

LENDING CLUB CASE STUDY

CHANDRIKA SHARMA
DEBARGHA GANGULY

PROBLEM STATEMENT

A real world scenario where a consumer finance company that specialises in lending various types of loans to its urban customers has to analyse various factors and make a decision for loan approval based on the applicant's profile.

An inadequate analysis backed with data can lead to either loss of customer or a loss of money to the company.

SOLUTION

Exploratory Data Analysis (EDA) is a data analysis technique that could help to understand and analyze the data set and provide with meaningful business oriented insights.

This will cover the following:

1. Univariate Analysis
2. Segmented Analysis
3. Bivariate/Multivariate Analysis
4. Derived matrices Analysis

BUSINESS UNDERSTANDING

1. The business objective is to identify patterns which indicate if a person is likely to default.
2. To identify various determining factors that could lead to a charged off customer.
3. To highlight high risk factors in lending business based on data set available.

OUR APPROACH

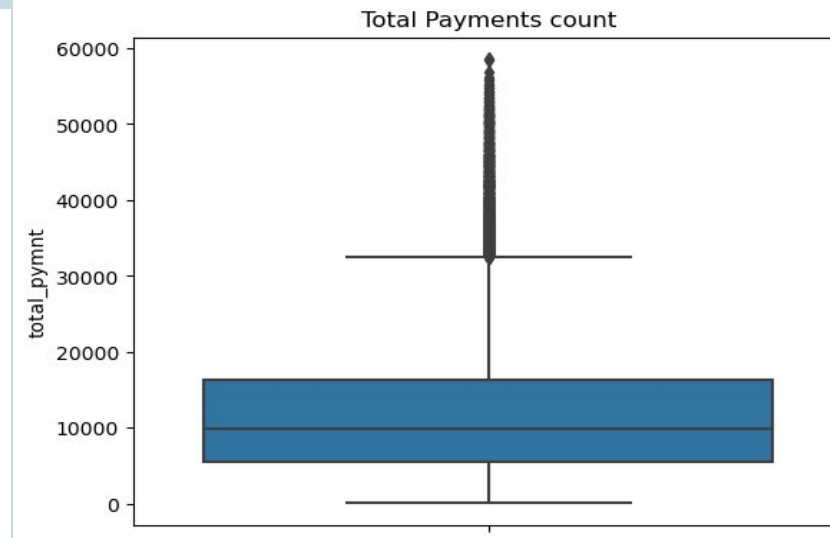
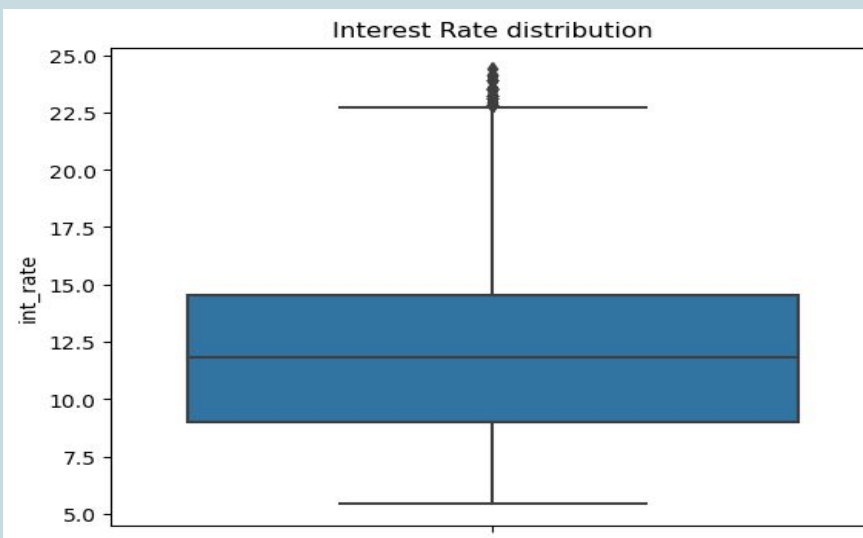
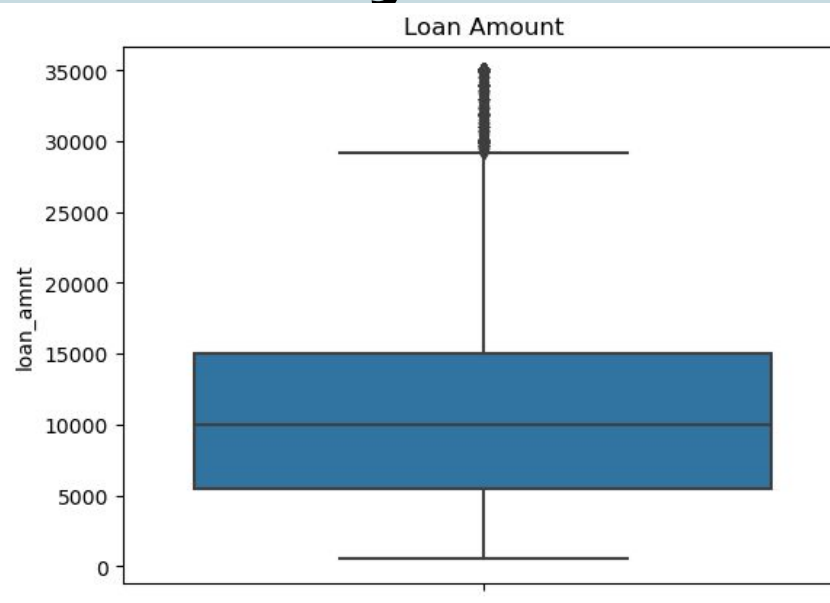
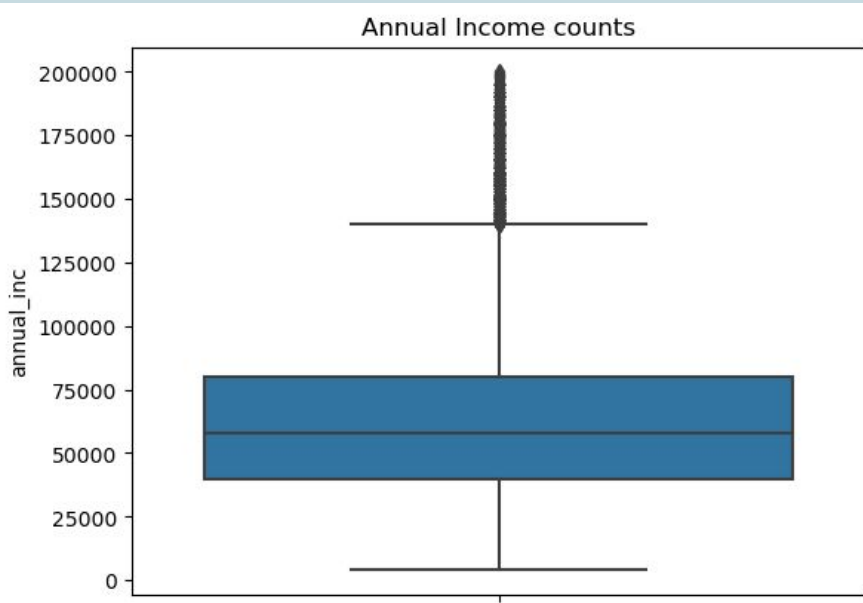
1. Going through the data dictionary provided and understand the meaning of the data given.
2. Gather some domain knowledge through internet.
3. Data cleaning by removing columns having all or maximum(> 50%) null values.
4. Removing irrelevant columns
5. Removing outliers as and when required
6. Standardizing the values and correct the data types.



