LENDING CLUB CASE STUDY

GROUP MEMBERS –

1. Ishika Singhal
2. Rhea Thomas

PROJECT ABSTRACT

1. AIM - To identify patterns which indicate if a person is likely to default and not likely to pay back the loan.
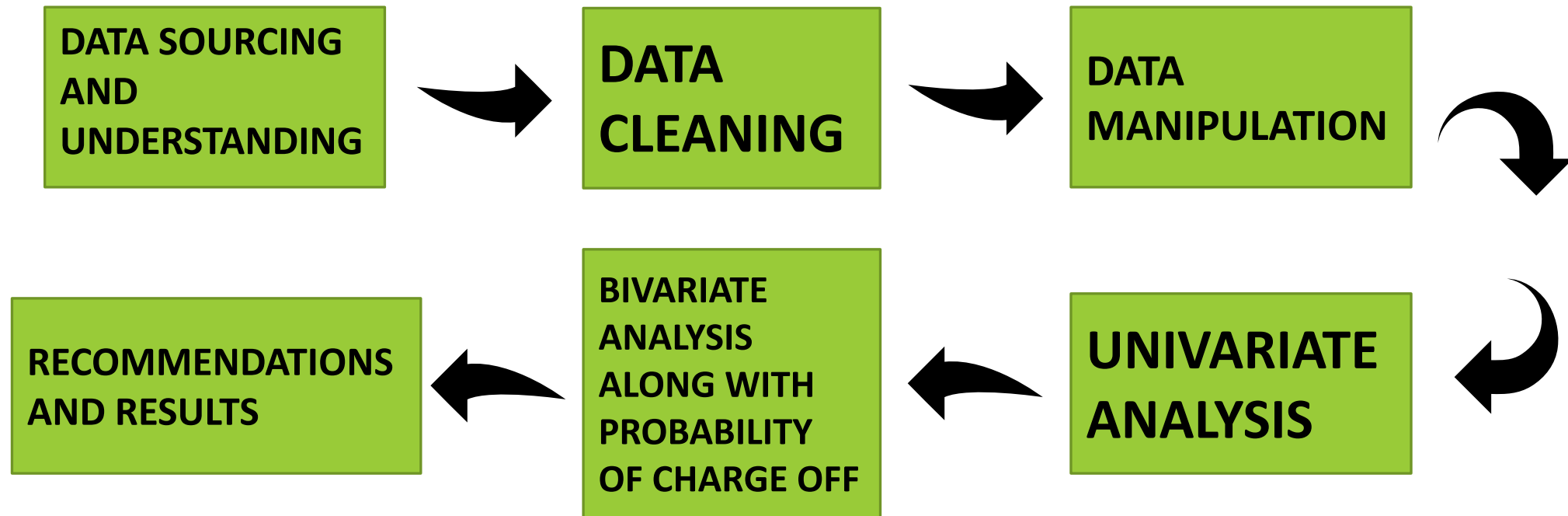
2. PROBLEM STATEMENT - When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
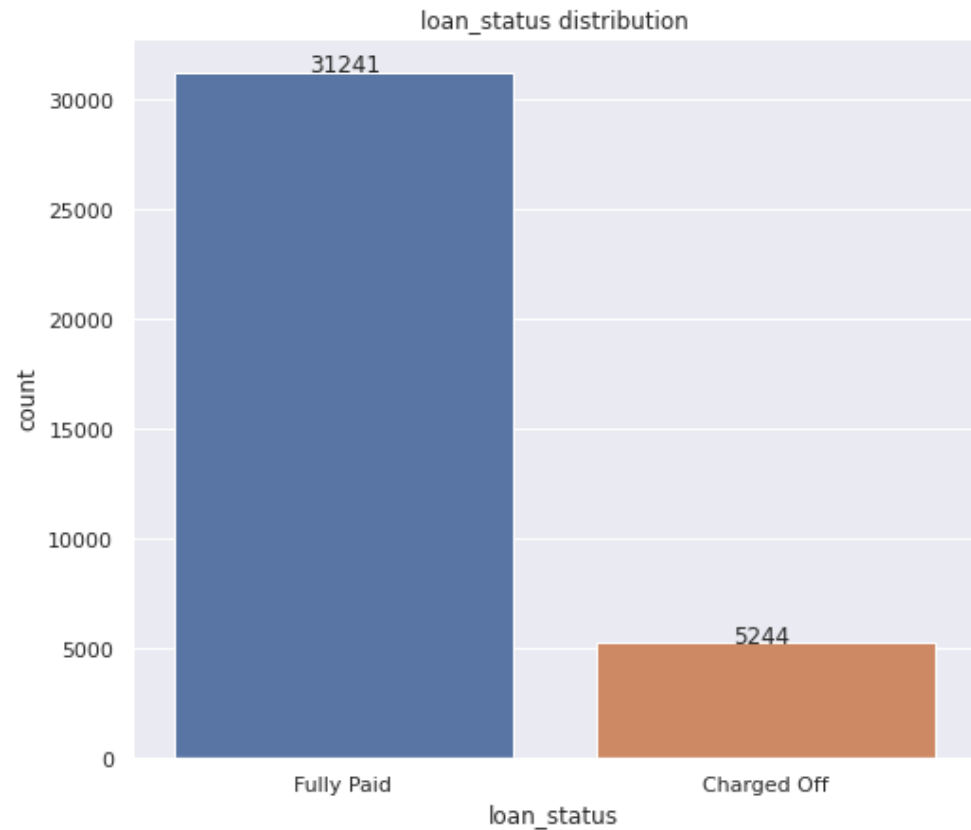
If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

So, company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
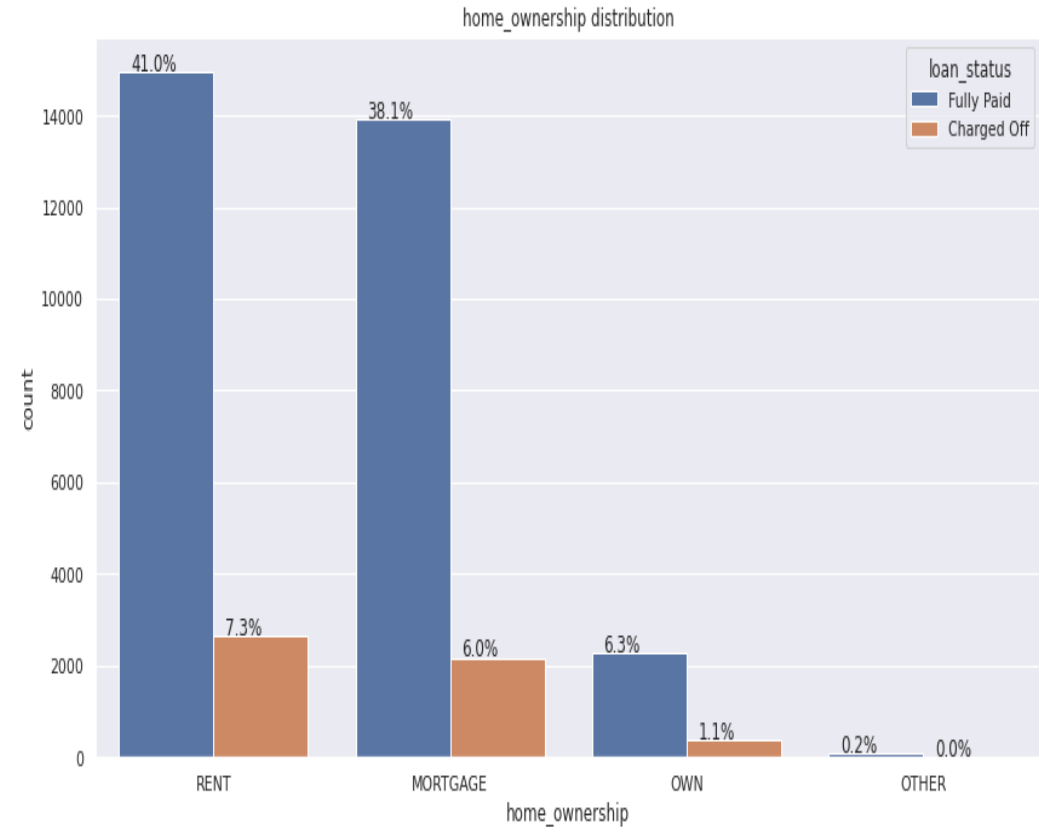
PROBLEM SOLVING METHODOLOGY

DATA SOURCING AND UNDERSTANDING → **DATA CLEANING** → **DATA MANIPULATION** →

RECOMMENDATIONS AND RESULTS ← **BIVARIATE ANALYSIS ALONG WITH PROBABILITY OF CHARGE OFF** ← **UNIVARIATE ANALYSIS** ←

UNIVARIATE ANALYSIS

loan_status distribution

home_ownership distribution

Around 85.6% loans are fully paid and 14.3% loans are charged off.

