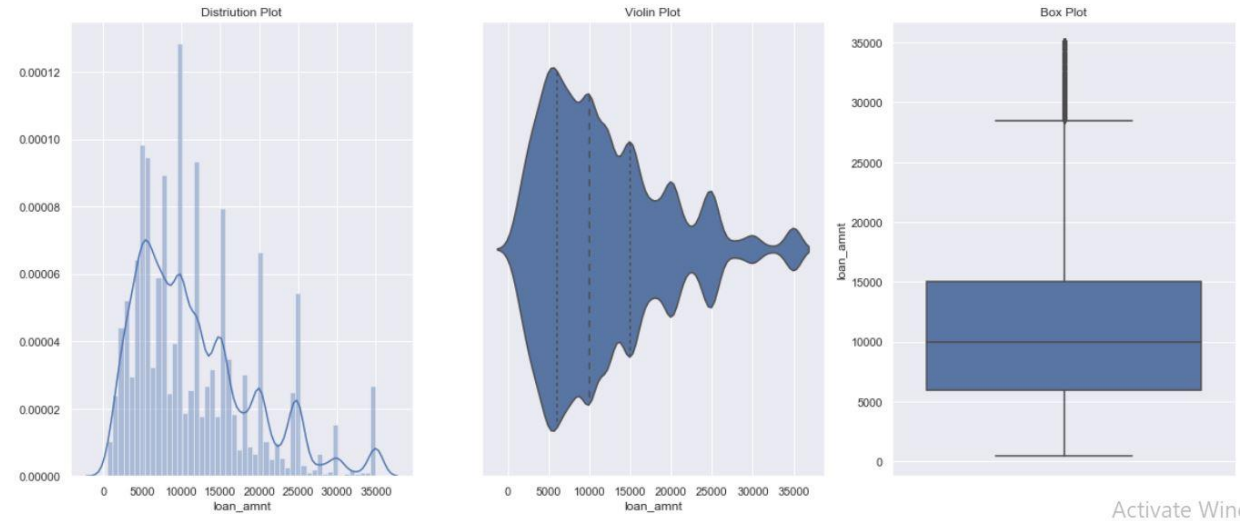
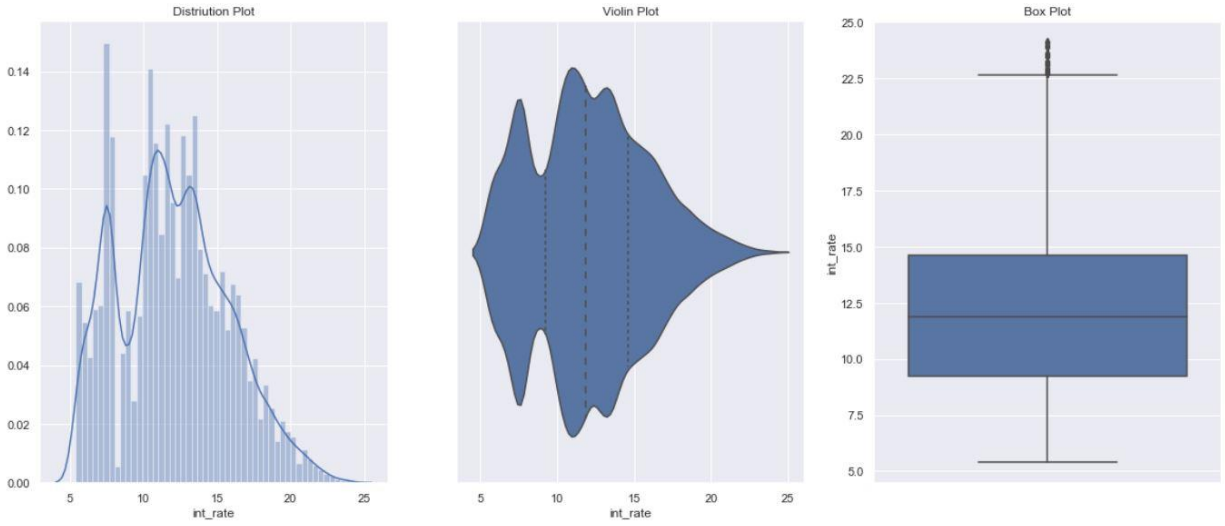
ANALYSIS RESULT OF DATASET

