

Lending Club Case Study

Anshuman Acharya

Anant BM

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 (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as '**charged-off**' are the '**defaulters**'.

Objective

- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.
- In other words, 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. And thus minimize the risk of losing money while lending to customers.

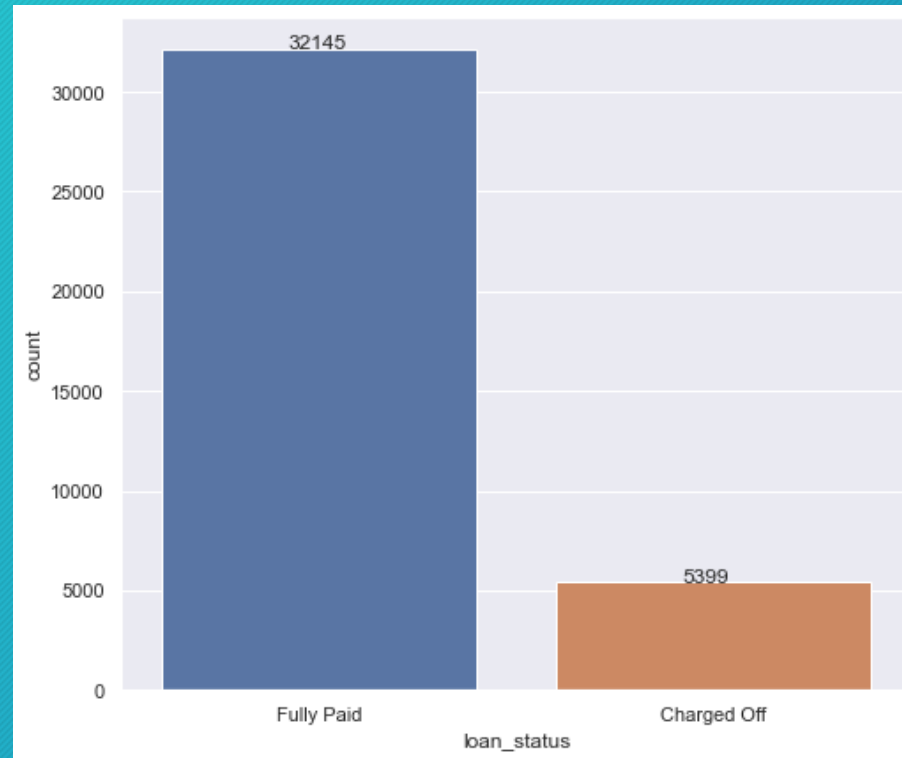
Important Columns used for Analysis

- loan_amnt
- funded_amnt
- funded_amnt_inv
- Term_in_months
- int_rate
- installment
- emp_length_years
- annual_inc
- dti
- total_acc
- total_pymnt
- total_pymnt_inv
- total_rec_prncp
- total_rec_int
- total_rec_late_fee
- recoveries
- pub_rec_bankruptcies
- loan_income_ratio
- issue_year