Most of the borrowers who have taken loan are living in rented home or mortgage their home.

purpose distribution
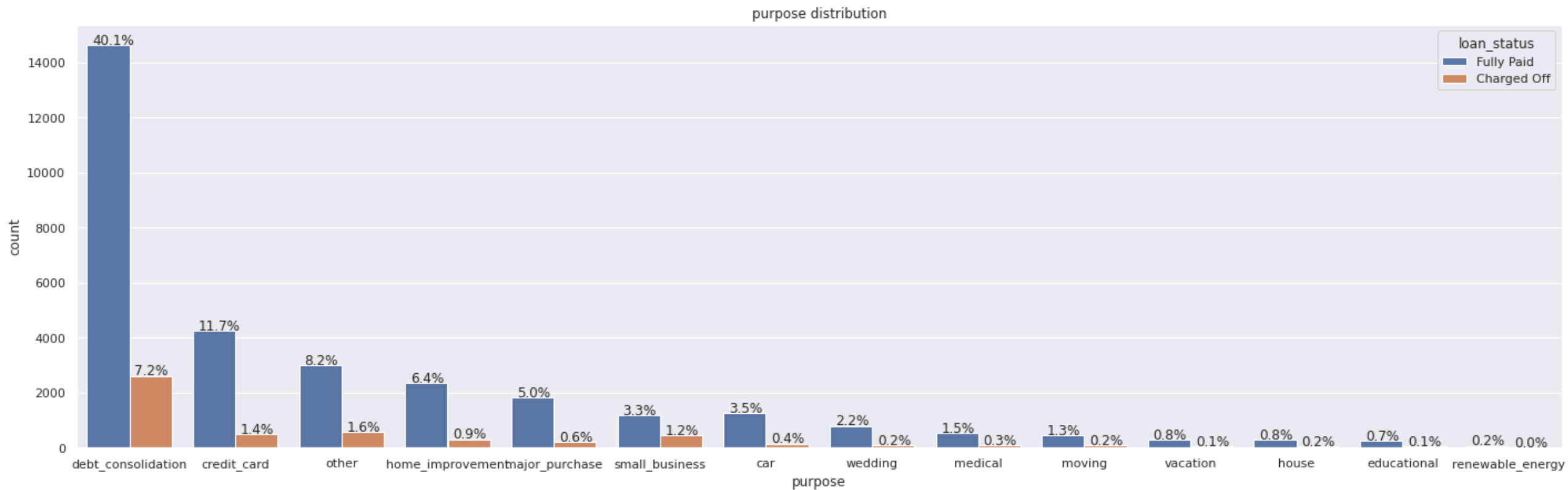
Approx 50% of the borrowers took loan for debt consolidation. Charged off loans percentage is maximum for the loan purpose of –

1. debt consolidation
2. paying credit card bills
3. others

term distribution
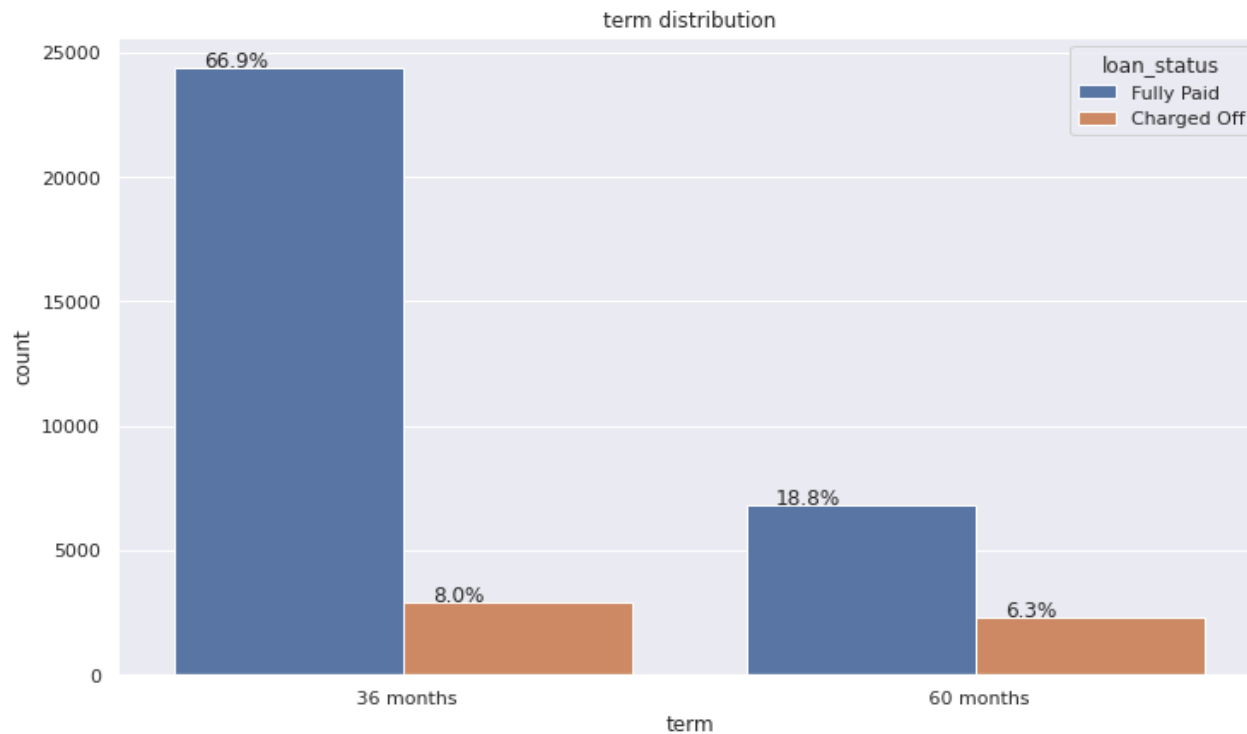
1. Maximum loans borrowers take loan for 36 months tenure.

2. The number of charged off loans is more for 36 months tenure as compared to 60 months.

3. Most of the borrowers are from CA, NY and FL.

4. Charged off loan percentage is also maximum for these states.

addr_state distribution

verification_status distribution

emp_length distribution
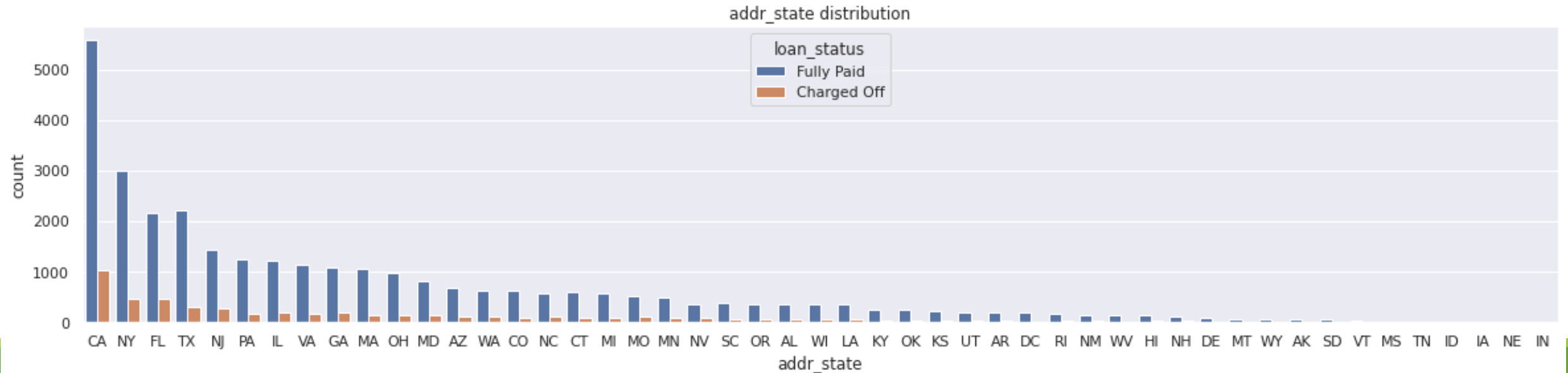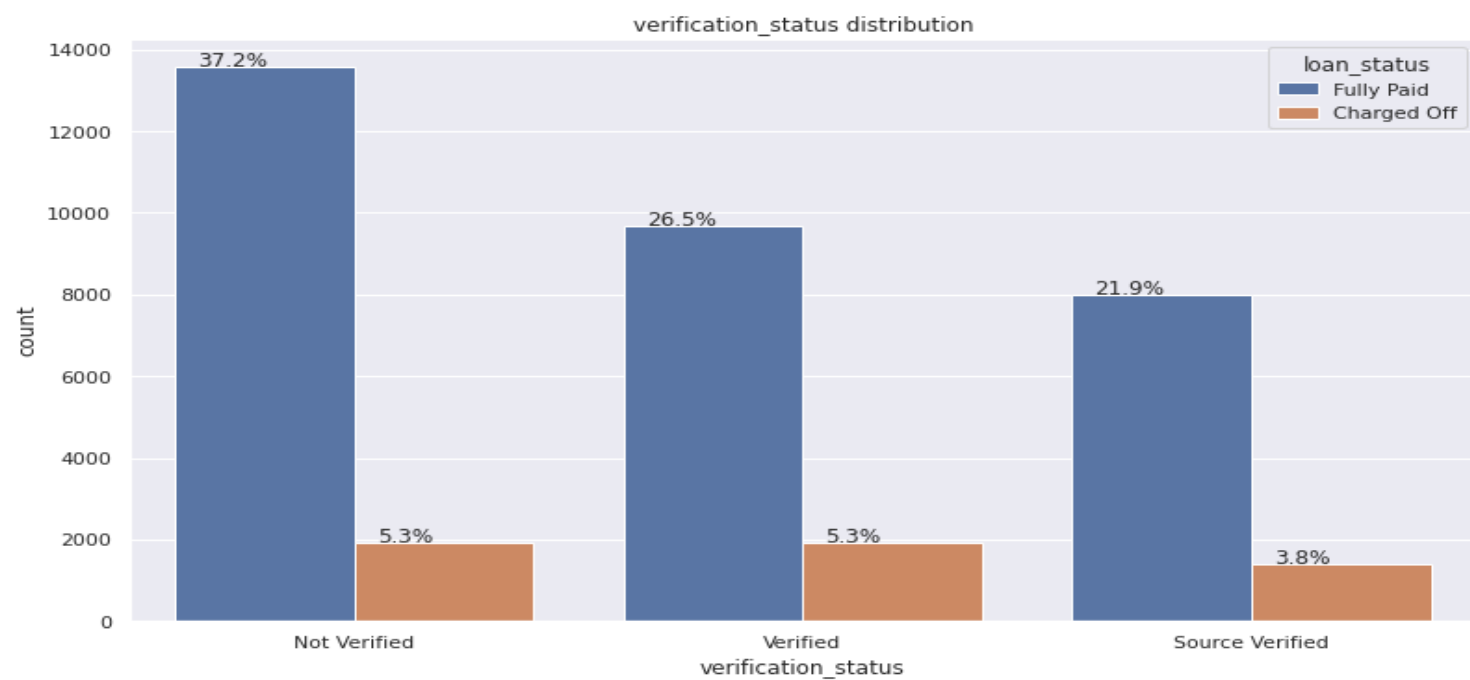
1. Maximum number of borrowers do not have their income verified.
2. Charged off loans are mostly from not verified and verified category of borrowers.