UNIVARIATE ANALYSIS

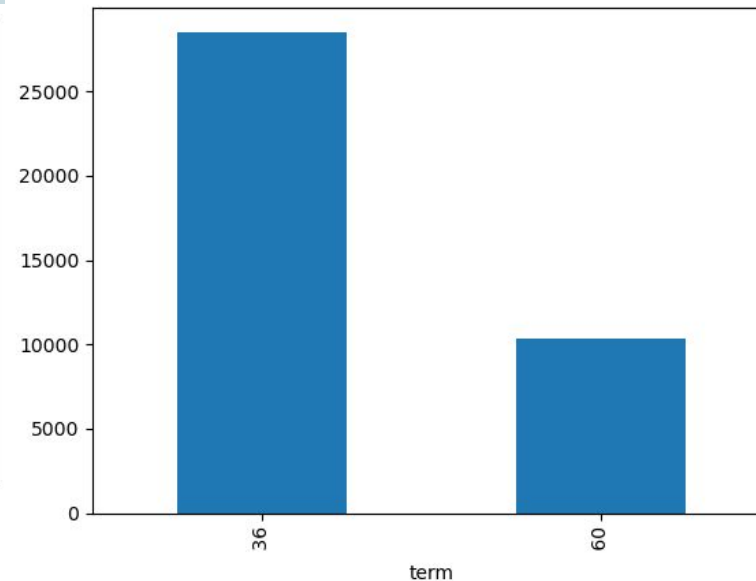
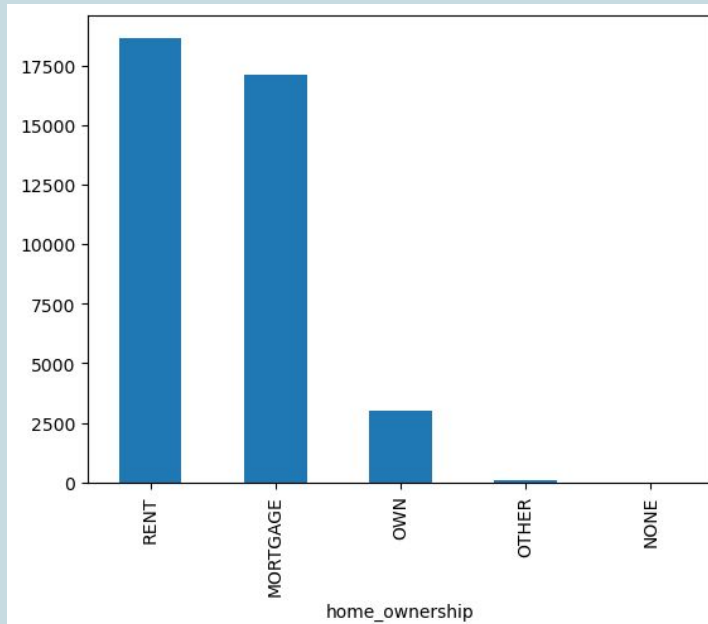
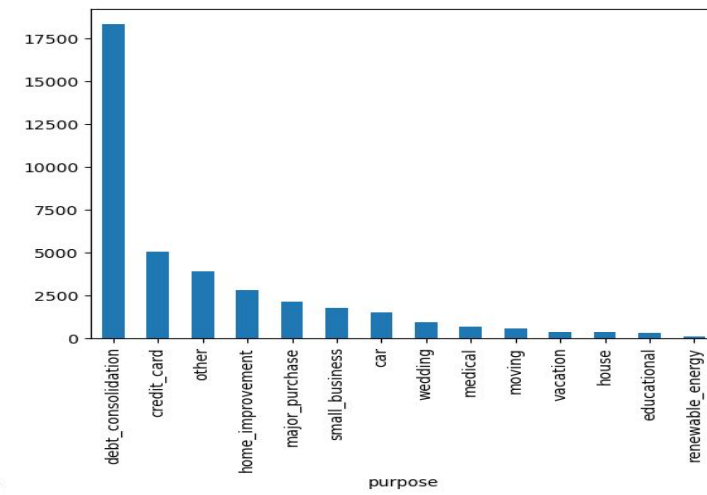
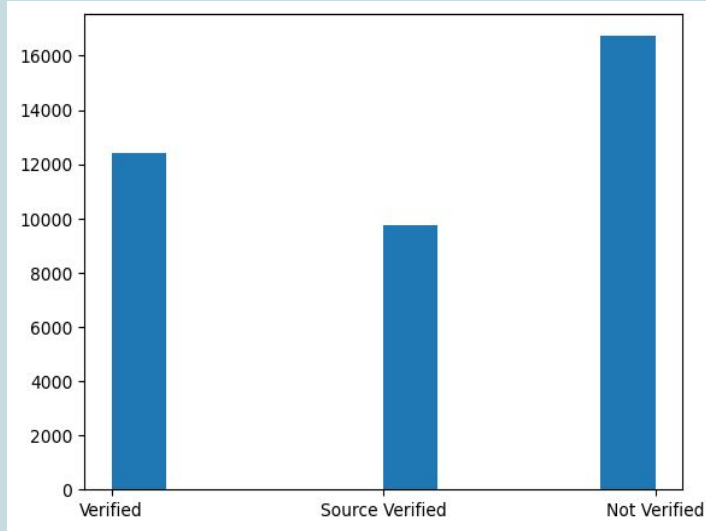
Quantitative variables analysis

- Annual_income range is between 40k to 200k anything above 200k is removed from analysis. Medium income is around 60k
- Median Loan amount is around 10k while while highest can go upto 35K. Loans greater than 30k are considered outliers and removed from analysis
- Interest rates mostly between 9-15 percent with median at 11.5%. Interest rate greater than 22.5% are considered outliers and removed from analysis
- Total payment median is around 10k and extends till 60k. Payments above 32k are outliers.