Steps used

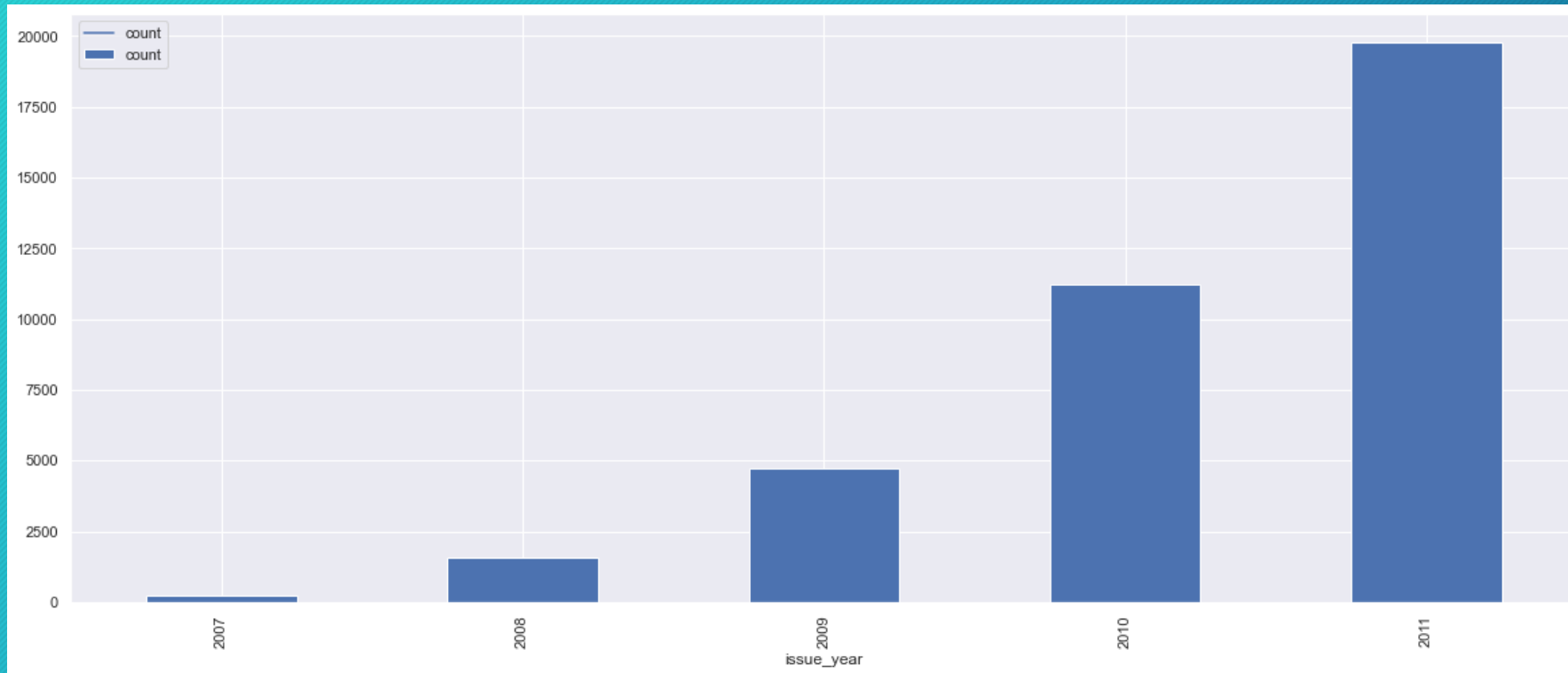
- Data Import
- Data Cleaning
- Data Analysis
- Insights

Default Rate



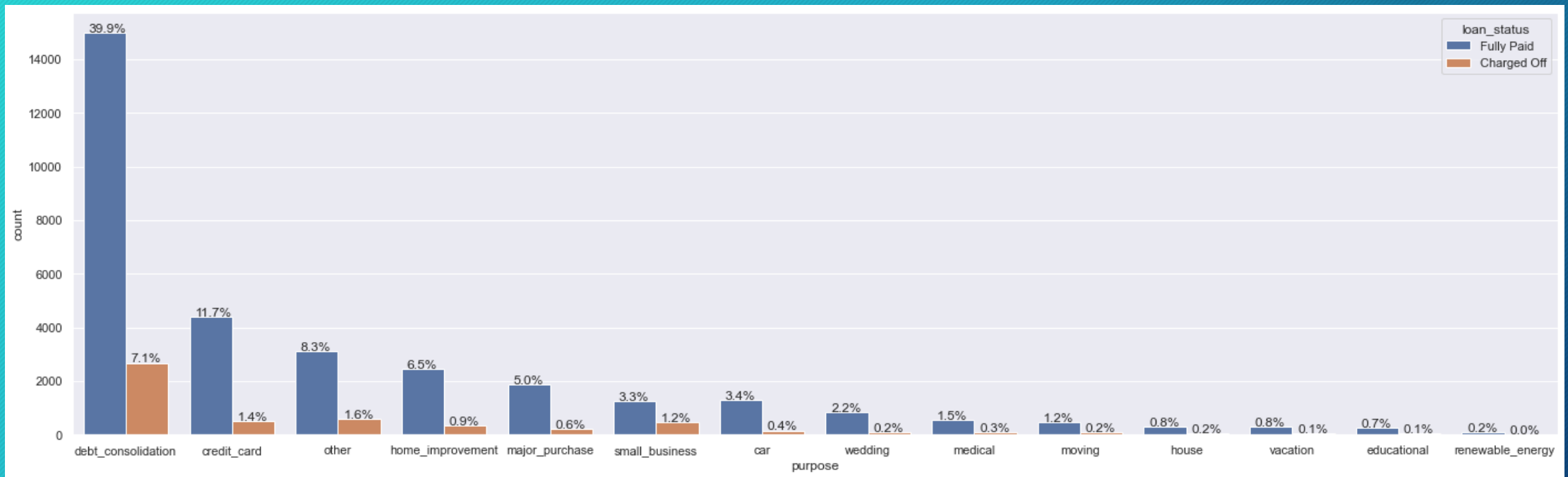
Here we can see that around 15% of loans are charged off

