Lending Club Case Study

Presented By:
Sai Krishna Somisetty
Mrityunjaya Shukla

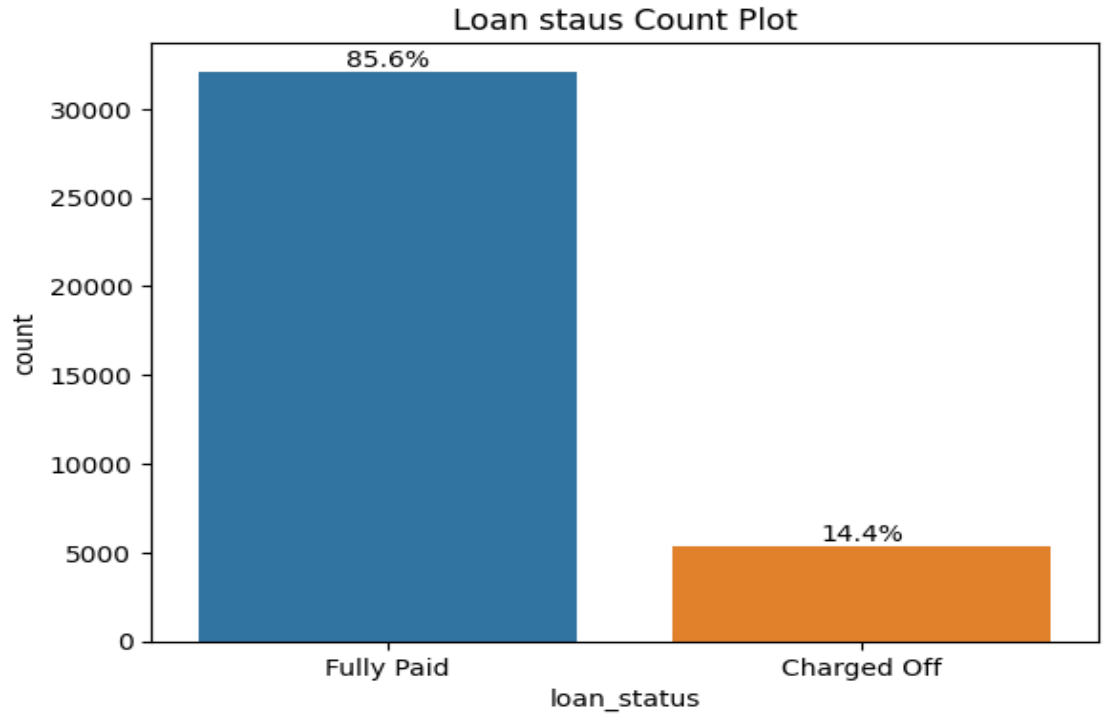
Brief overview of available dataset

- 39717 data records with 111 parameters
- Some parameters are all null
- All loan application types are INDIVIDUAL
- Certain columns like interest rate, employee experience etc are having suffixes in all values
- 2 loan statuses are unique across the dataset - 'Fully Paid', 'Charged Off'