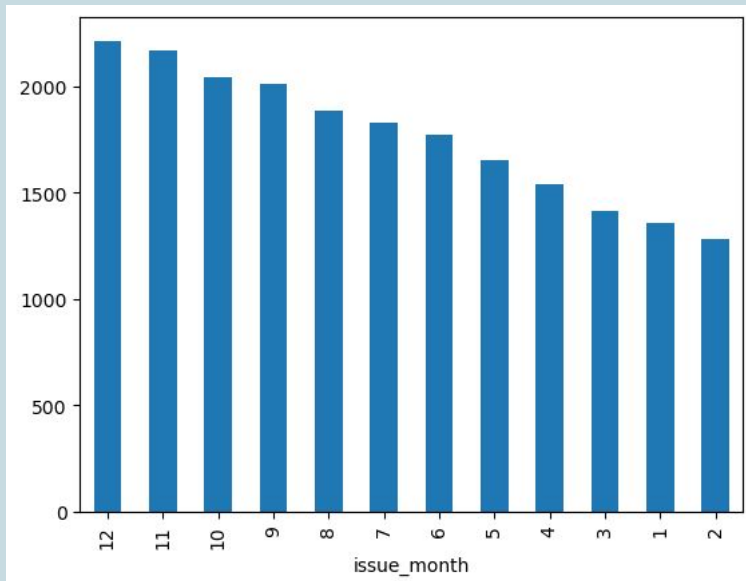
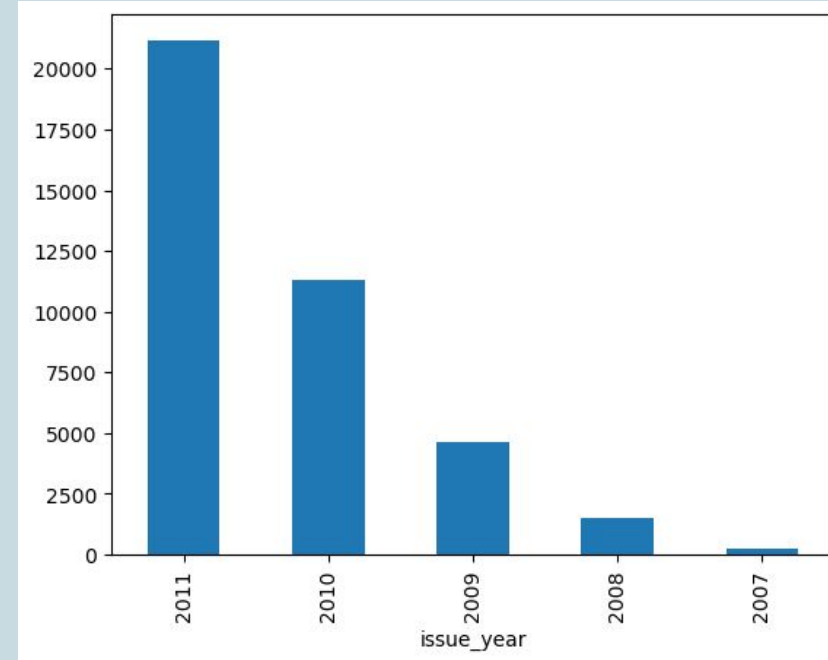
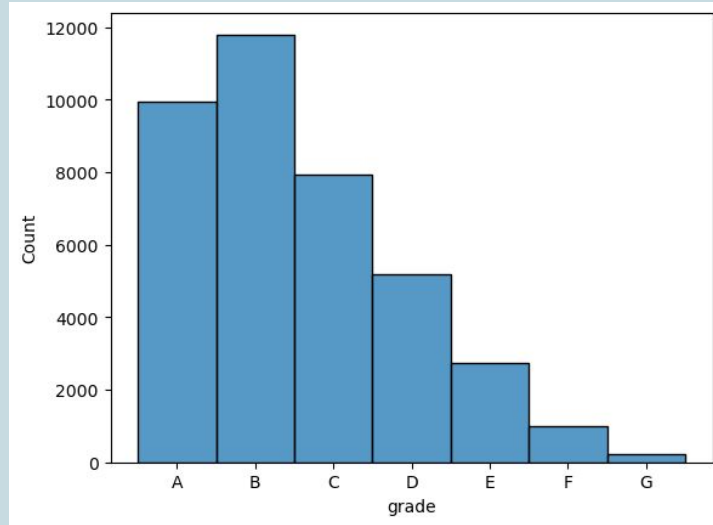
Total payment inv and funded amount follows similar distribution