Year wise loan strength



Loan applications are increasing year on year

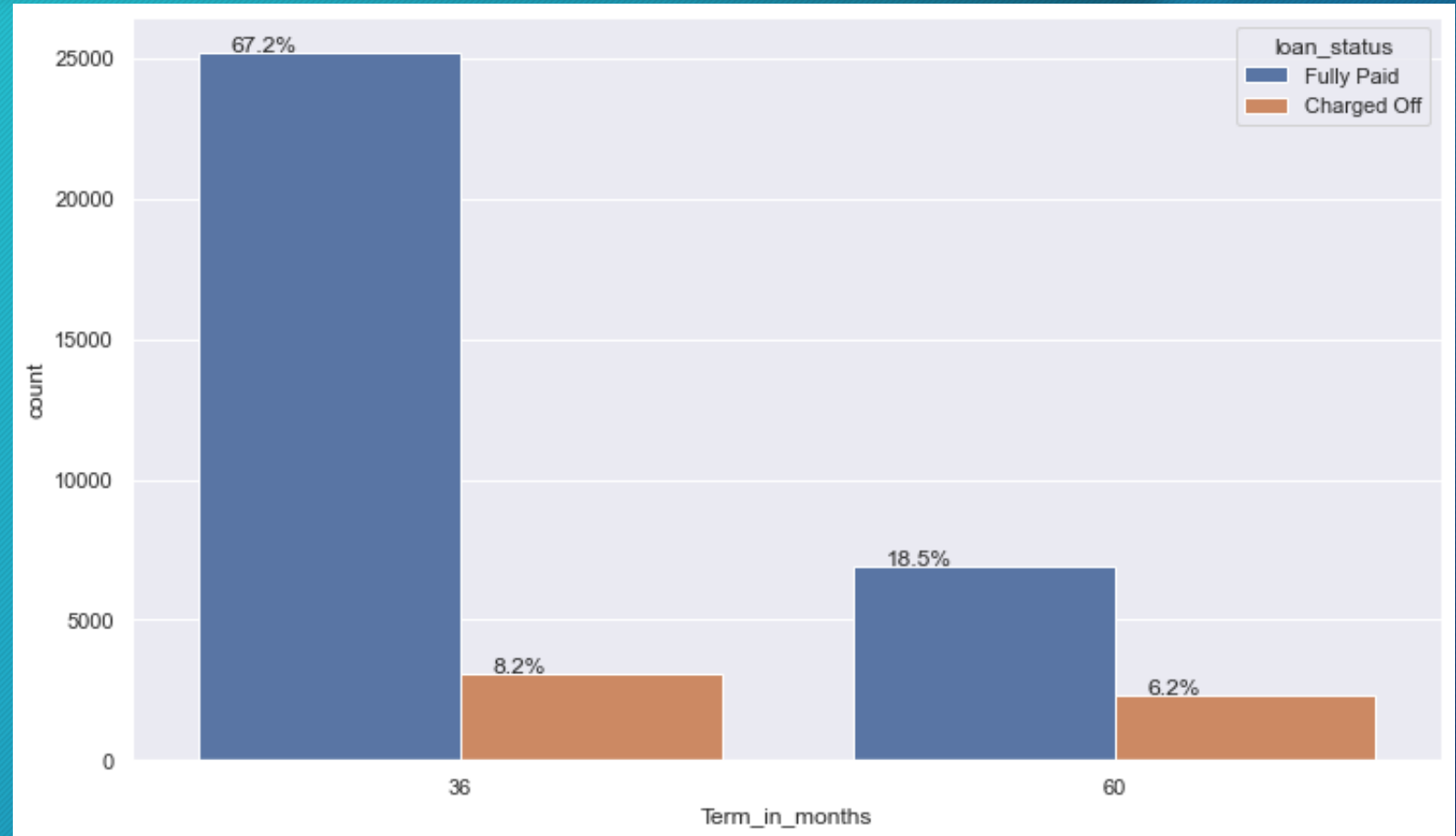
Loan Purpose



Around 60% of the applicants applied loan for Debt Consolidation