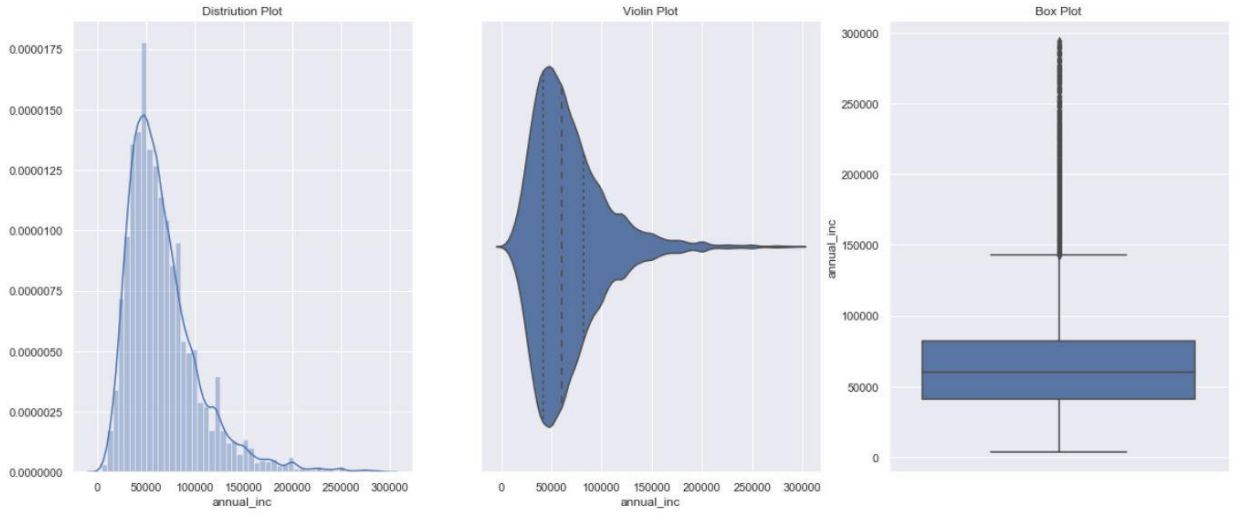
Univariate Analysis:



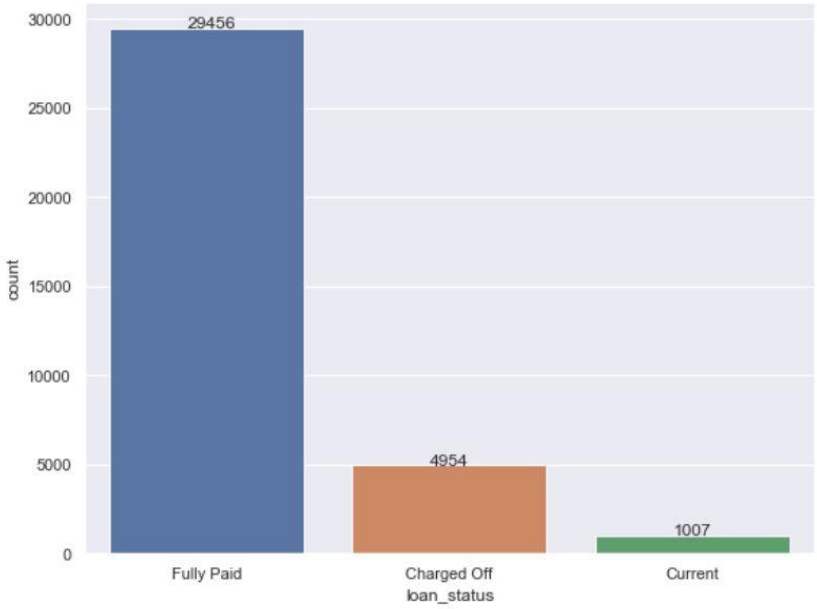
Most of the loan amounts are distributed between 8000 to 20000 USD