3. Maximum borrowers are from 10+ category.
4. The maximum number of borrowers with charged off loan have experience of either 10+ or < 1 year.
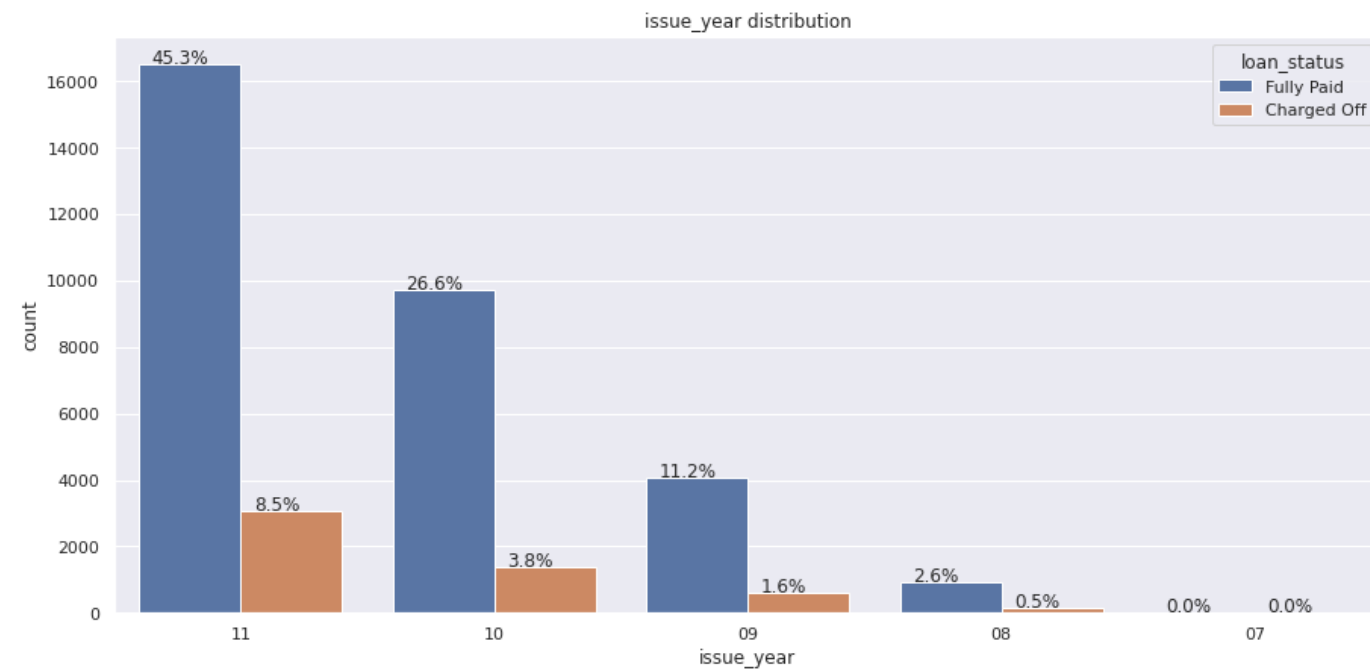
grade distribution

sub_grade distribution
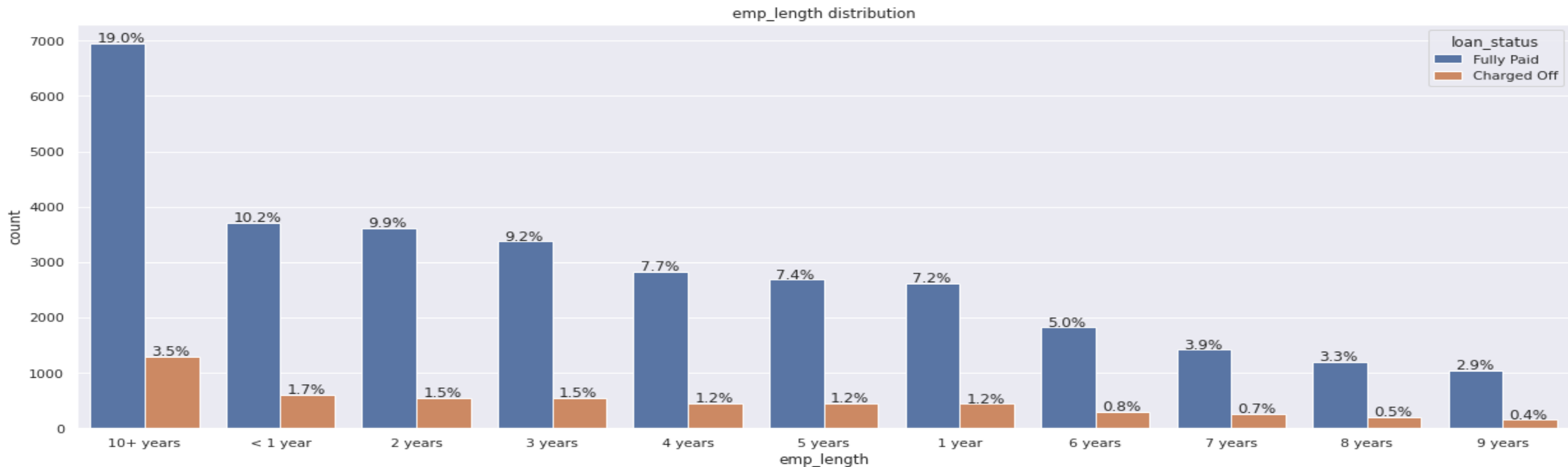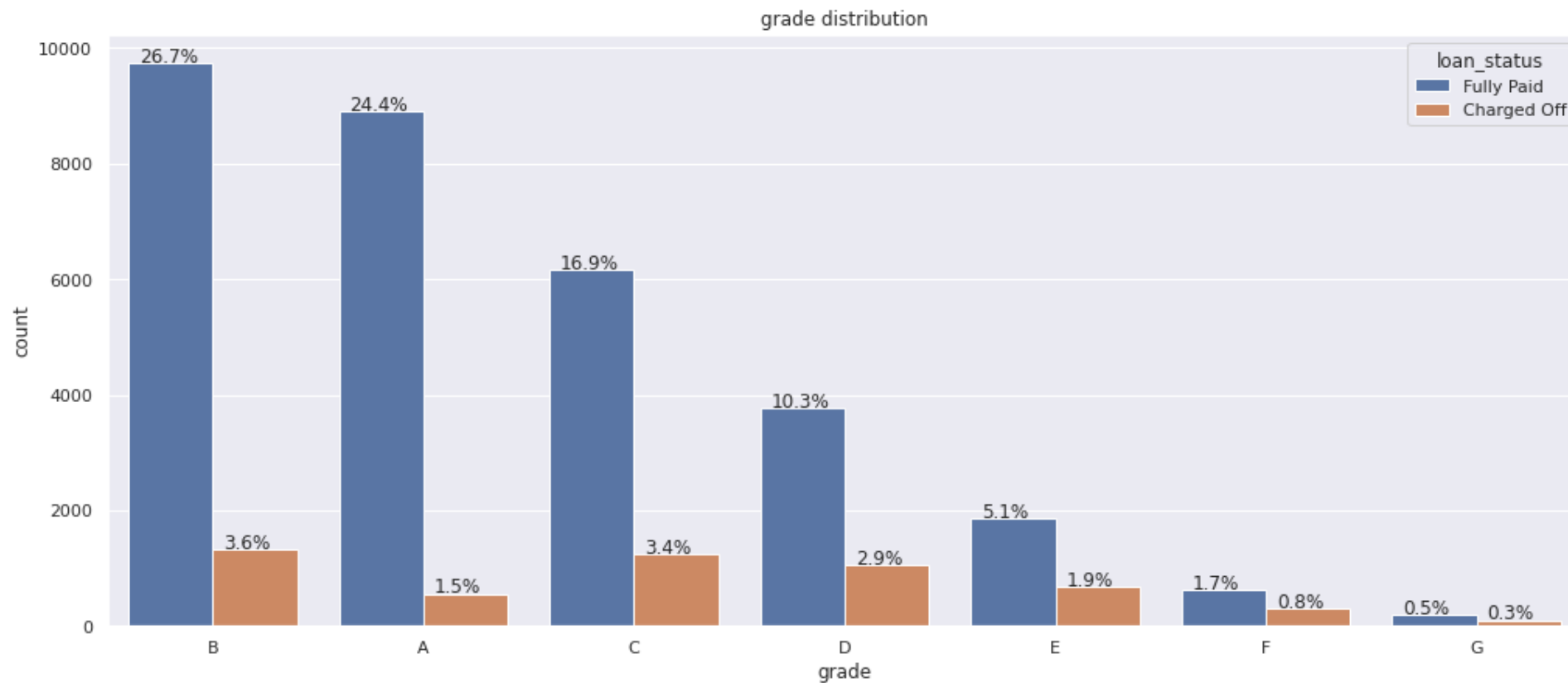
1. Maximum number of charged off loans are from B, C and D grade.