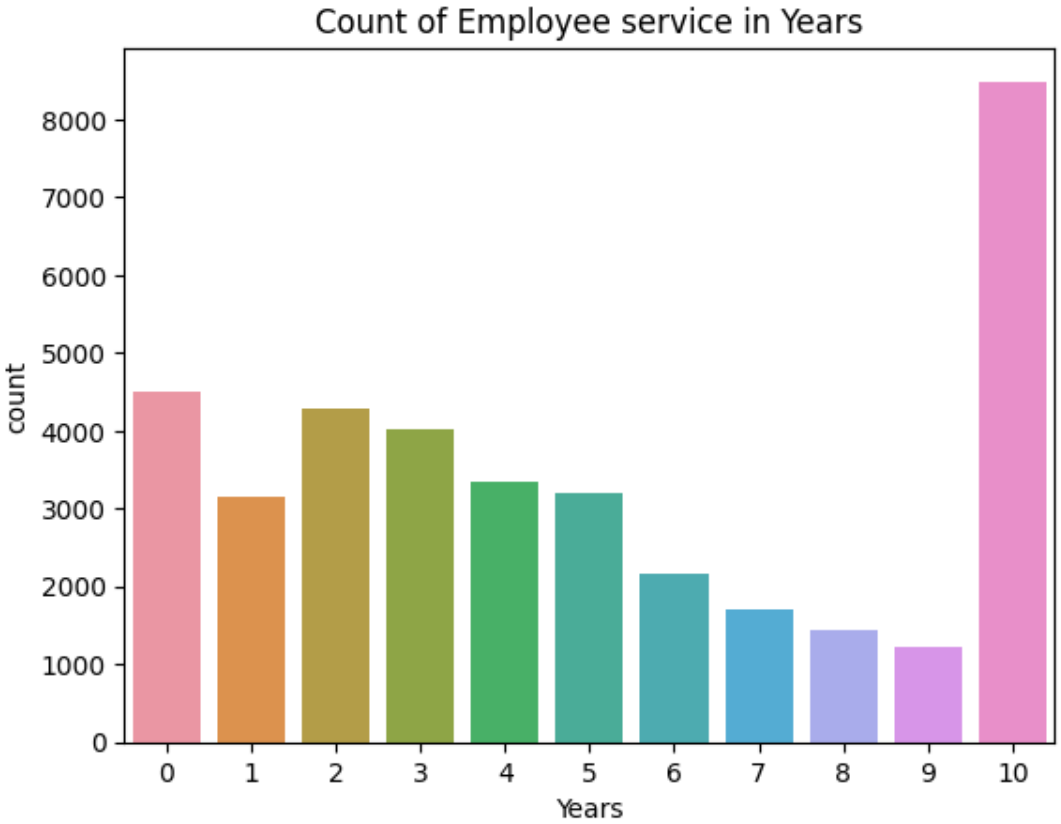
Data Cleaning

- Considered the loan status, which are **Fully paid** and **Charged Off.**
- Deleted all the columns which contains empty values or NA.
- Removed the rows which contains empty cells.
- For the better analysis grouped the **emp_length** 10+ Years to **10** & **<1 Year to 1**
- Created two new columns **‘Year’**, **‘Month’** from ‘**issue_d**’ date column.
- Columns **'emp_length', 'int_rate’** habing data type as Object, **type casted** to numeric.
- Used Sort function to sort the column **‘Grade’** for the better view.