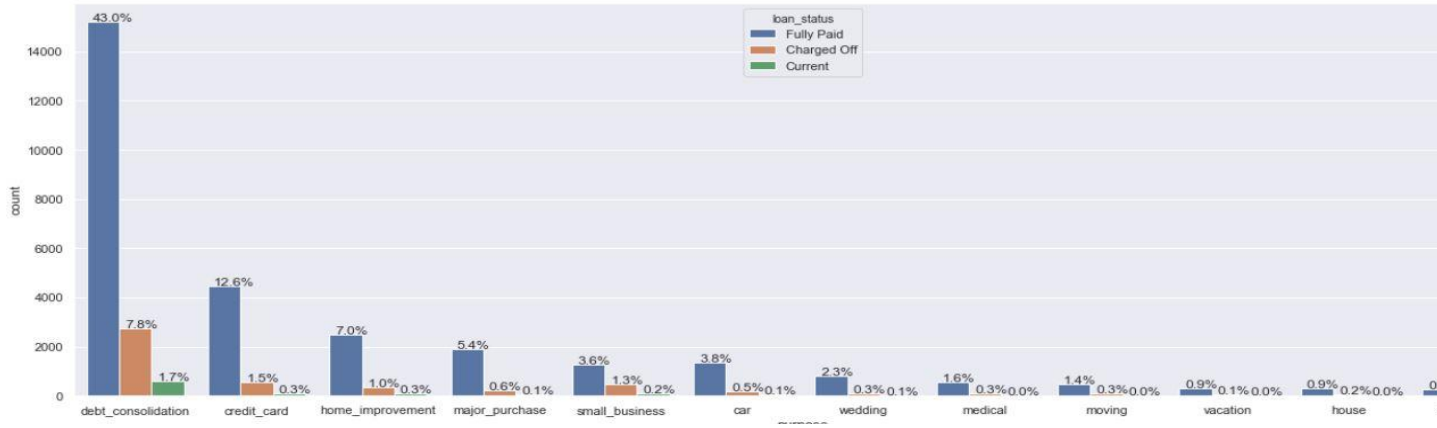
Most of the interest rate are distributed between 10% to 16%



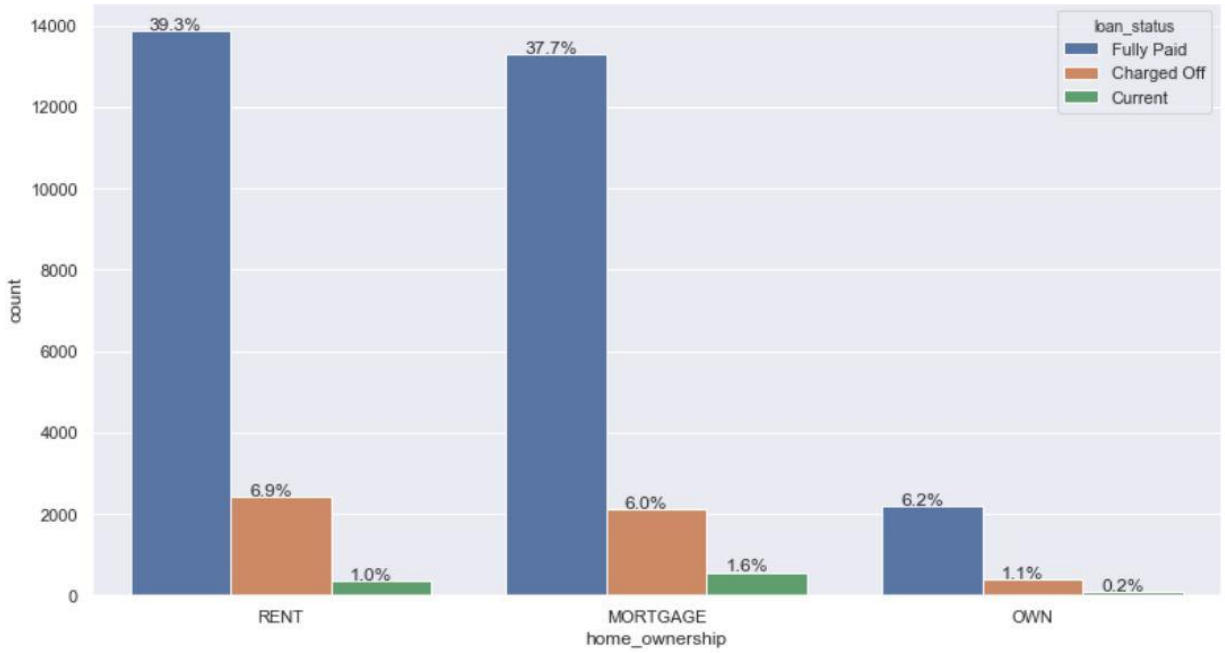
Most of the applicants earns between 40000 to 90000 USD annualy