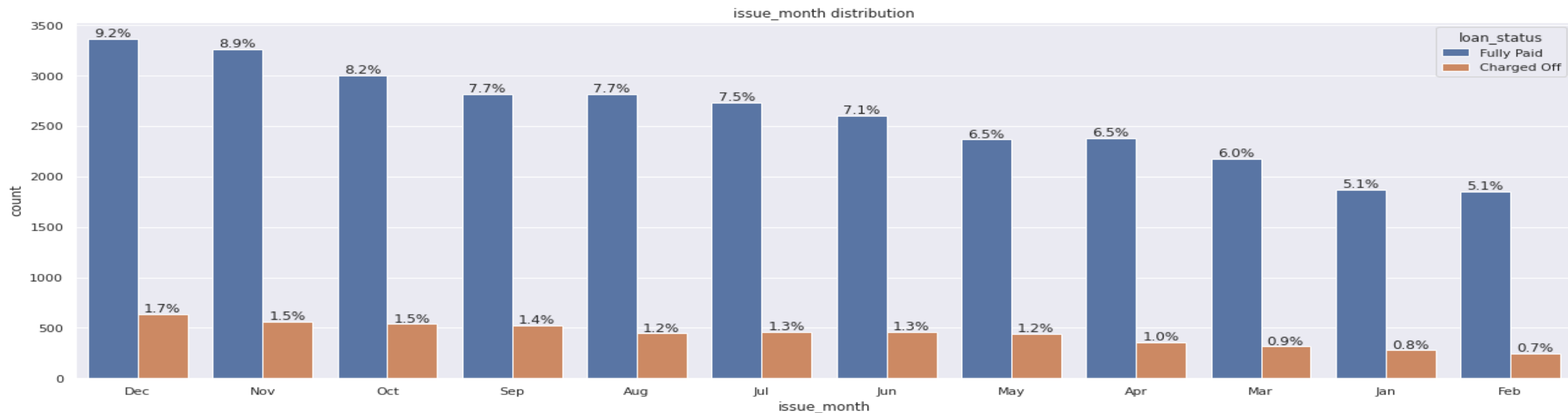2. Maximum number of charged off loans are from sub grade B3, B4 and B5.
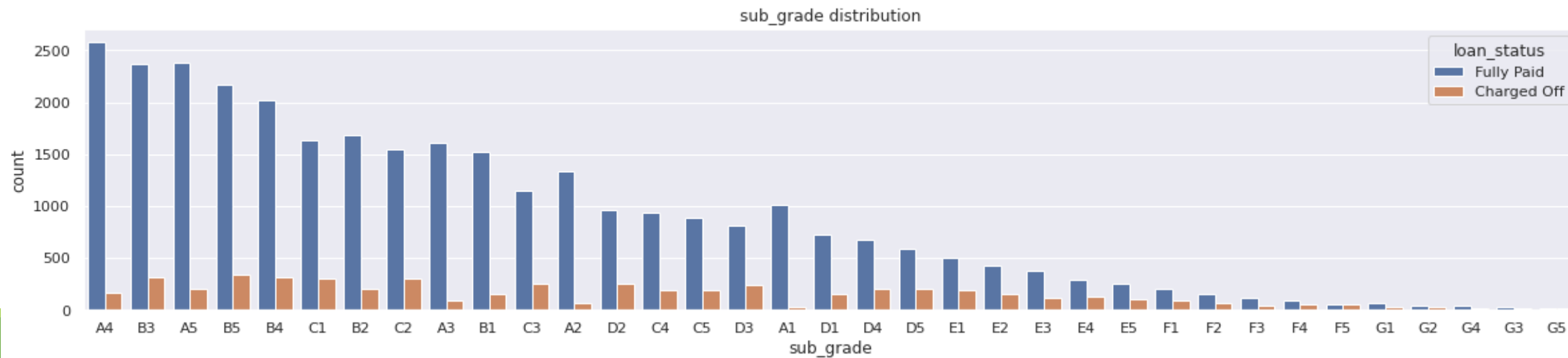
1. The number of loans issued are almost doubled every year

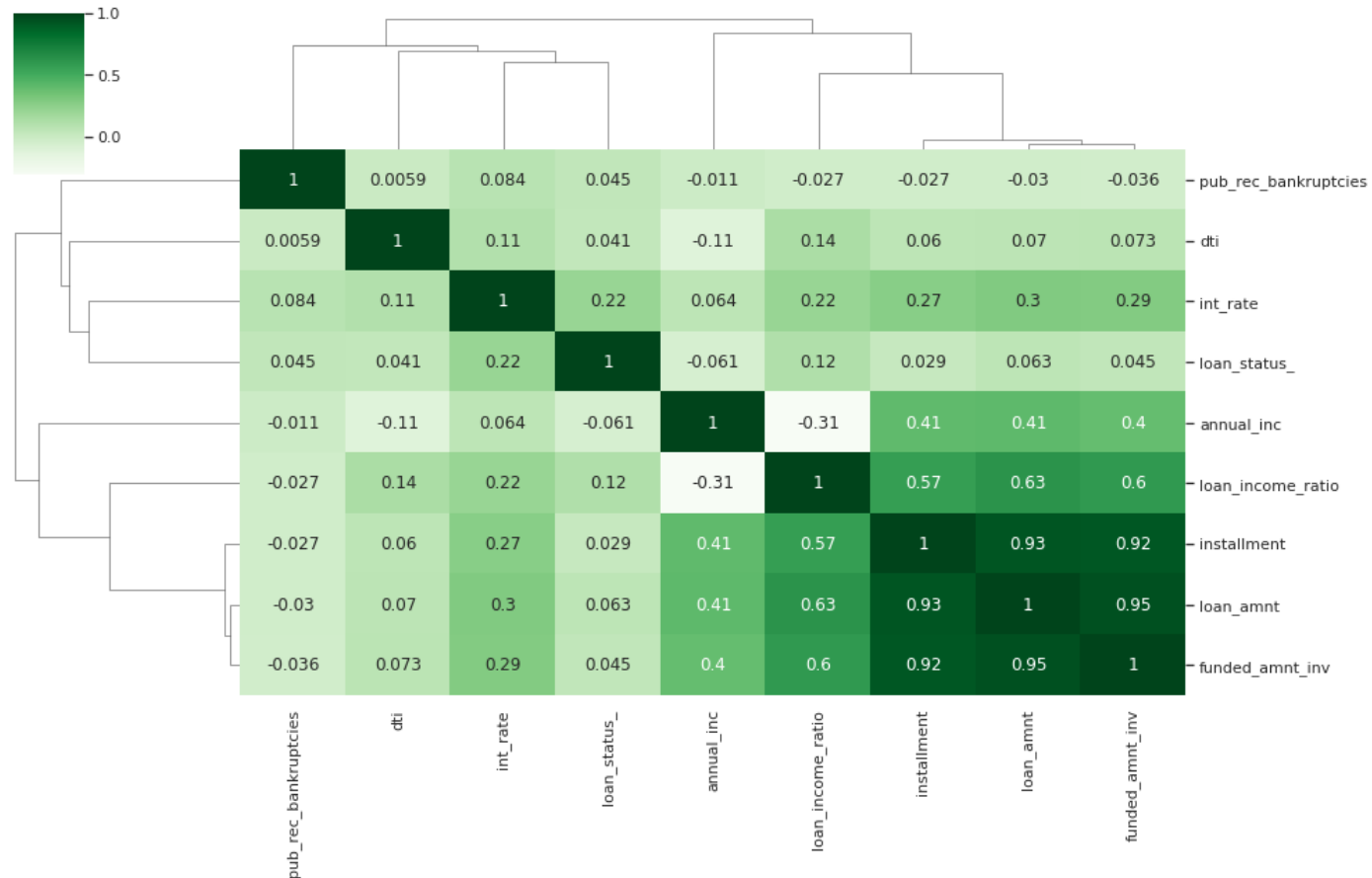2. The maximum number of charged off loans were issued in the year 2011.