Unordered categories



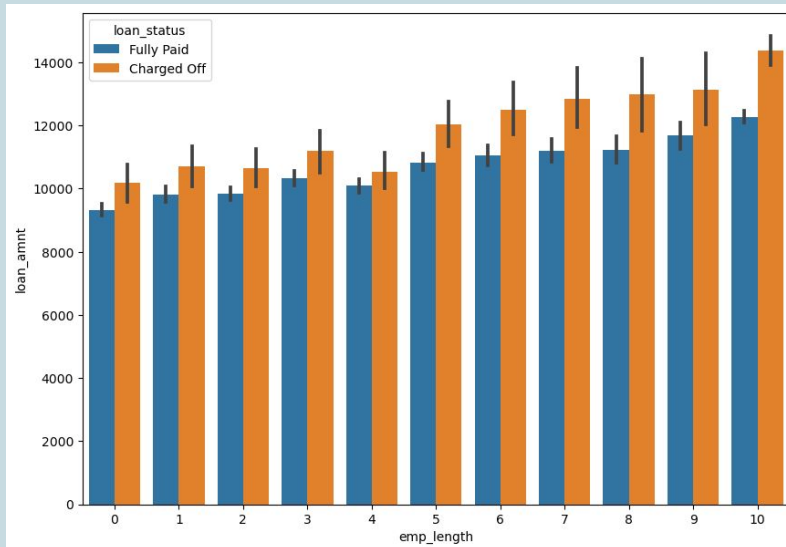
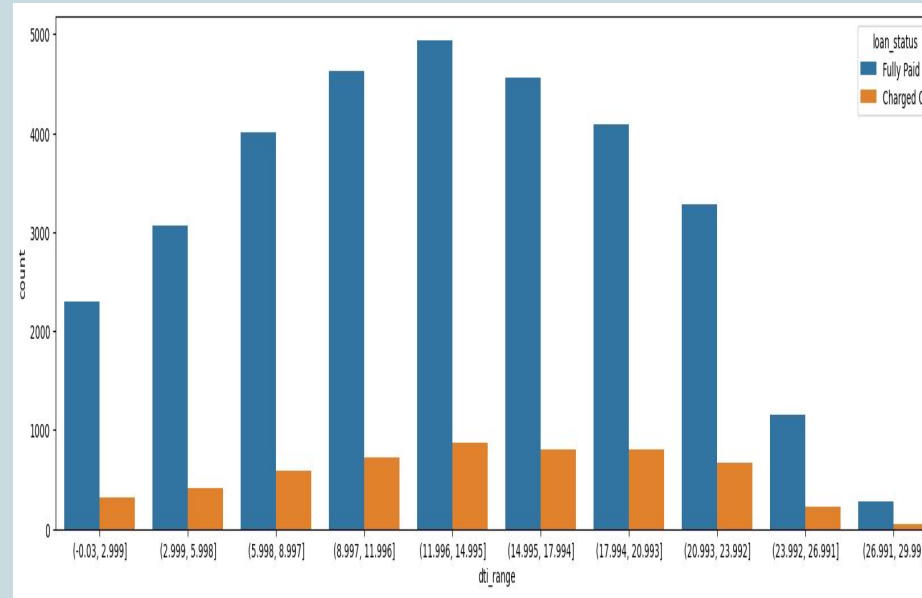
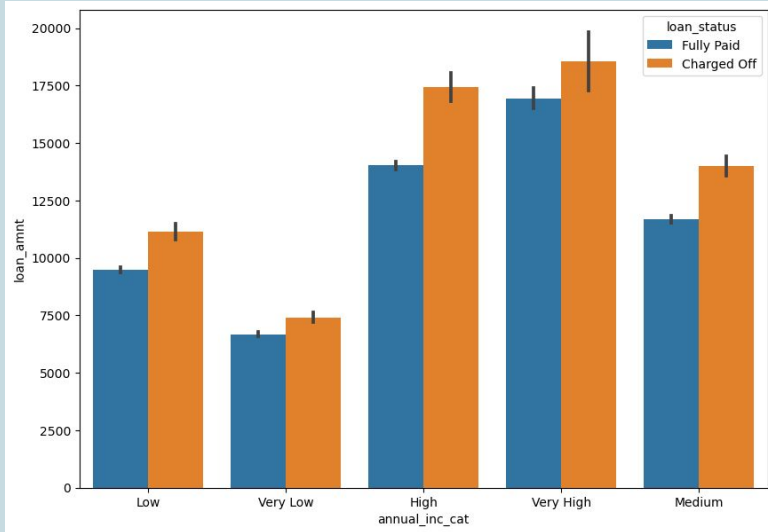
- Around 40% unverified loans. This should be a worrying sign for the company
- Most loans are taken for debt_consolidation
- Most people taking loans are living on rent
- Most loans are for shorter duration

Ordered categories