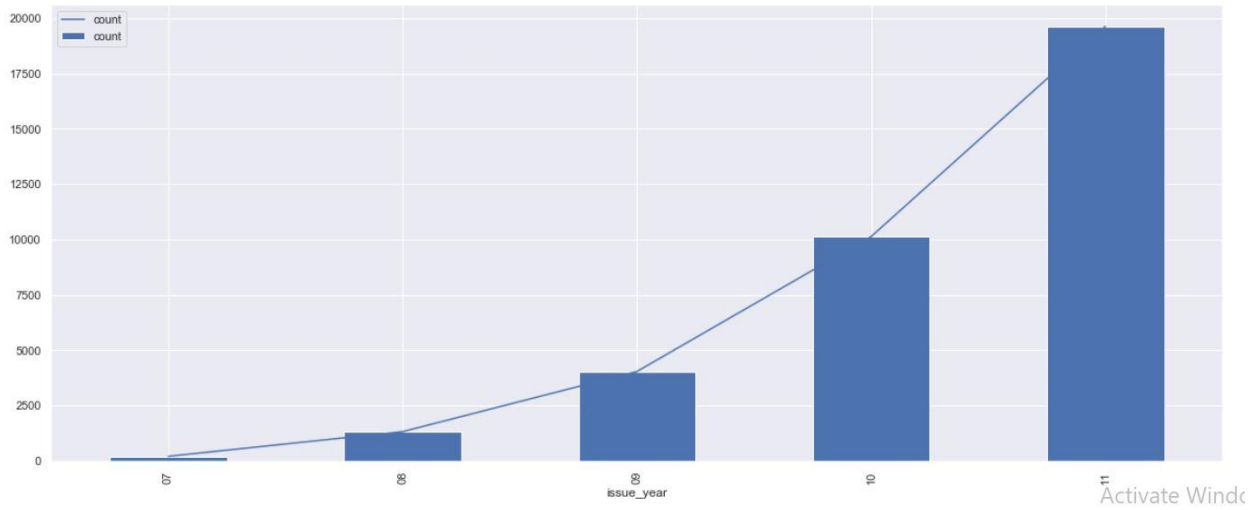
5 % of the applicants Charged Off



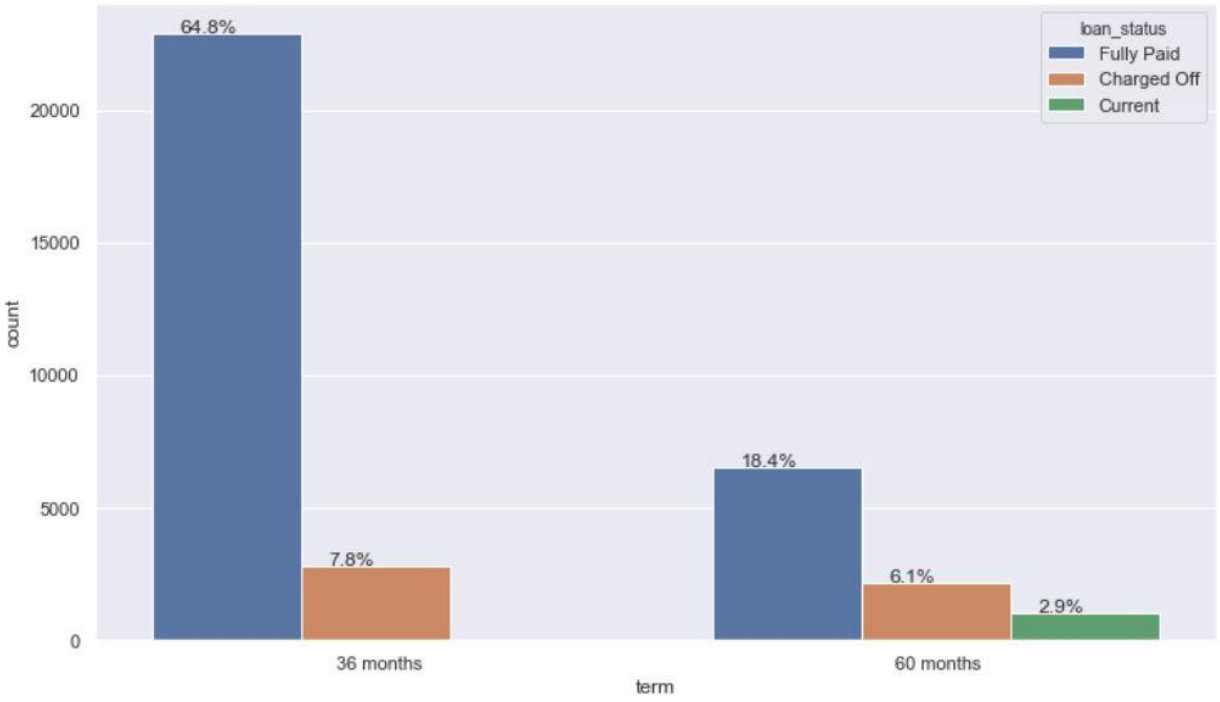
Approx 60% of the applicant applied loan for paying their other loan (Debt Consolidation)



40% of applicant are living in rented home whereas 52% applicants were mortgaged their home.