3. The number of loans issued increase towards the end of the year

4. The maximum number of charged off loans are borrowed at the end of year.
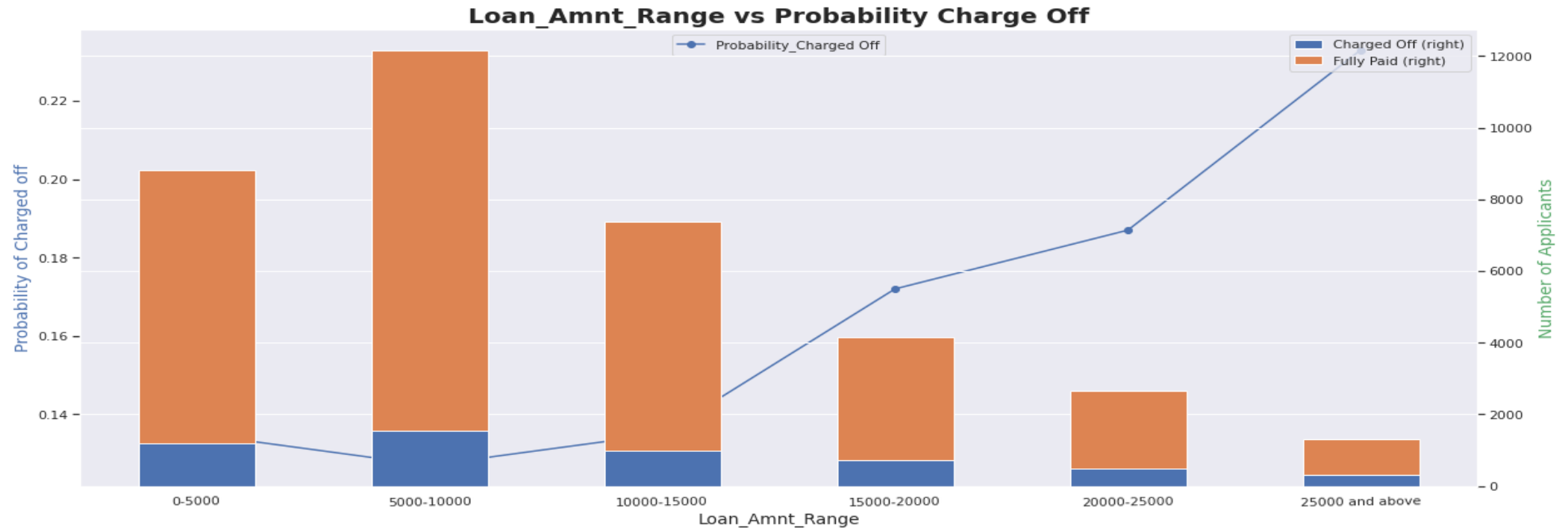
CORRELATION BETWEEN CONTINUOUS VARIABLES



Correlation between Columns

1. loan_amnt, funded_amnt_inc, installment are highly positively correlated.

2. annual_inc and dti are negatively correlated.

BIVARIATE ANALYSIS WITH PROBABILITY OF CHARGED OFF
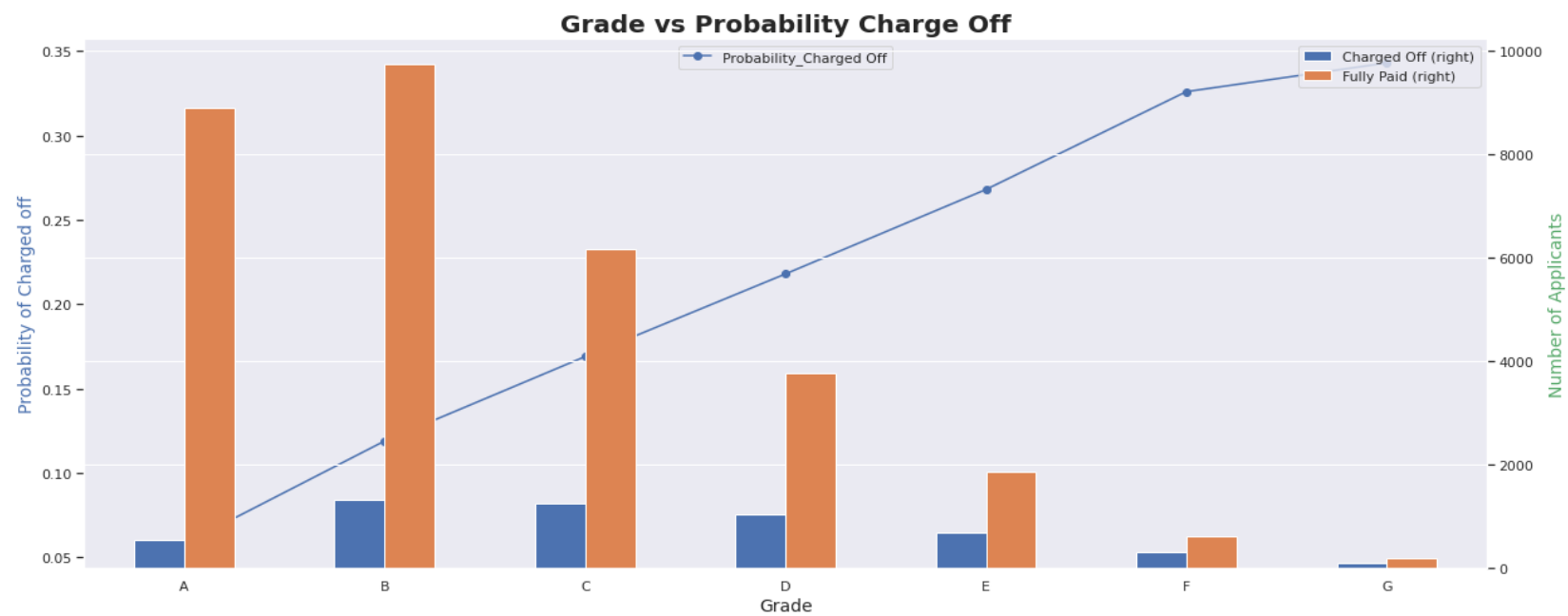

Loan_Amnt_Range vs Probability Charge Off

1. As the loan amount is increasing the probability that an applicant will be a defaulter is increasing
2. The maximum probability for a loan to get charged off is for loan amount of 25k and above

Term vs Probability Charge Off
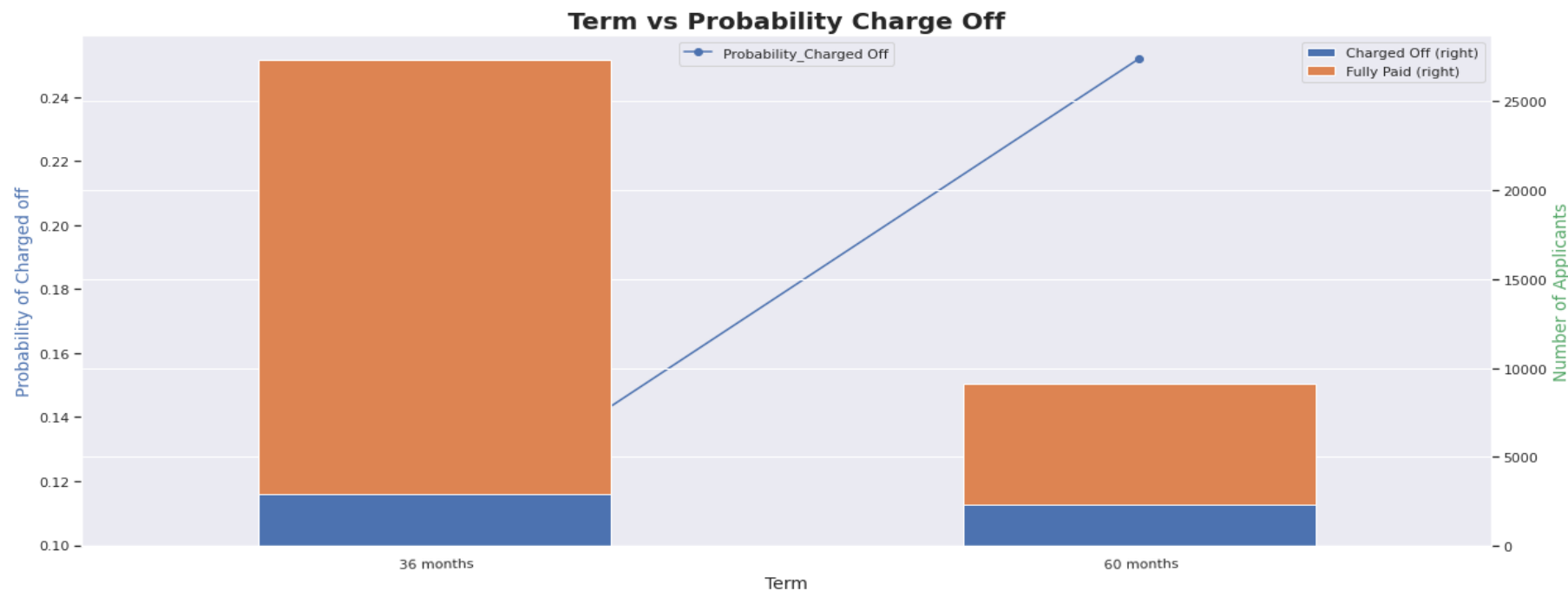
1. The loans with tenure of 60 months are having the maximum probability of being charged off