Majority of employees are able to fully pay loan

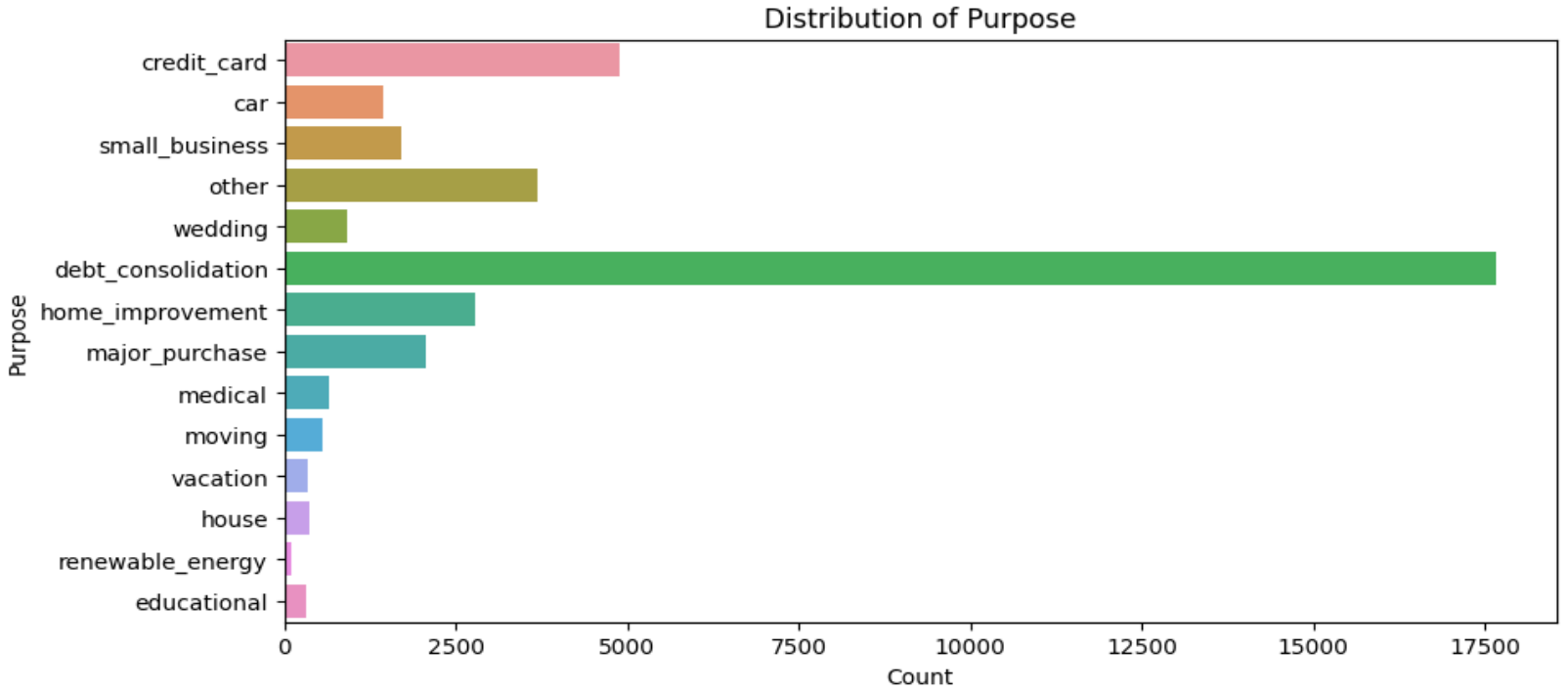


Employees (prospective lendees)



- Top 2 lendees are > 10 years and < 1 year experience
- Loan taking appetite mostly is reducing with experience

Loans category wise distribution

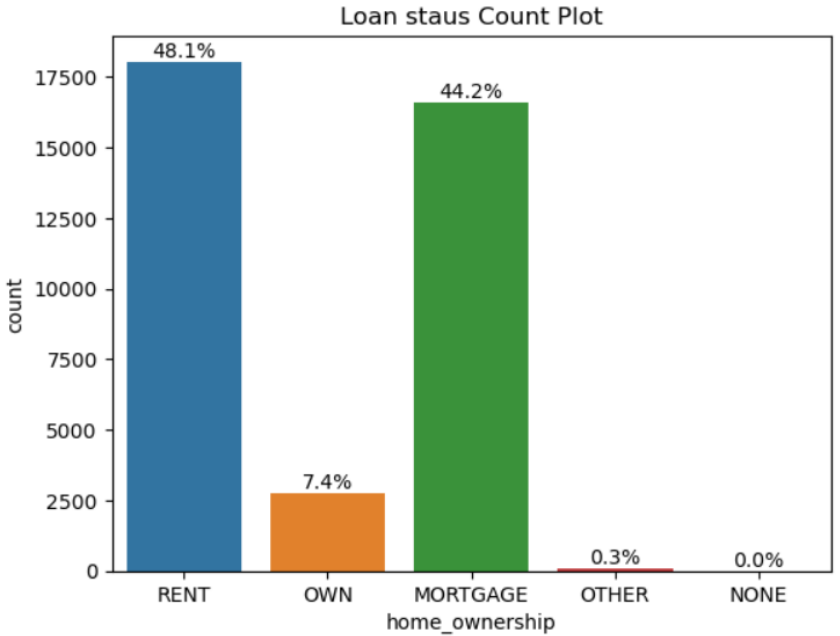


- Debt consolidation is 3.5 times more than second highest share category (credit card)
- Those who take loans take more loans to cover the previous loans