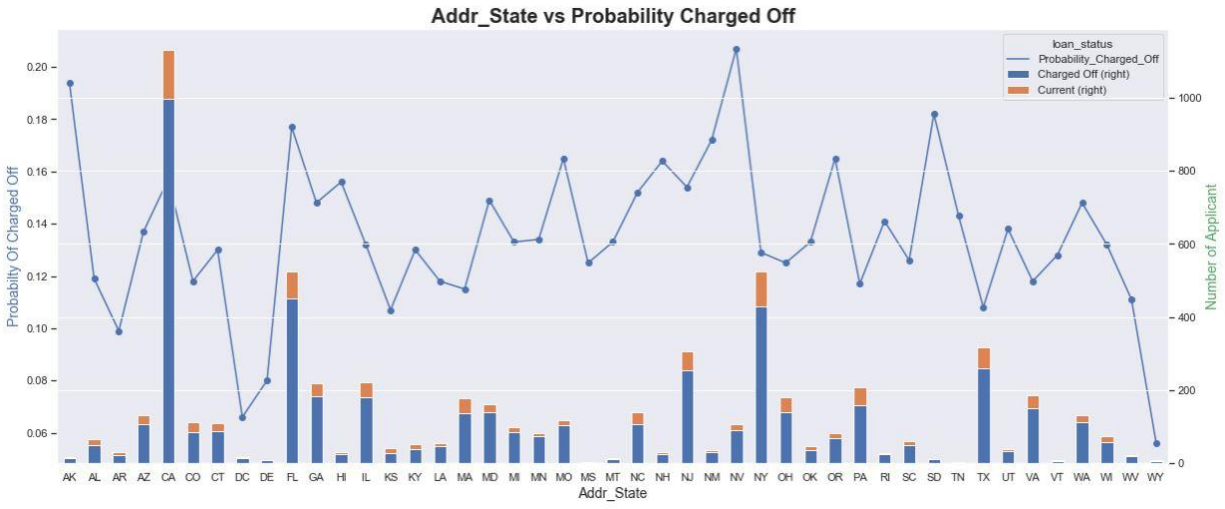
Loan Applicants are increasing year on year, approx. 47% of loan applicants received loan in 2011.



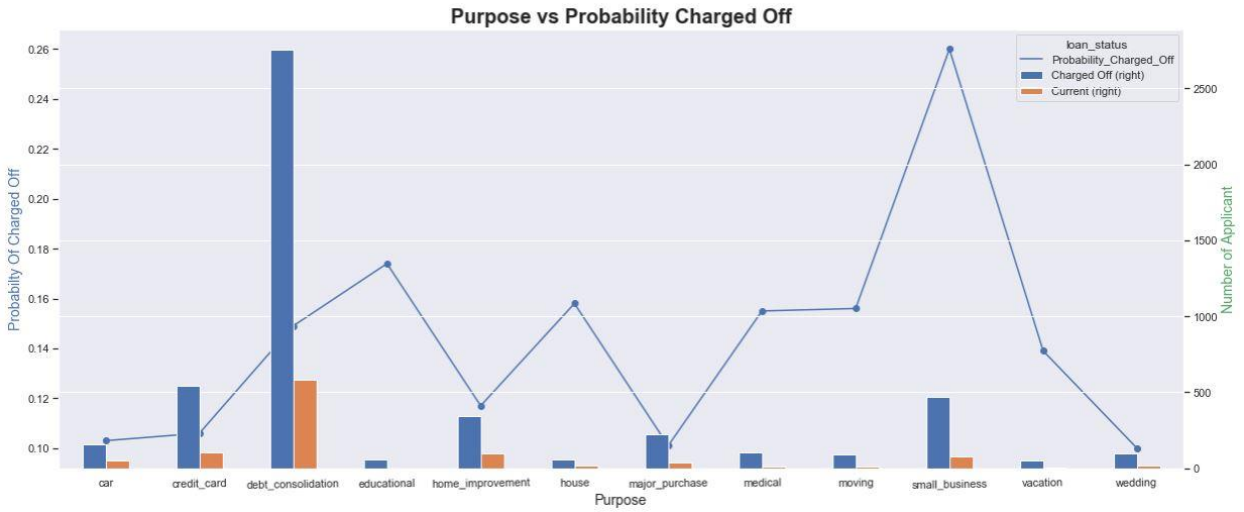
70% of applicants applied loan for 36 months term period.