Home_Ownership vs Probability Charge Off
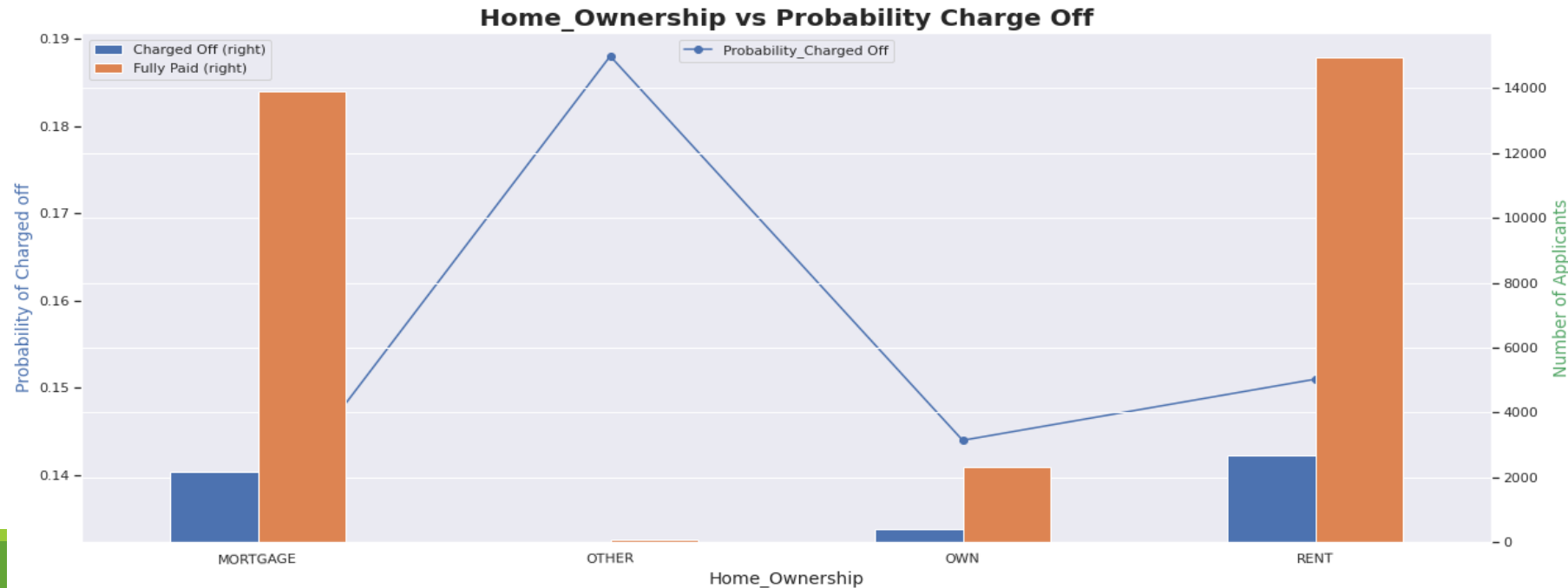
1. Applicants who have home ownership as others or rent are most likely to be a defaulter

Grade vs Probability Charge Off
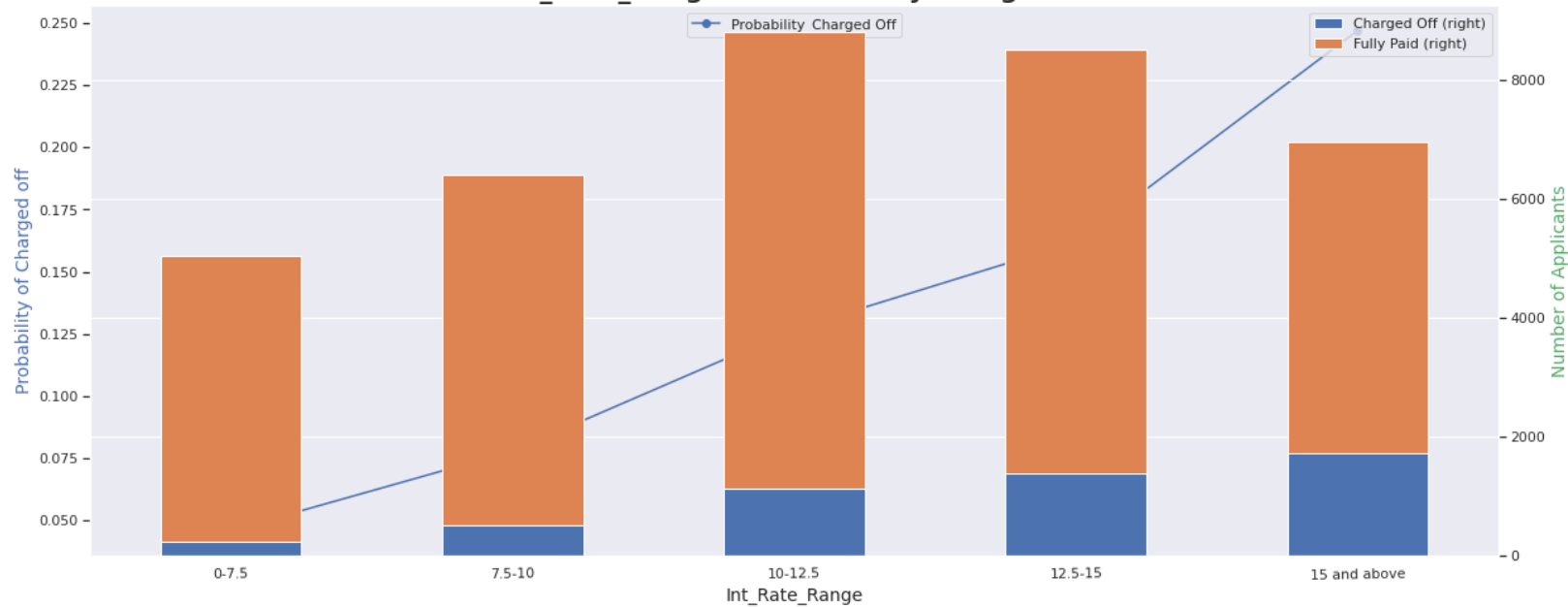
Sub_Grade vs Probability Charge Off