- Number of loans increasing from A to B and then decreasing gradually. Very few loans for grade F, G which is a good sign for the company
- Number of loans issued has increased exponentially from 2007 to 2011 and December sees highest number of loans issued probably due to festive season and holidays

Segmented analysis

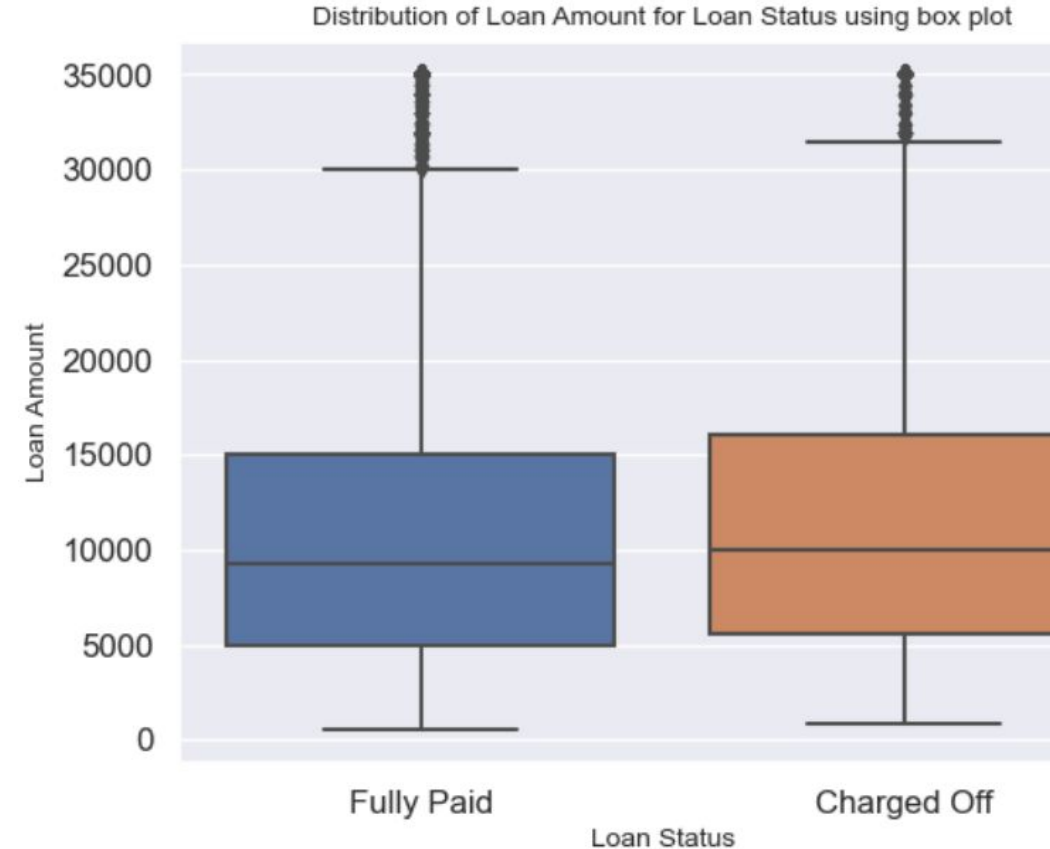
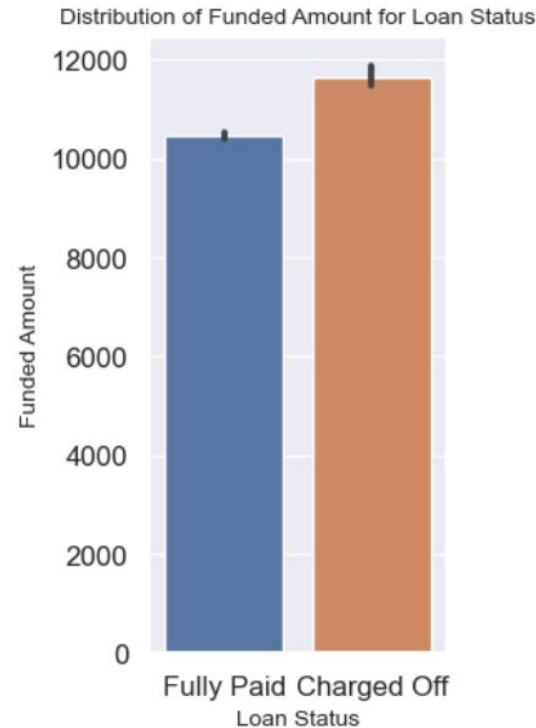
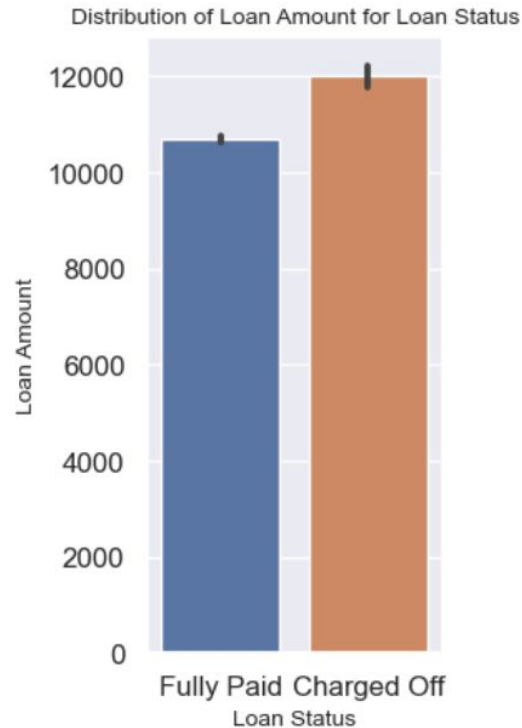


