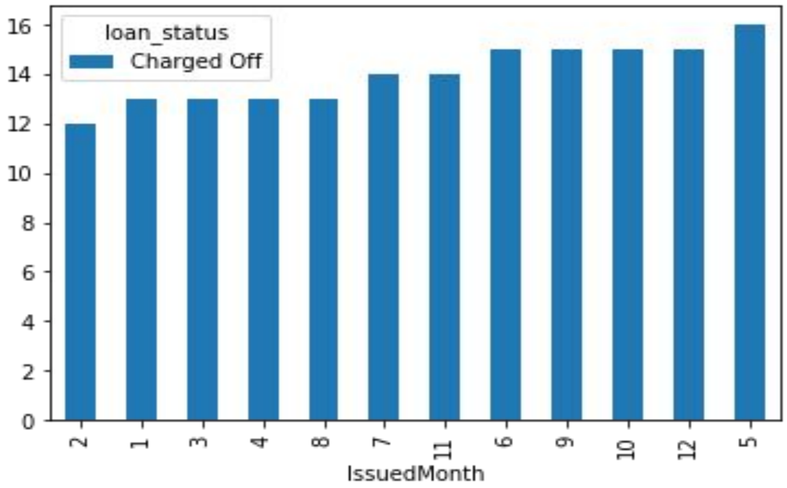


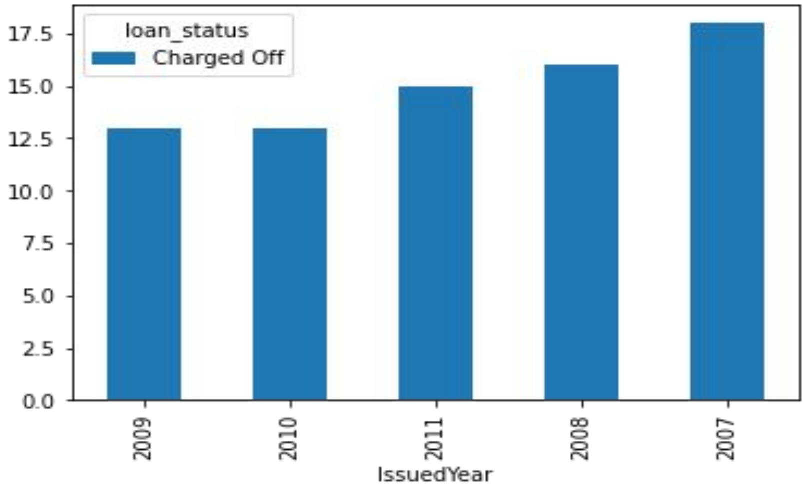
- **Small business, Renewable Energy and Education loan has the highest rate of charge off**
- **Renewable Energy and Education loan seem to have lower interest rate, indicating that they are cheaper. Perhaps this is due to government policies to promote these loans. Perhaps, this is also the reason, that there are more defaults in these two**

Observation :

- **LC should be wary of small business loans. Also, Renewable Energy and Educational Loan**
- **LC can consider setting interest rate based on the loan purpose, particularly education and renewable energy loans**