Bivariate Analysis:

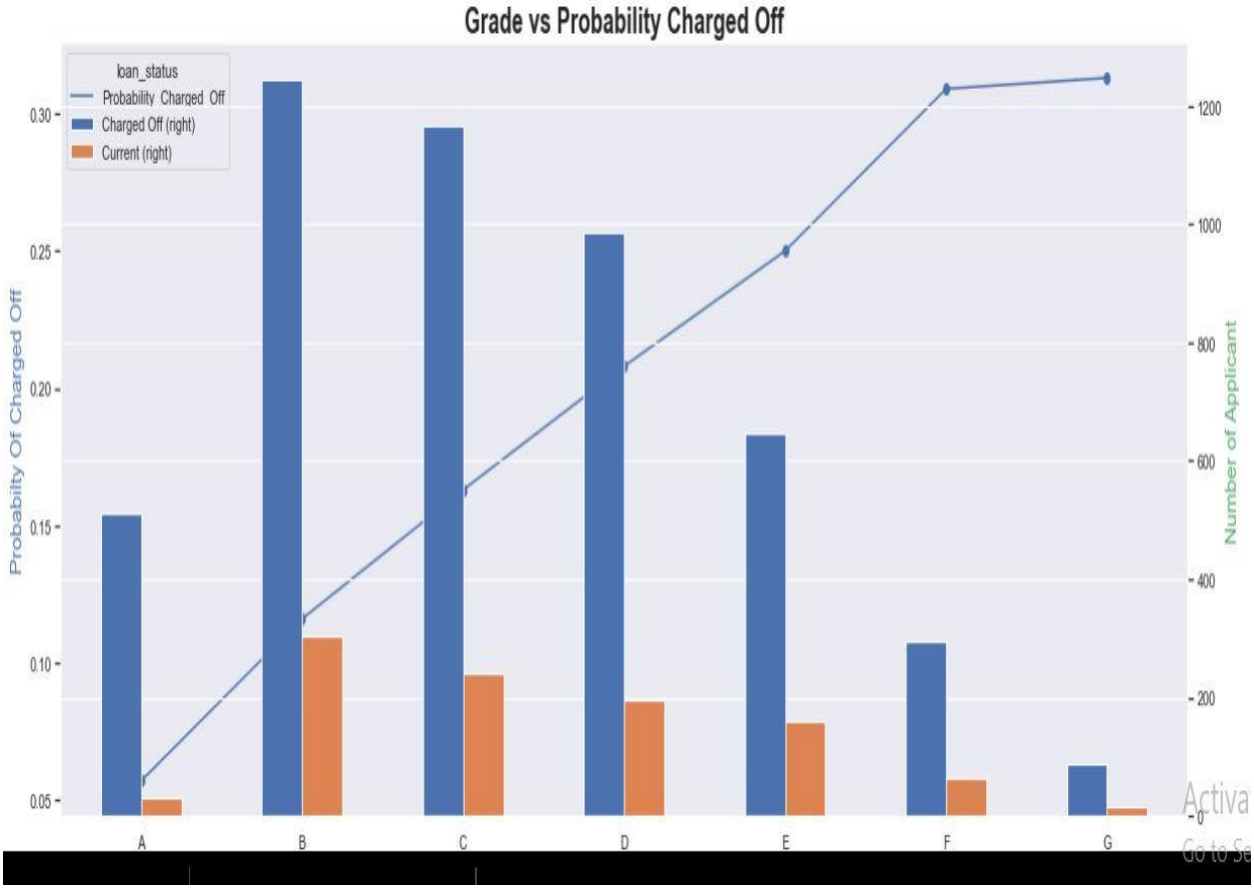
It is clear from the heat that how ‘loan_amnt’,’funded_amnt’,&’funded_amnt_inv’ are closely interrelated. So we can take any one column out of them for our analysis.