- Annual income, dti and emp_length values has been segmented and analysed for loan_amount
- Very high income group (>95 percentile has most charged off loans however ratio of total to charged off loans remains same for all segments
- dti range 12-15 has maximum number of loans but charged off loans doesn't change much based on dti
- Employees in 10+ years category has maximum charged off loans



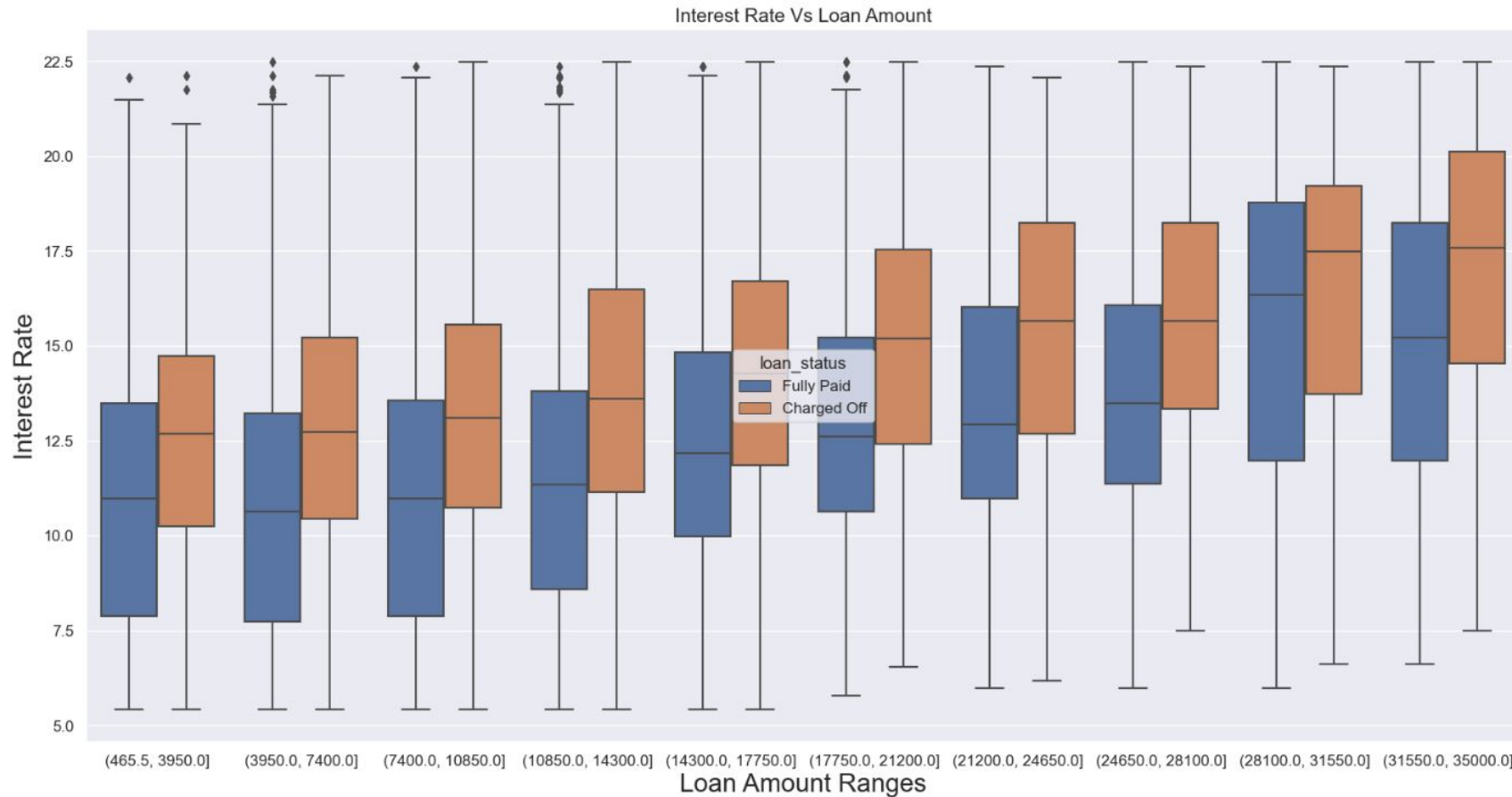
BI/MULTI VARIATE ANALYSIS

LOAN AMOUNT VS LOAN STATUS



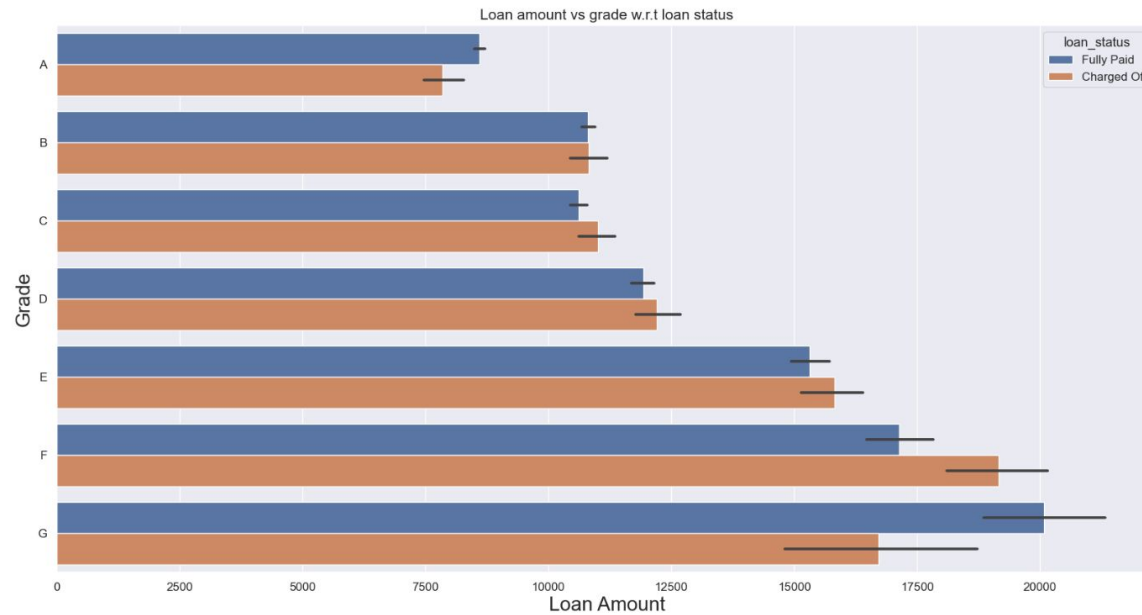
1. There is no major difference between loan amount and funded amount
2. An average of loan amount given across charged off status is slightly more than that of fully paid.
3. The loan amount ranges from USD 500 to 35K. Most of the clients were given loan amounts from 5K to 16K

INTEREST RATE VS LOAN AMOUNT



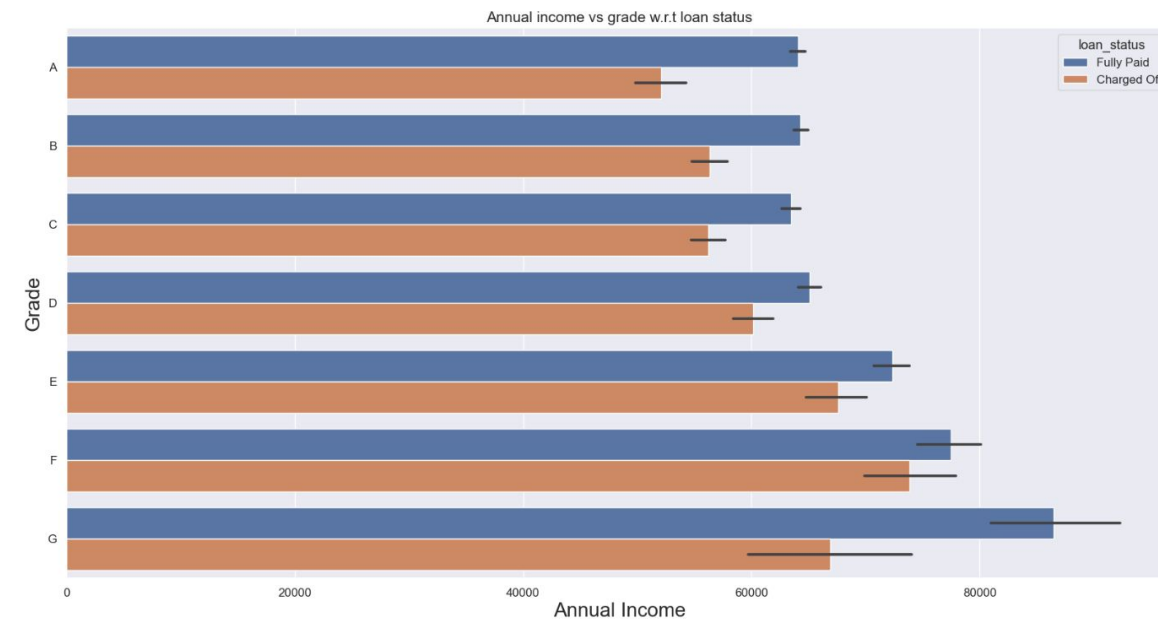
1. Charged off loan have a significant high interest rate compared to fully paid category across all loan amount ranges.
2. Also, people tend to default more when interest rate is 14-17.5%
3. People applied for loan amount between 28K - 35K at a rate of interest 17.5% or above tends to default more.