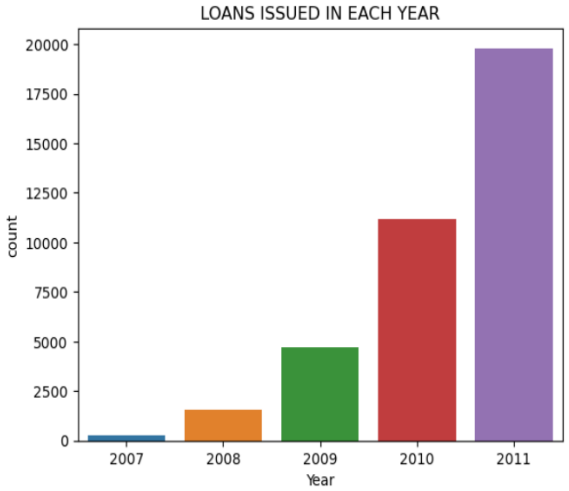
House Ownership

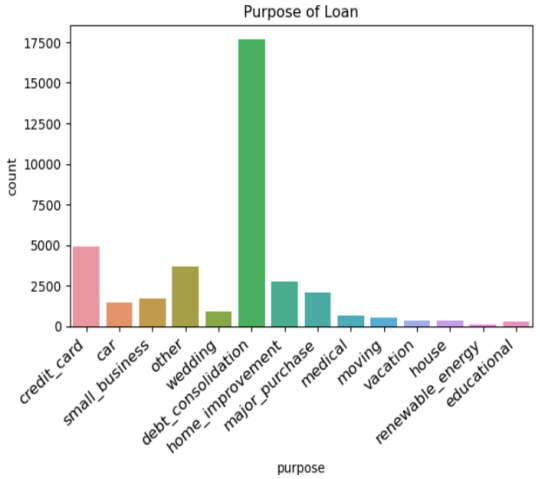
1. Most of the customers are either in **Rented** house or **Mortgaged** their house.

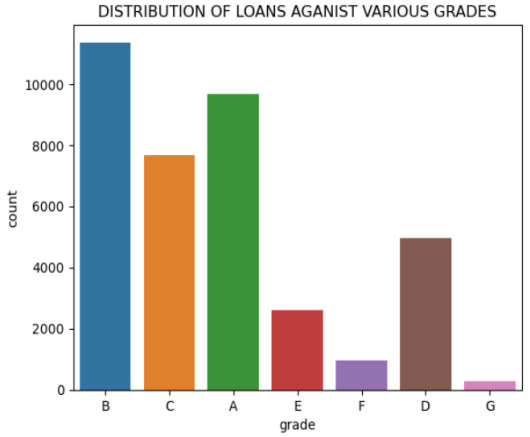
2. People who owns the house are less likely to apply for a loan.



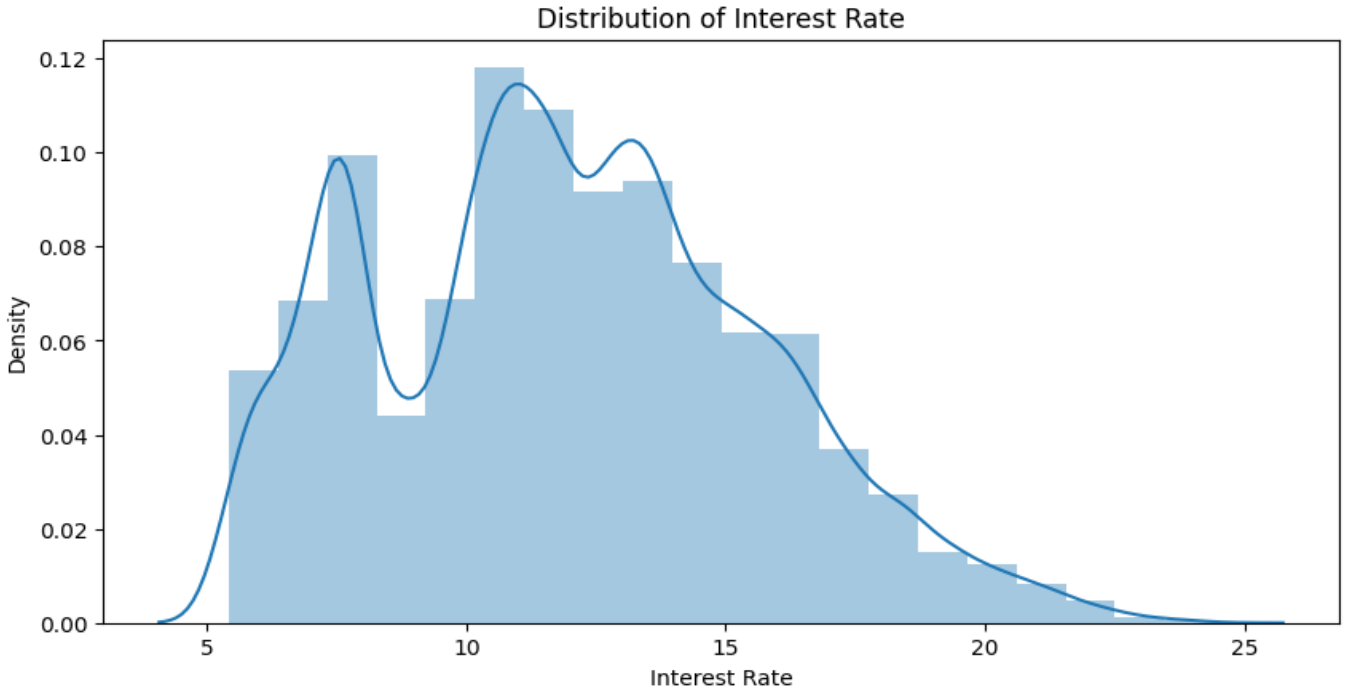
1. Most of the Loans are Issued in the Year **2011.**
2. Majority of people applied for loan to accumulate their **debt_Consolidation**
3. High number of loans are applied by customers of Grade of **B.**