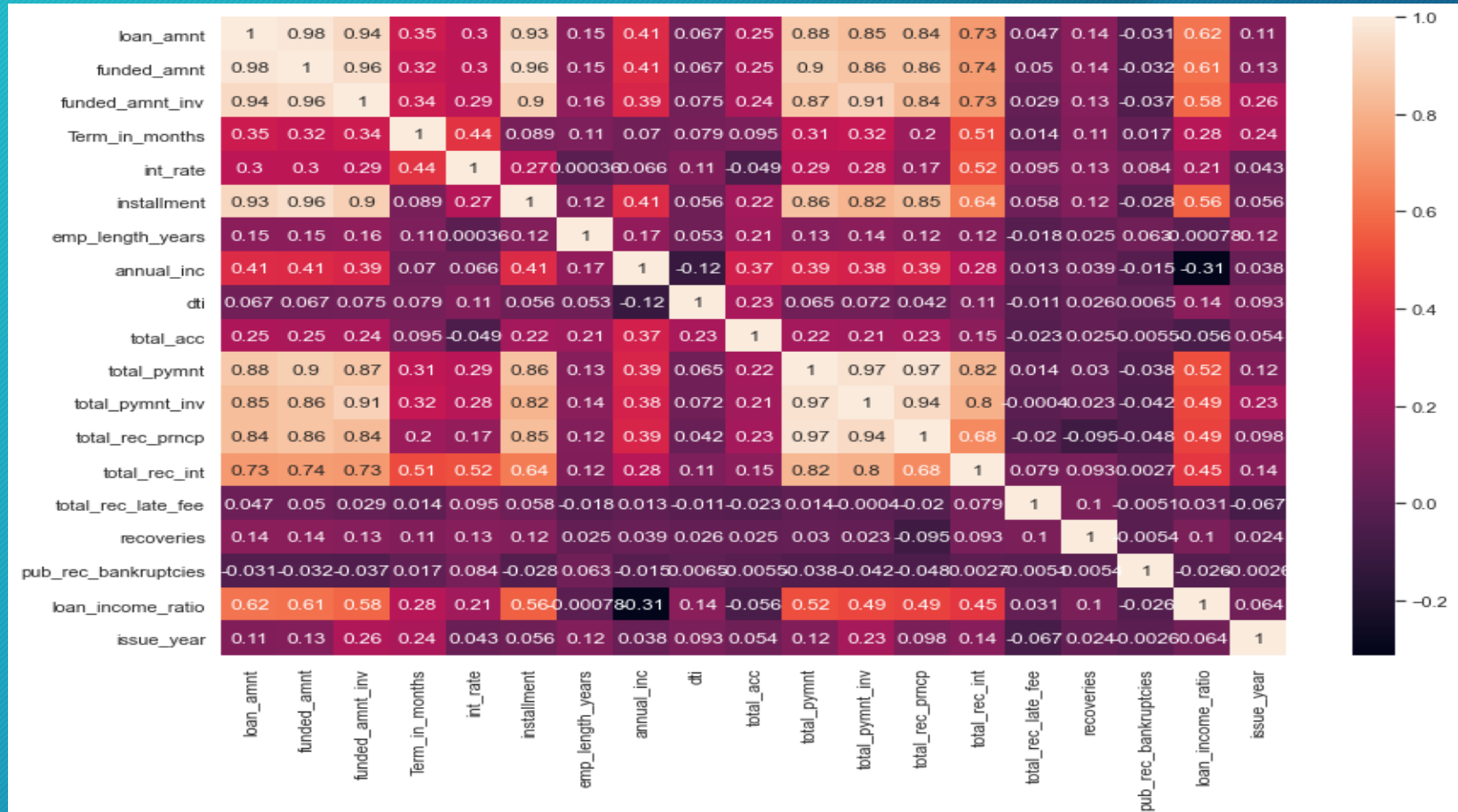
Loan Term

Around 75% of applicants applied loan for 36 months term period. if compared with in group, the long term have high default