GRADE VS LOAN AMOUNT & INCOME W.R.T LOAN STATUS

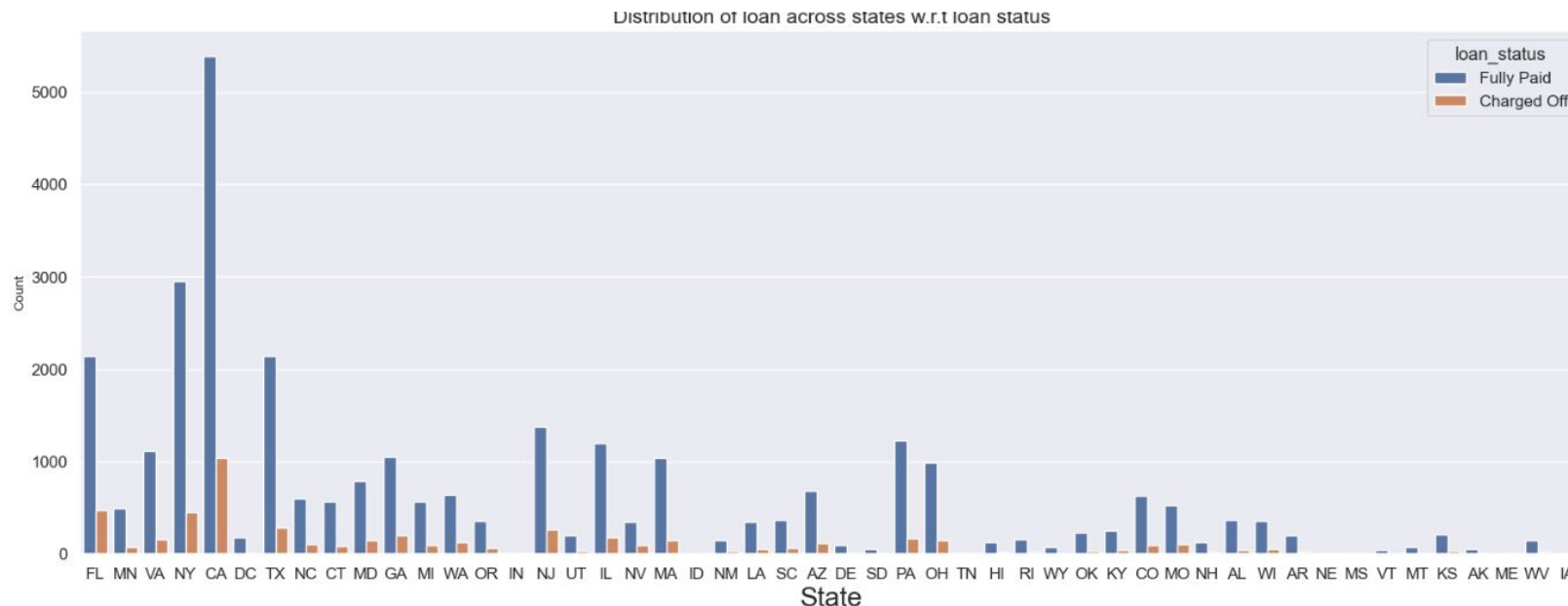


1. Loan amount and interest rate is high for grades from E to G. Thus, loans in these grades are at higher risk of default.
2. Chances of default is least for grade A, followed by B, C, D

1. As grade G is high risk loan, it is given to high annual income group and for high term.
2. For grade G, the people with annual income less than equal to 70K and interest rate more than 20% tend to default. Hence at high risk.



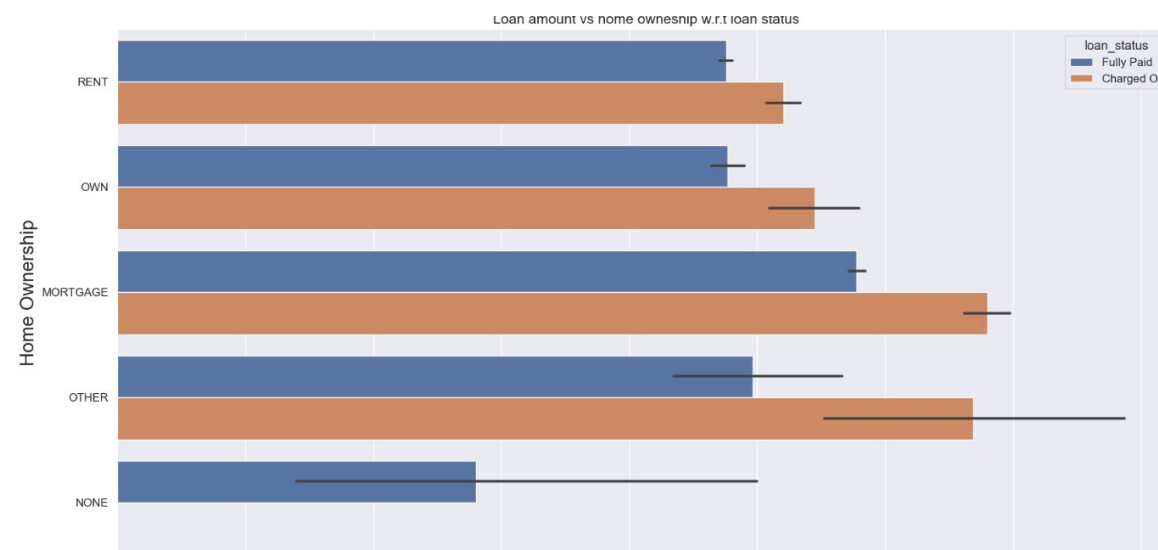
Analysis with home ownership & state



A high count of loan is recorded for the metro cities such as California, New york.

Loan amount of more than 12K with mortgage and int rate more than 11.5% could be a high risk factor.

Home ownership should be identified before approving loan as there are many defaults seen in other category. Whereas no defaults in none criteria might be for people living in their parent's home.