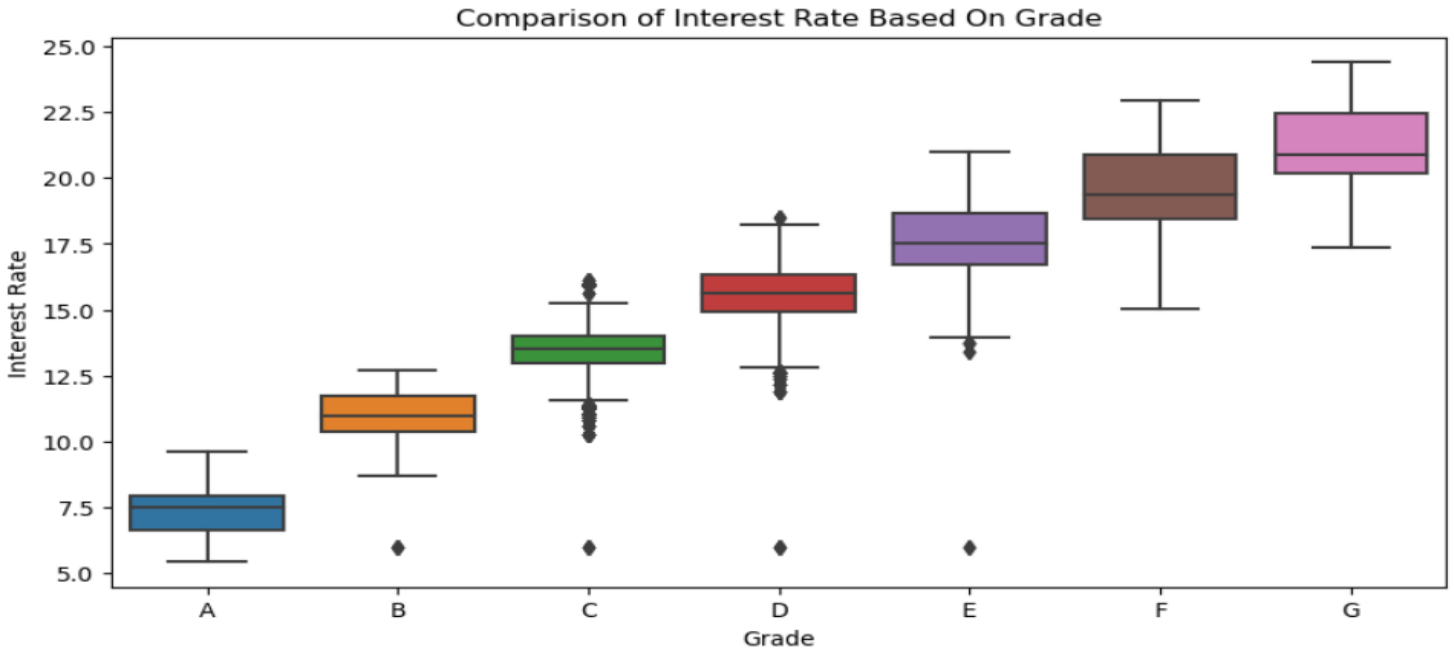


Loans interest rates wise distribution

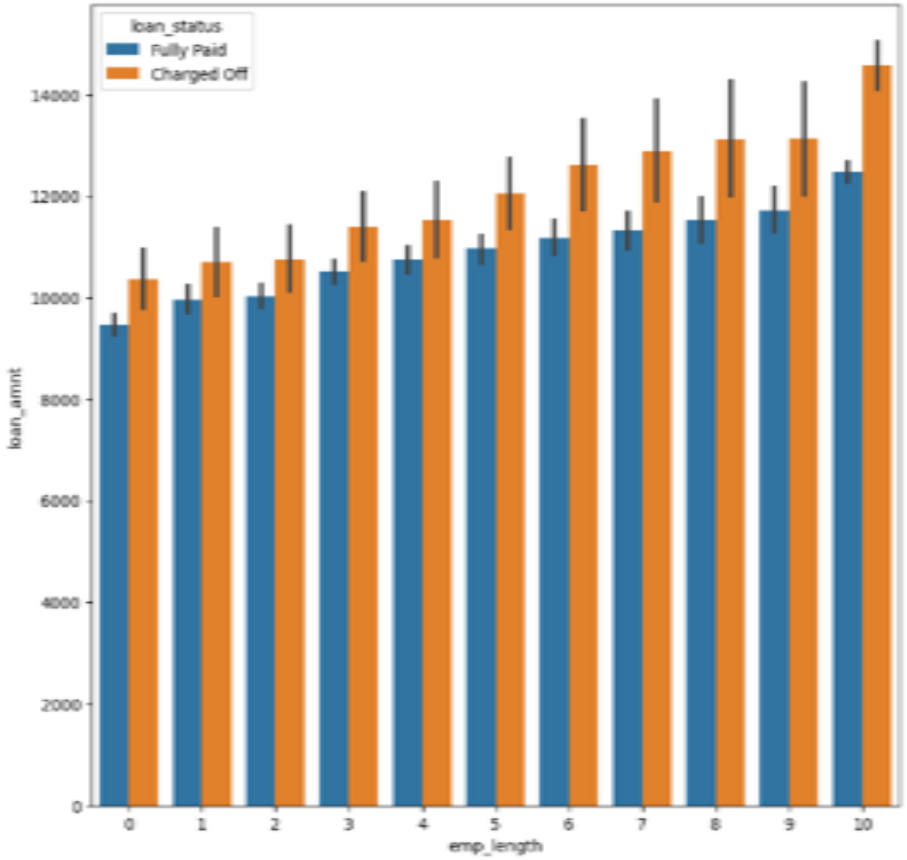