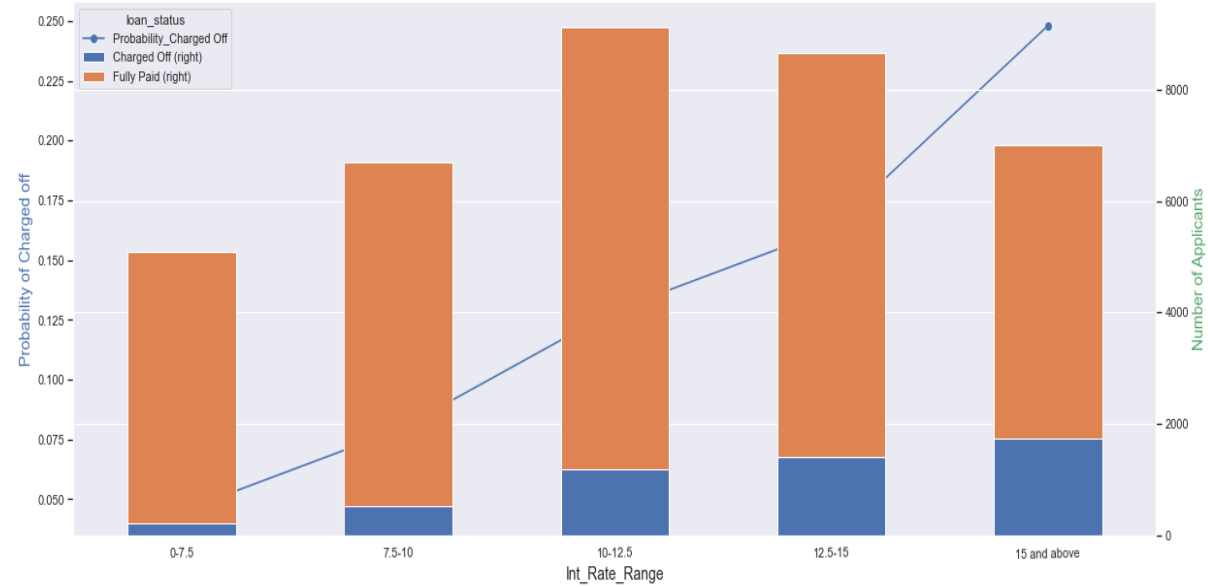


Loan Correlation Matrix

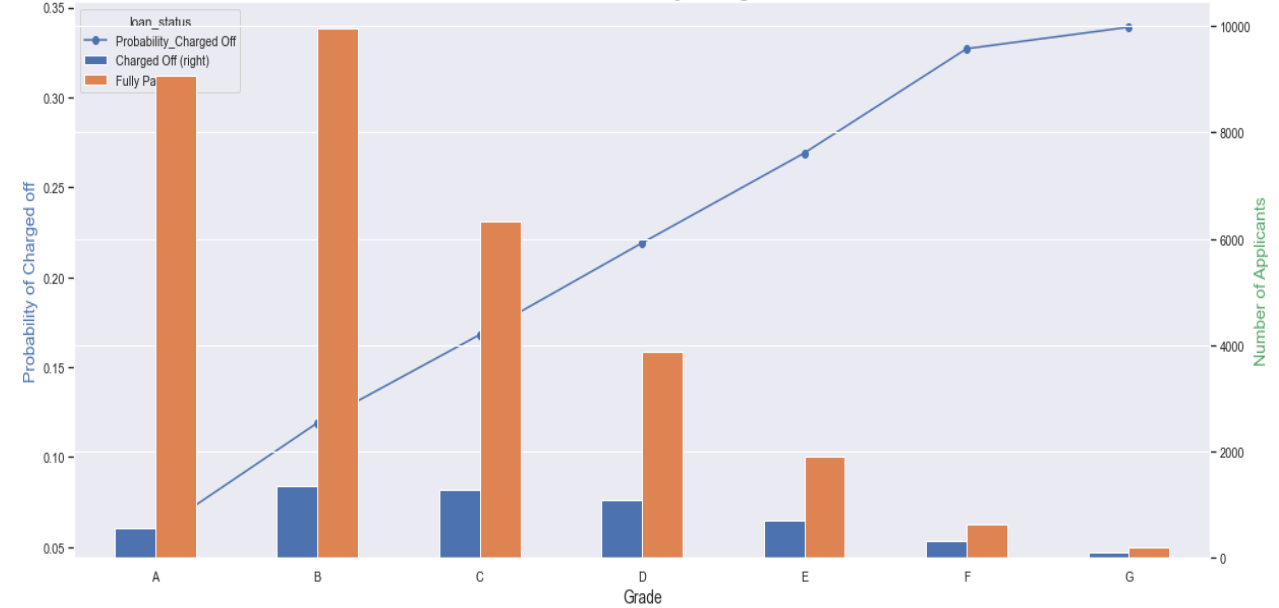
It is clear from the Heatmap that how 'loan_amnt', 'funded_amnt' & 