1. As we move from Grade A to G, probability that applicant will be a defaulter is increasing.

2. Subgrades F5, G3 and G5 are having maximum probability of being a defaulter.
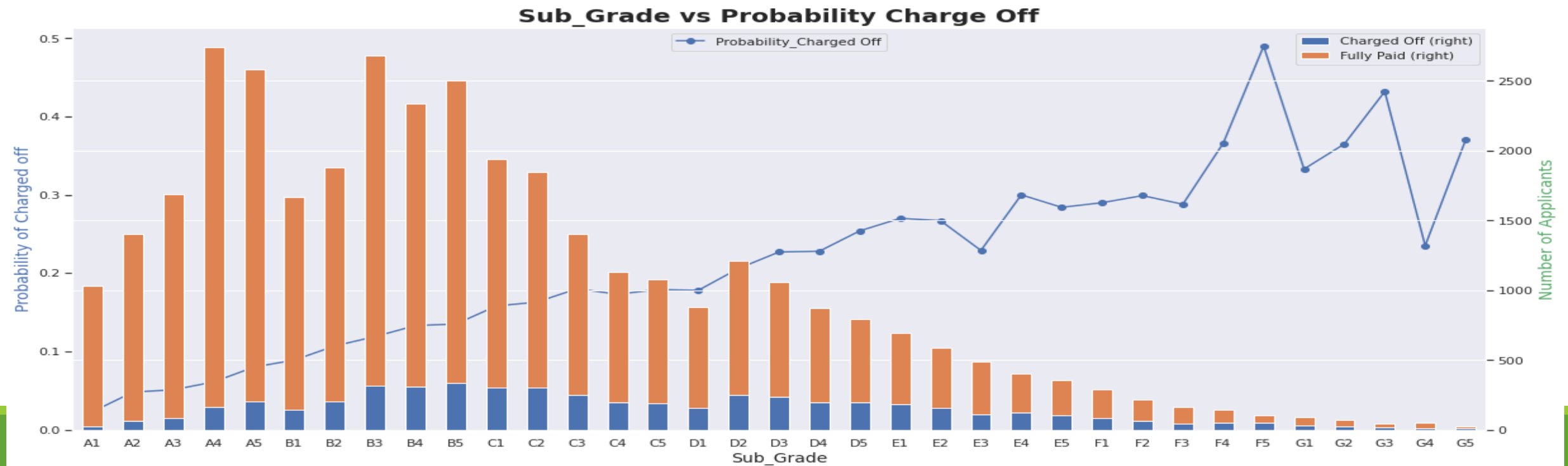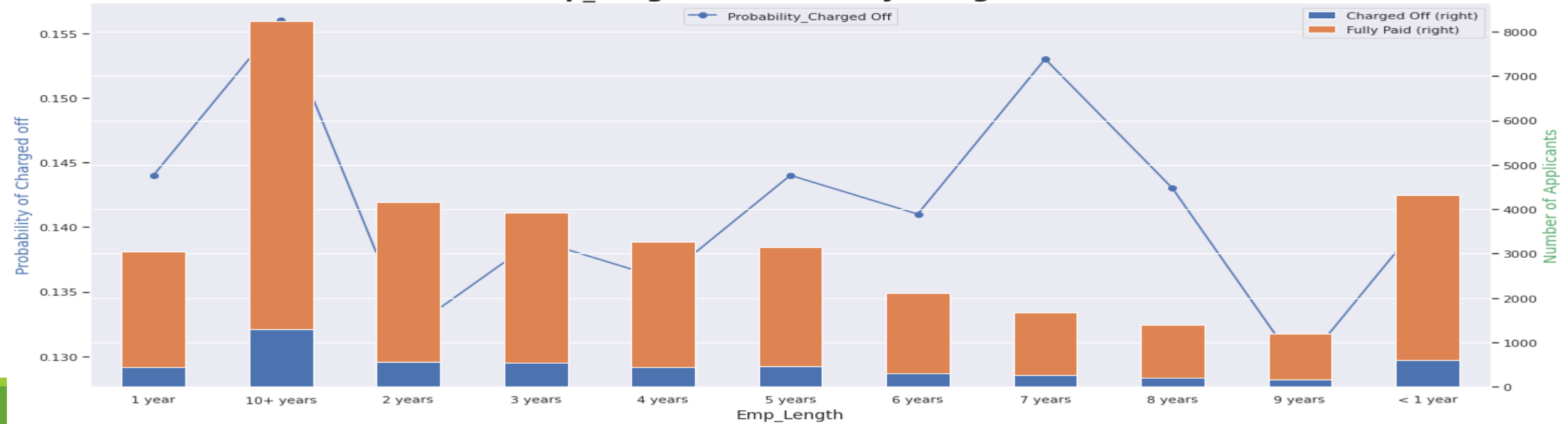
Int_Rate_Range vs Probability Charge Off
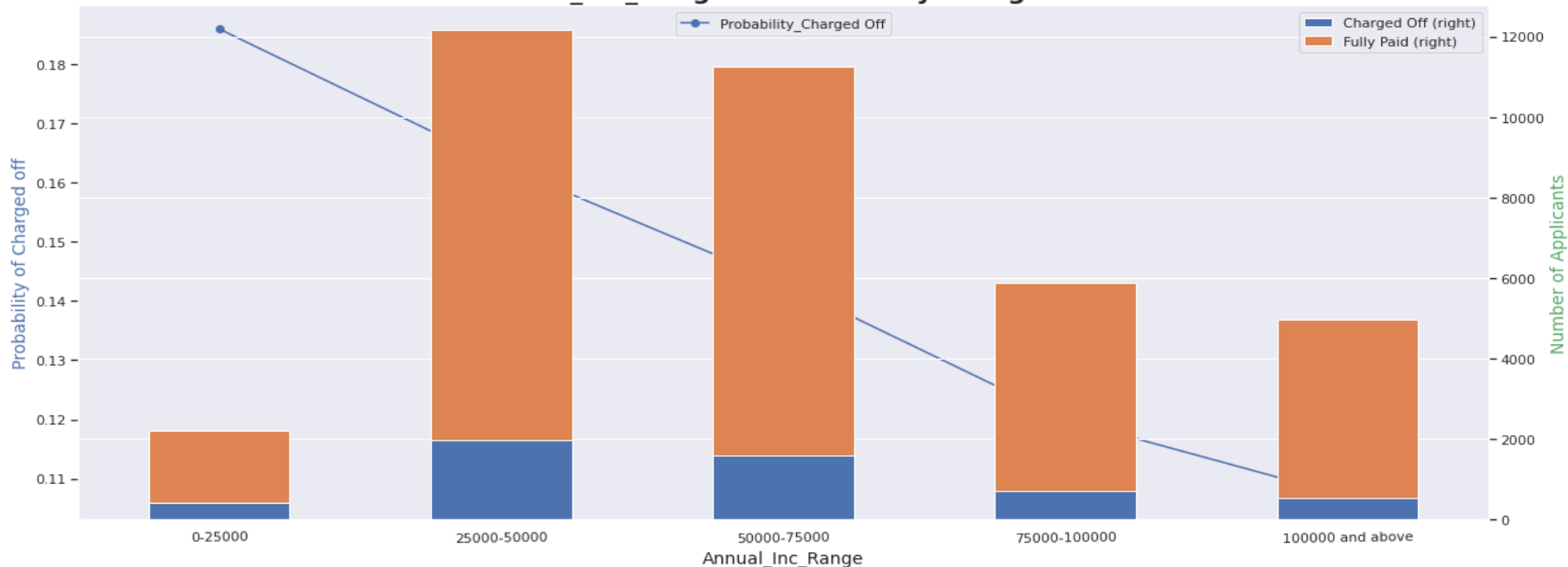
Emp_Length vs Probability Charge Off

1. As the interest rate is increasing the probability that an applicant will be a defaulter is increasing

2. The highest probability for a loan to get charged off is at an interest rate of 15 and above

3. Applicants with experience of 10+, < 1 and 1 year are most likely to be defaulters
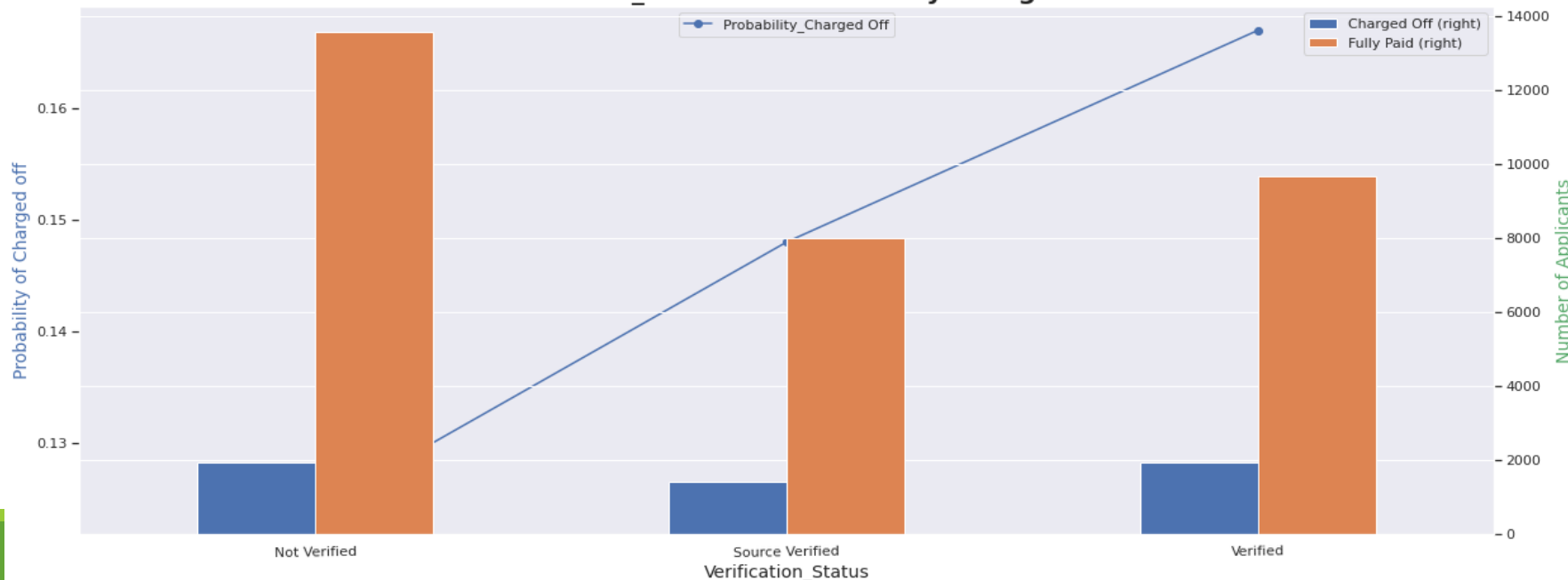
Annual_Inc_Range vs Probability Charge Off