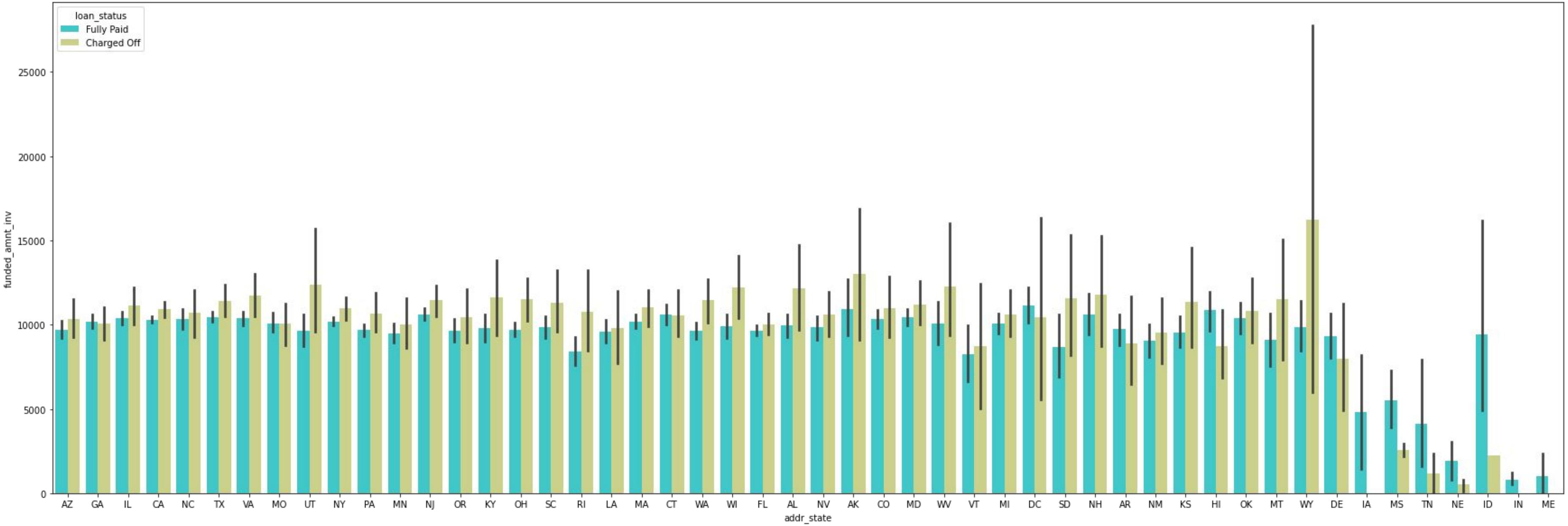
Loan Issued Date





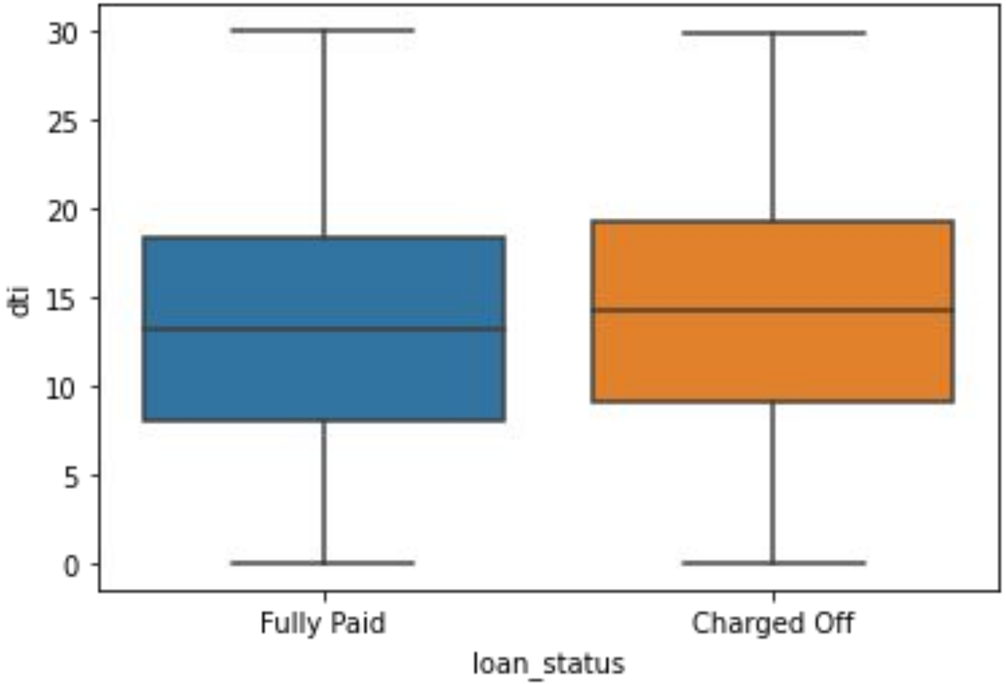
- **Most of the disbursed on the month of may**
- **From 2007 to 2011 the charge off is higher in 2007**
- **No discernible trend in terms of defaults, which can be used to predict future course of action**

Loan Location (State)



- **The State WY (Wyoming) has the highest disbursal of loan amount**
- **And the Charge Off rate is also higher**
- **So the WY state has high risk of Charge Off**

Debt to Income



- **Debt to income(dti) : A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations**

- **When the debt to income rate raises the chances of charge off risk aslo increases**

Conclusion / Recommendations

Low Risk Loans

- Disbursement Amount = Loan Requested
- Tenure is shorter, i.e. = 36 months
- Interest Rate is Lower
- Annual Income is between 200-500K
- Home Ownership is either of OWN, RENT or MORTGAGED or NONE
- Employment Length is between 2-8 yrs, but not 7 years
- Public Records and Bankruptcies are low
- Purpose is Car Loan, Major purchase or Wedding
- State / Zip Code is ....
- Number of Open Accounts is ....
- Debt to Income Ratio is ....
- Loan Inquiries in last 6 months are less than ....

High Risk Loans

- Disbursement Amount < Loan Requested
- Tenure is longer, i.e. = 60 months
- Interest Rate is higher
- Annual Income below 50K or above 500K
- Home ownership is OTHER, i.e. not known
- Employment Length is too low, or too high or 7 years
- Public Records and Bankruptcies are high
- Purpose is Small Business, Education or Renewable Energy
- State / Zip Code is ....
- Number of Open Accounts is ....
- Debt to Income Ratio is ....
- Loan Inquiries in last 6 months are more than ....