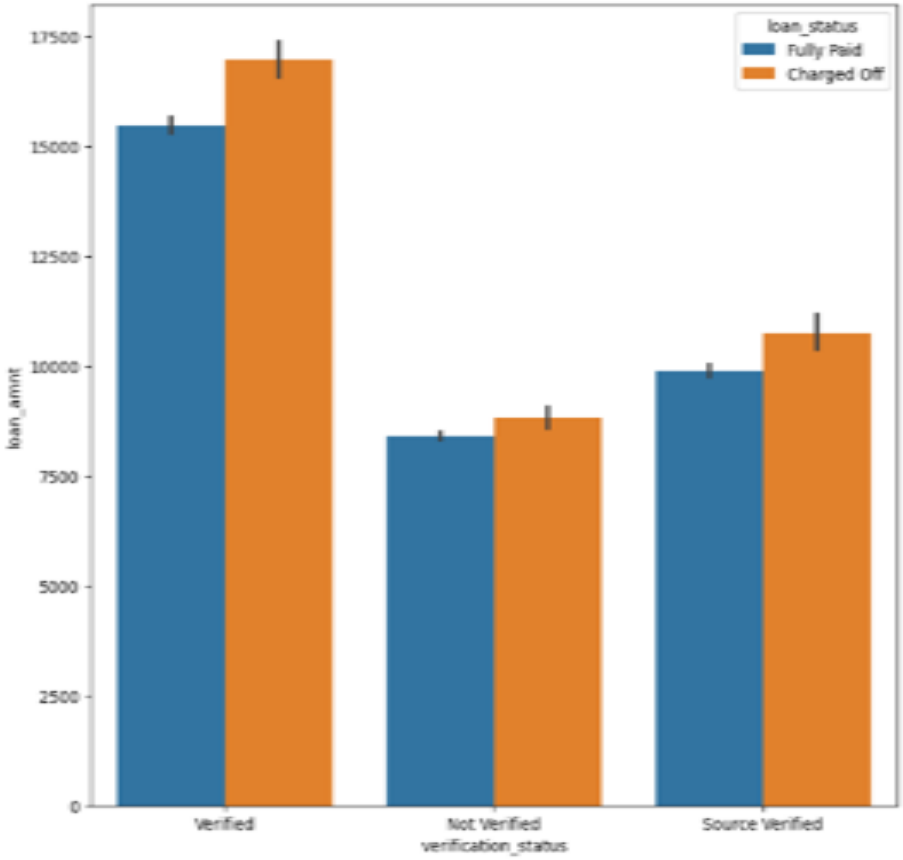
- 11-15% interest rate most spread

Box Plot for Interest rates on each grade