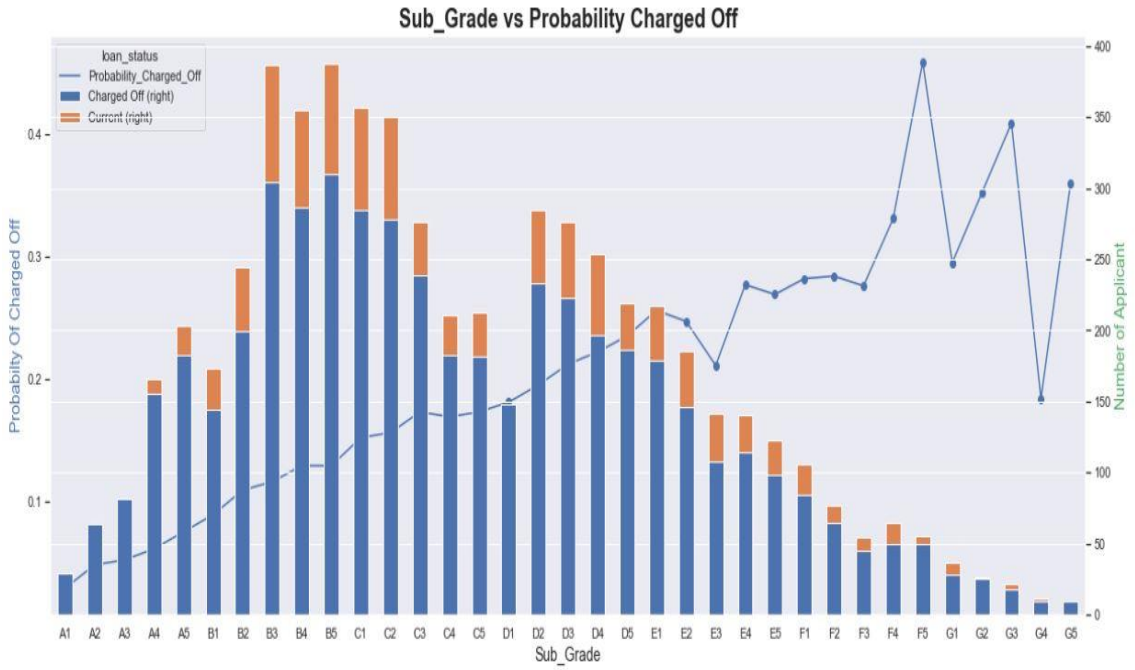


There are multiple States/Provinces with high probability of charge, highest being ‘NV’ at 7%

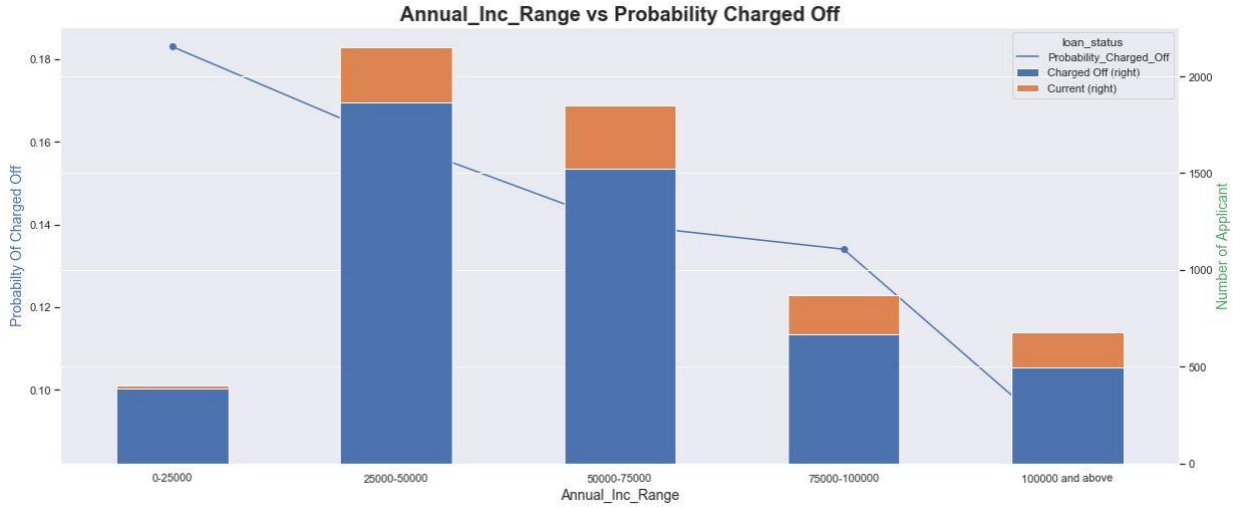


Applicants who has taken the loan for small business has the highest probability of charge off of 14%. So bank should take extra caution like take some asset or guarantee while approving the loan for purpose of ‘small business’