'funded_amnt_inv' are closely interrelated and hence are used for analysis.



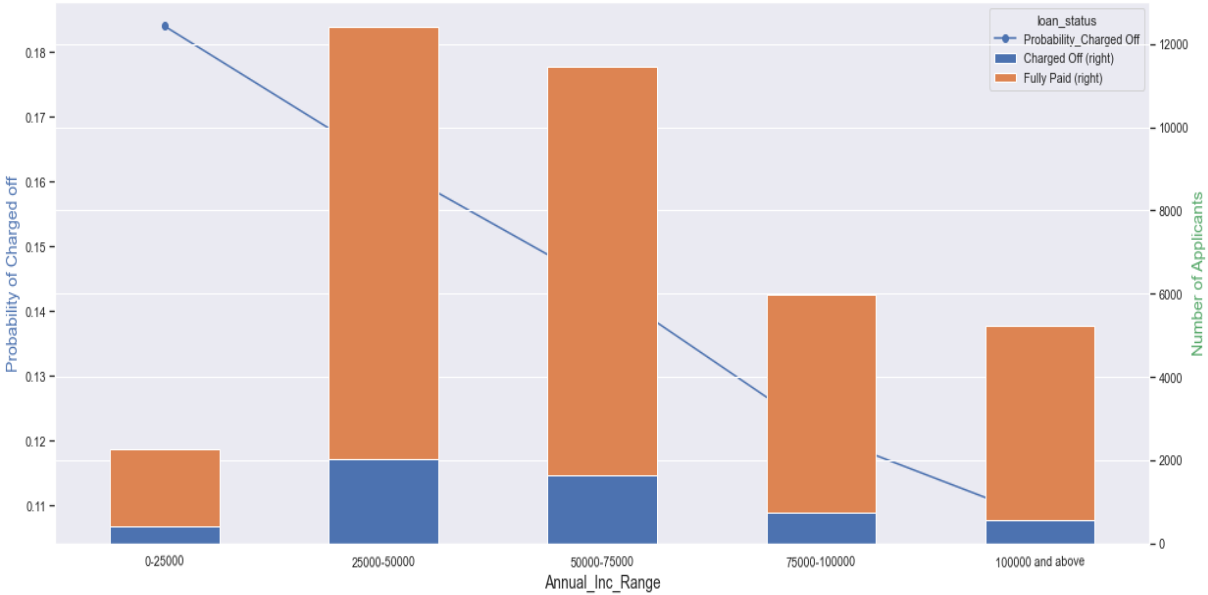
Int_Rate_Range vs Probability Charge Off