Employee attributes vs defaulting

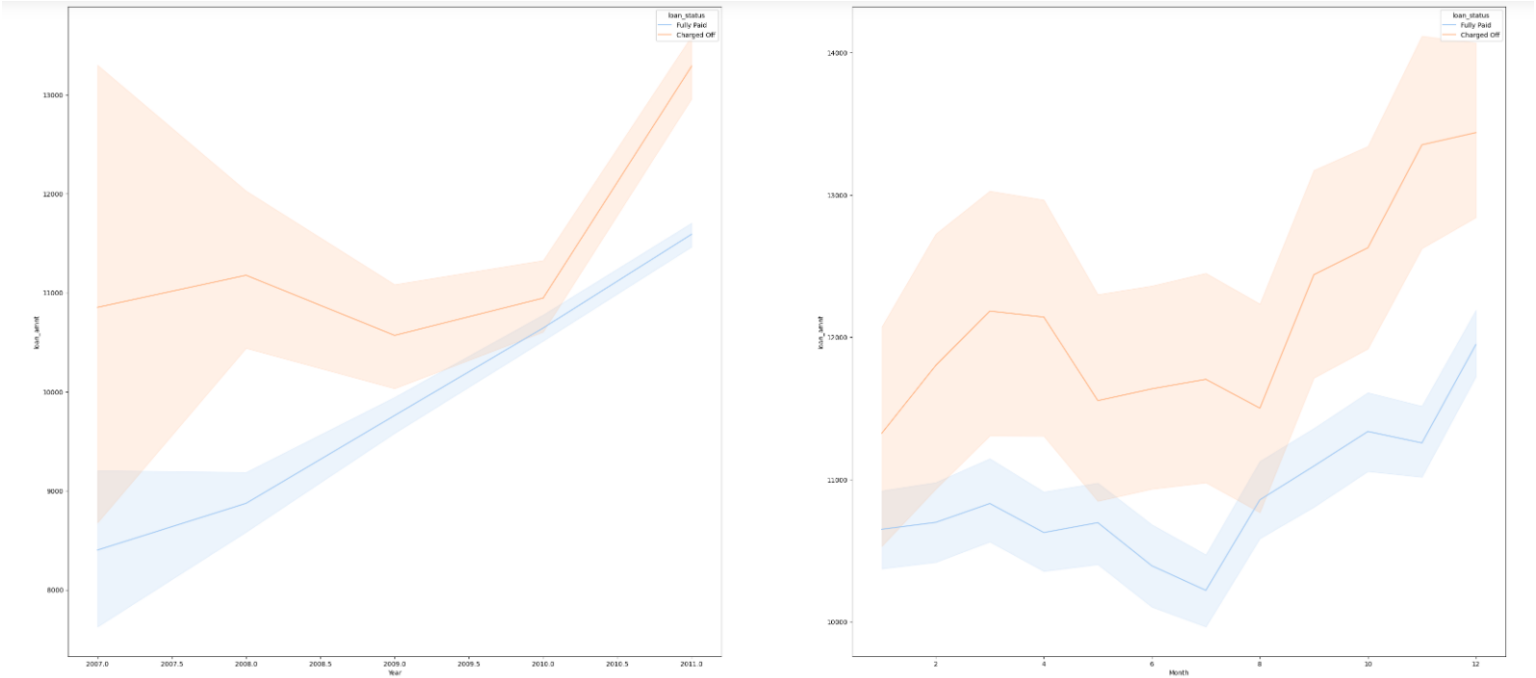




Employee attributes vs defaulting correlation (contd)