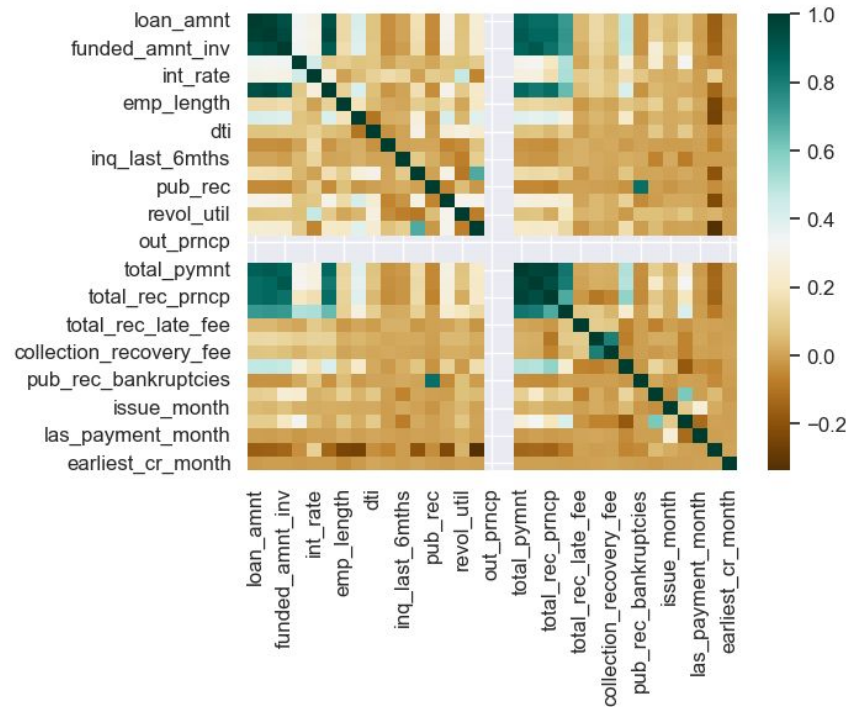


Some more analysis and its findings



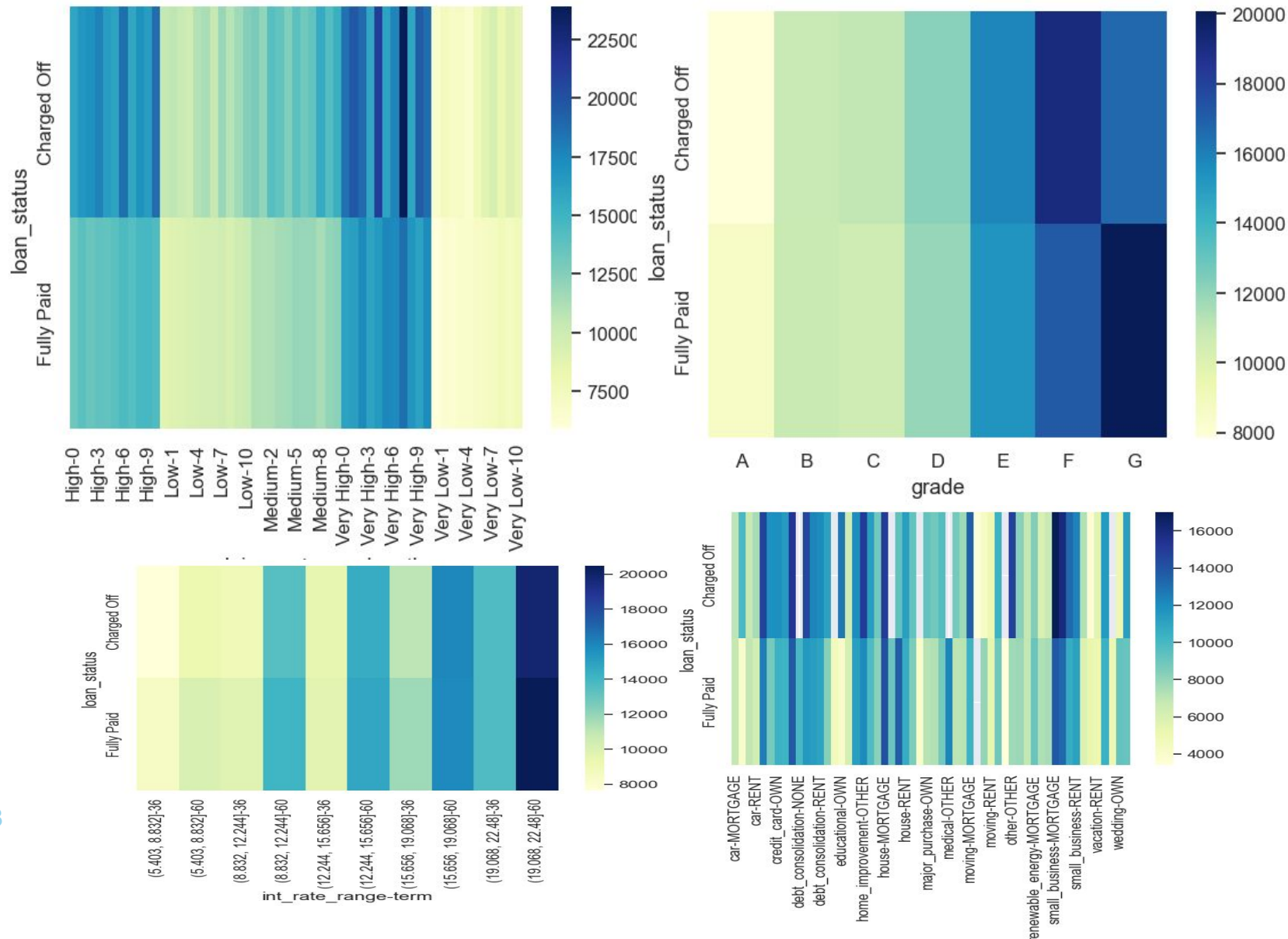
1. For open credit lines from 6-8, the chances of getting loan is higher.
2. Greater the revolving balance lesser the chances of getting loans.
3. People having more than 10 years of credit history has a higher chances of getting loans sanctioned.
4. More loans are given for 0 or 1 number of delinq in past 2 yrs.
5. Delinq 2yrs more than 1 the amount of loans given decreases drastically which is ideal for the LC.
6. Loans disbursed for a max DTI range is 29. DTI higher than 30 loans are not disbursed.
7. Mostly loans are approved for people having 0 public record bankruptcies.
8. Verified applicants get approval for higher amount of loan.
9. Greater the revolving balance lesser the chances of getting loans.

Correlation Analysis



- Loan amount and funded amount are highly correlated for total amount and total payments and total received principal amount which is expected.
- Loan amount, funded amount revolving balance, total amount paid has a decent (~40%) positive correlation with annual income
- debt ratio(dti) is negatively correlated to annual income

Multivariate heatmaps



- Average charged off loans are highest in the very high income group(>95 percentile) and employment period 6-9 years
- Grade F has the highest average charged off loans
- Interest rate (19-22) range and term 5 years has highest charged off loans
- Highest average charged off loan is for purpose as 'small_business' and its even higher when the person has Mortgage home
- Interestingly Loans taken for education having own house has very less default

Conclusions

The key factors identified to make decision for loan approval could be:

1. Annual Income
2. Grades
3. Verification Status
4. Public record bankruptcies
5. Delinq 2 yrs
6. DTI
7. Credit line history

Recommendations

A few considerations to avoid or minimize charged off accounts:

1. Loan amount higher than 30K at an interest rate equal or above 15% is a high risk account and hence must be thoroughly investigated before taking a decision.
2. For a loan amount of more than 30K, an annual income of the borrower must be equal or greater than 70K.
3. To avoid defaults, the debt to income ratio of the borrower must be less than 30
4. Home ownership must be checked and verified. Mortgage and rent is one of the risk factor.
5. Loans falling under grades E,F,G are seen to have a high risk associated. For grade G, people with annual income less than equal to 70K and interest rate more than 20% tend to default. Hence at high risk.
6. An employment length is 10+ yrs and loan amount of 12k-14k is a high risk account.
7. A loan amount of more than 15K taken for a purpose of small business is at high risk.
8. Borrower having more than 1 public recorded bankruptcies and delinq count is at high risk.
9. Borrower having a high amount of outstanding balance in his revolving accounts is at a high risk.