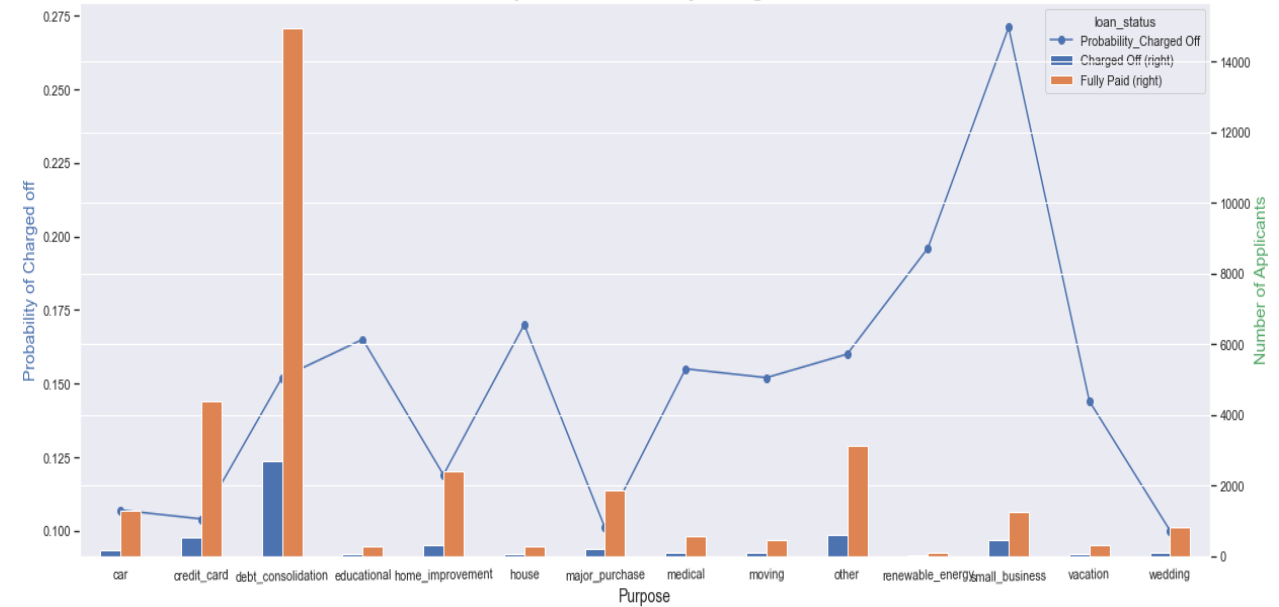
Grade vs Probability Charge Off



Annual_Inc_Range vs Probability Charge Off



Purpose vs Probability Charge Off



Observations

- Applicants who have taken the Loan for 'small business' have the highest probability of charge off of 27%. So bank should take extra caution like take some asset or guarantee while approving the loan for purpose of 'small business'
- There are multiple States/Provinces with high probability of charge, highest being 'NV' at 22%
- Applicants with more than 10 years of experience are more probable of charged off.
- As the interest rate is increasing the probability that person will default is increasing with highest of 9% at 15% & above bracket.
- As the annual income is decreasing the probability that person will default is increasing with highest of 18% at (0 to 25000) salary bracket.
- As we move from Grade A to G, probability that person will charged off is increasing.

Conculsion

Top-5 Major variables to consider for loan prediction:

1. Purpose of Loan
2. Interest Rate
3. Employment Length
4. Term
5. Grade