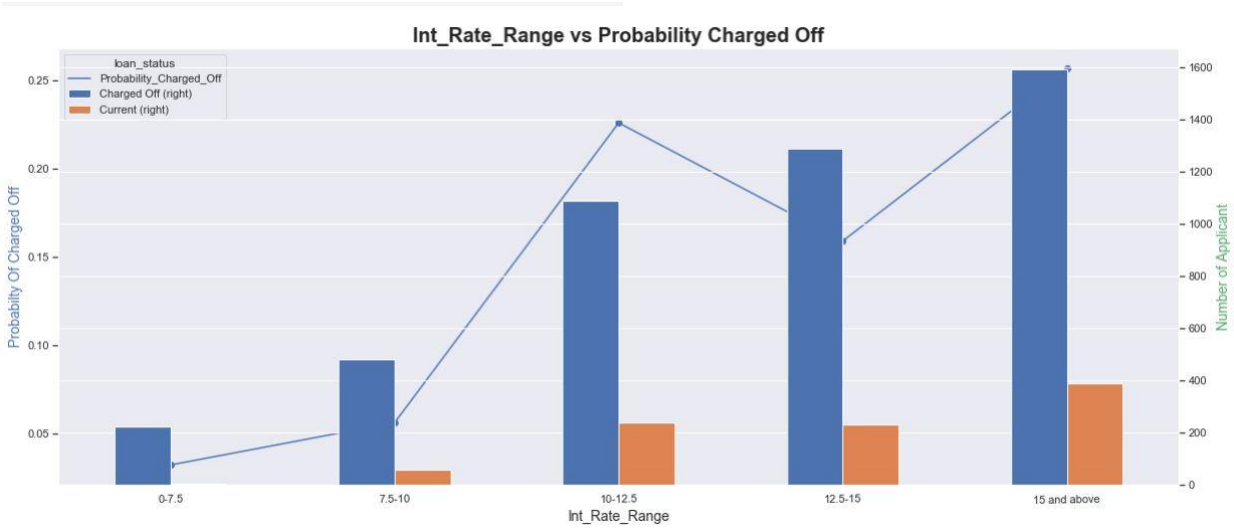




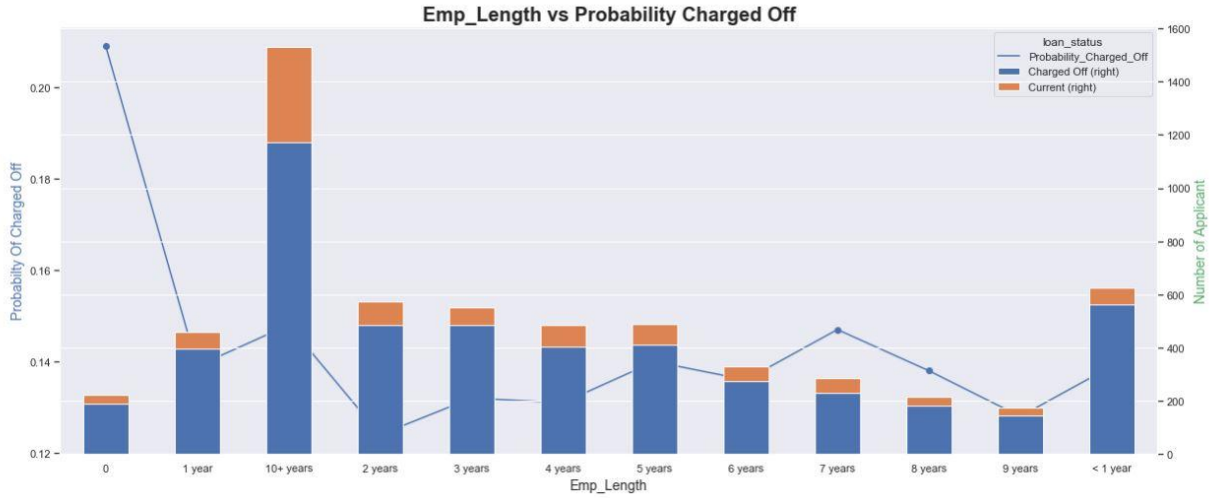
As we move from grade A to G , Probability that person will charged off is increasing.



As the annual income is decreasing the probability that person will be default is increasing with highest of 7% at (0-25000) salary bracket



As the interest rate is increasing the probability that person will default is increasing with highest of 9% at 15% and above bracket



Applicant who are self employed and less that 1 year of experience are more probable of charged off

Top 5 Major Variables to consider for loan prediction are:

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