- Charged off loan amounts are more than fully paid loans
- The difference percentage increases with employee experience
- Keeping track of the percentage difference over time for every year of experience will give a rough idea of threshold for fully paid loan amount

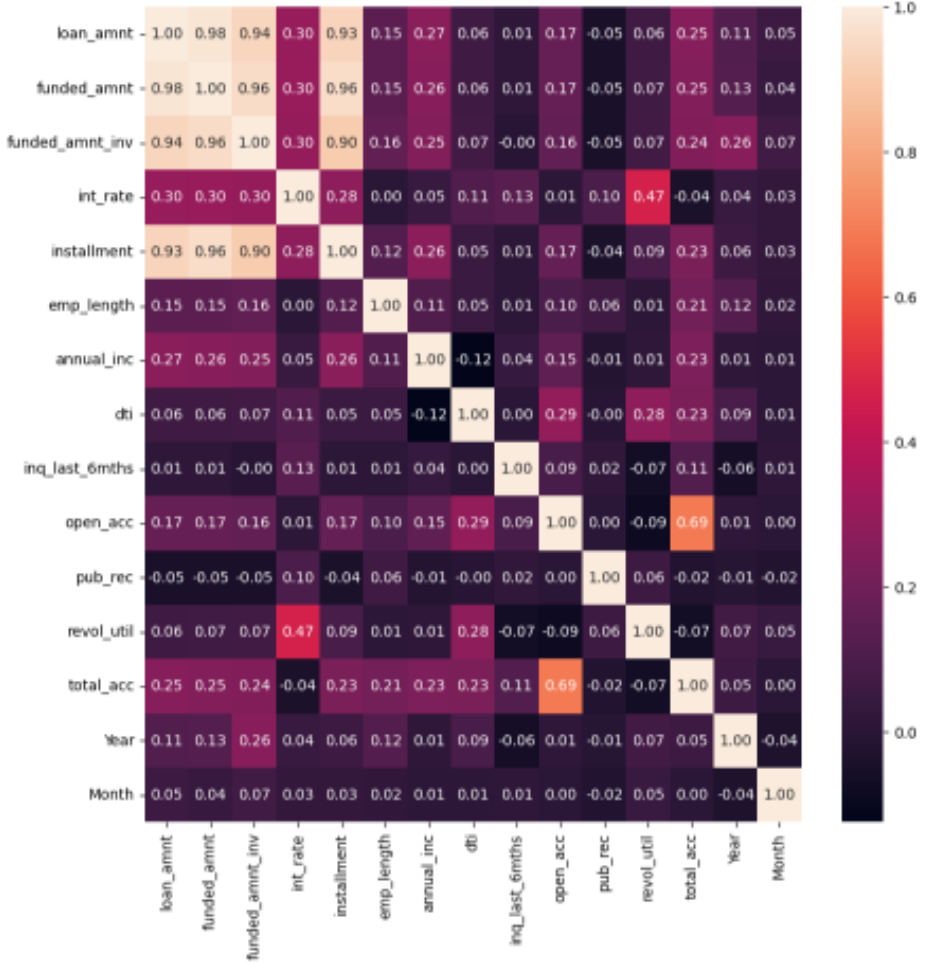
Line graph: Which shows loans issued in each month and a Year



Insights from correla

- Installment amounts increase almost at par with loan amount, so the loan payment duration will mostly be the same across loan cycles

- So the bigger the loan, the larger EMI discipline to be maintained. One miss and default probability increases. Hence smaller loans (to calculate as described in previous slide) should be aimed



Insights from correlation map - 2

- Loan amount and interest rate have a feeble but not ignorable correlation(0.3), the more the loan amount it, interest rates are likely follow.
- With annual increment, the debt to income ratio decreases, resulting in lesser risk of defaulting. Hence more experience results in dissolution of default risk.
- However taking loan at any stage is susceptible to risk afresh at that experience level

Conclusion

- Greater caution should be exercised while giving loans to <1 years and >10 years experienced employees
- Grade G loans are of highest interest rates. High interest loans are susceptible to greater defaults. Hence amounts of grade G loans should be kept small.
- The loan amount should be decided by reviewing the trend of loan disbursement (for fully repaid loan amount) in previous years and the corresponding loan amount increment rate