1. As the annual income is decreasing the probability that applicant will be a defaulter is increasing

2. The highest probability that an applicant will be a defaulter is of the one with 0 to 25k annual income.

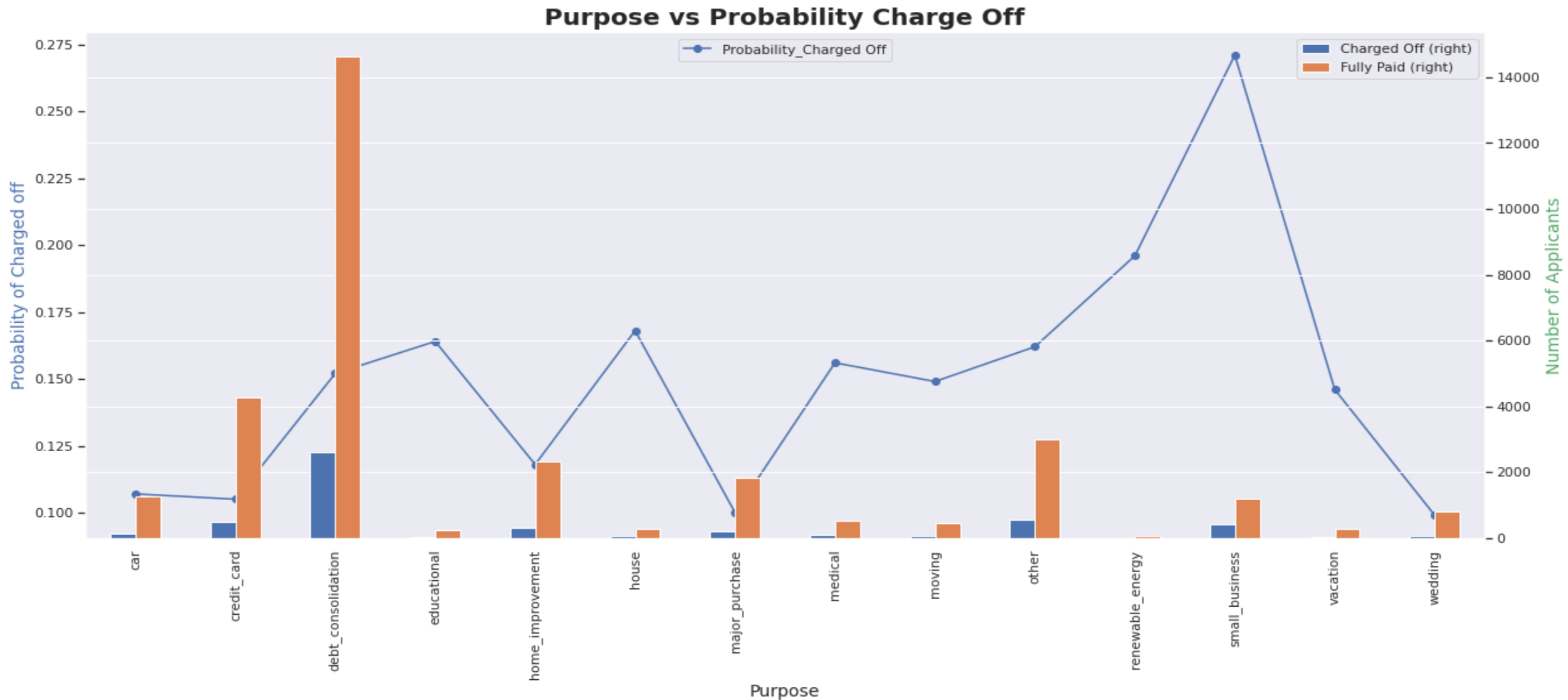**Verification_Status vs Probability Charge Off**

1. Applicants having their income verified are most likely to be a defaulter

Issue_Month vs Probability Charge Off
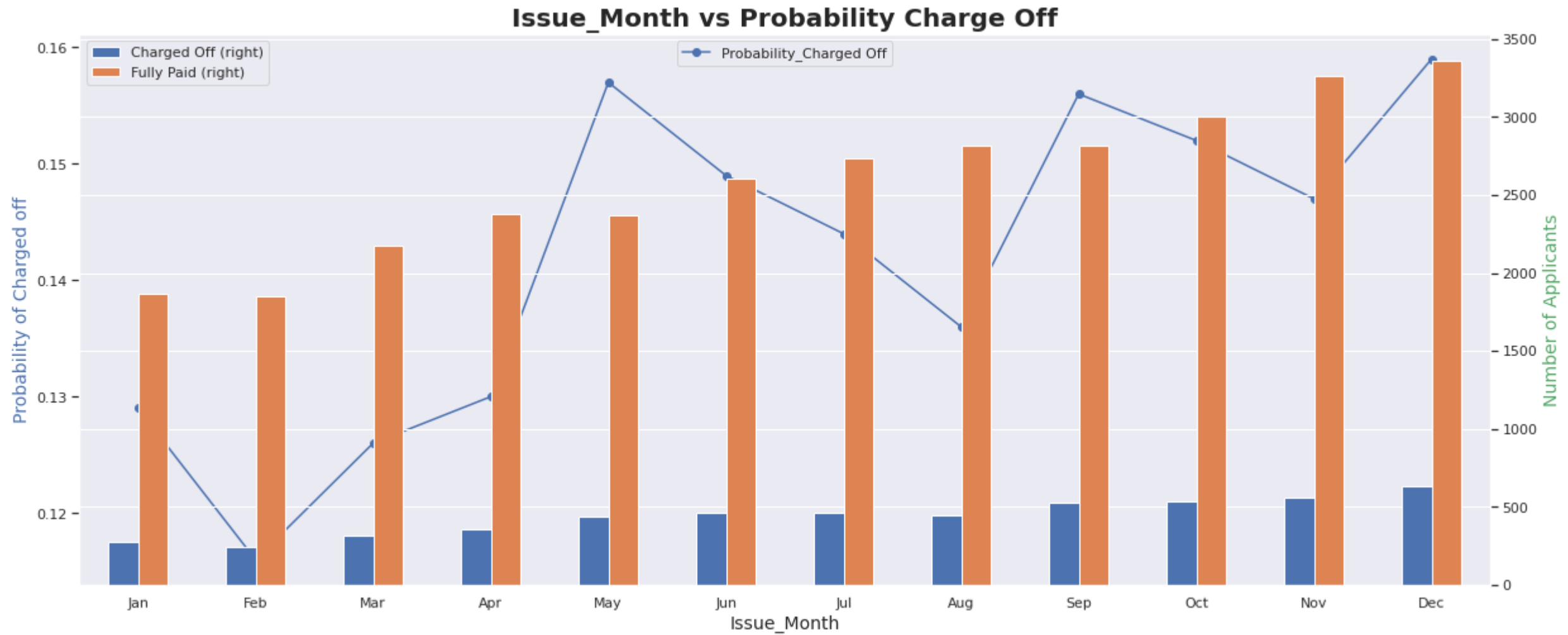
1. Loans issued in the month of Dec, May and Sept are most likely to be charged off

Purpose vs Probability Charge Off

1.Applicants who want to take loan for 'small business' have the highest probability of being defaulters.
2.So bank should take extra caution like take some asset or guarantee while approving the loan for purpose of 'small business'

Addr_State vs Probability Charge Off
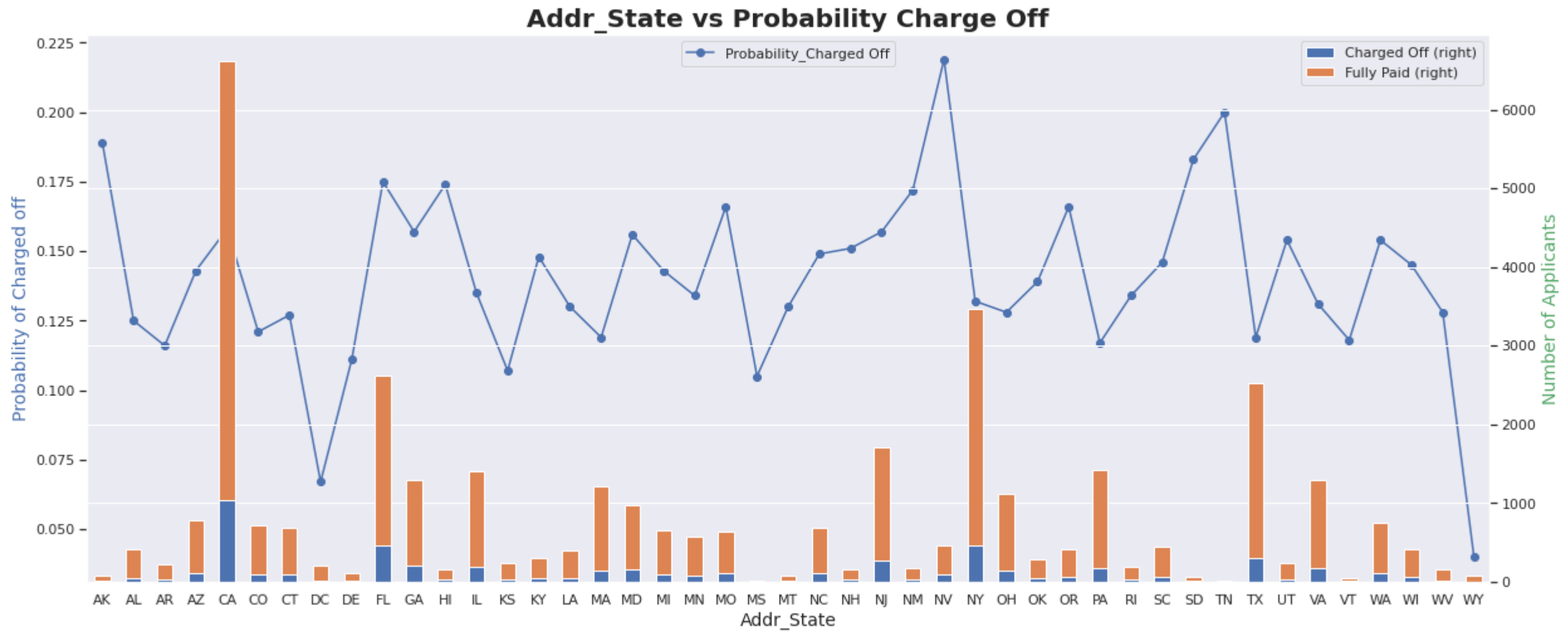
1. Applicants from NV, TN, AK and SD states are most likely to become defaulters

Dti_Range vs Probability Charge Off
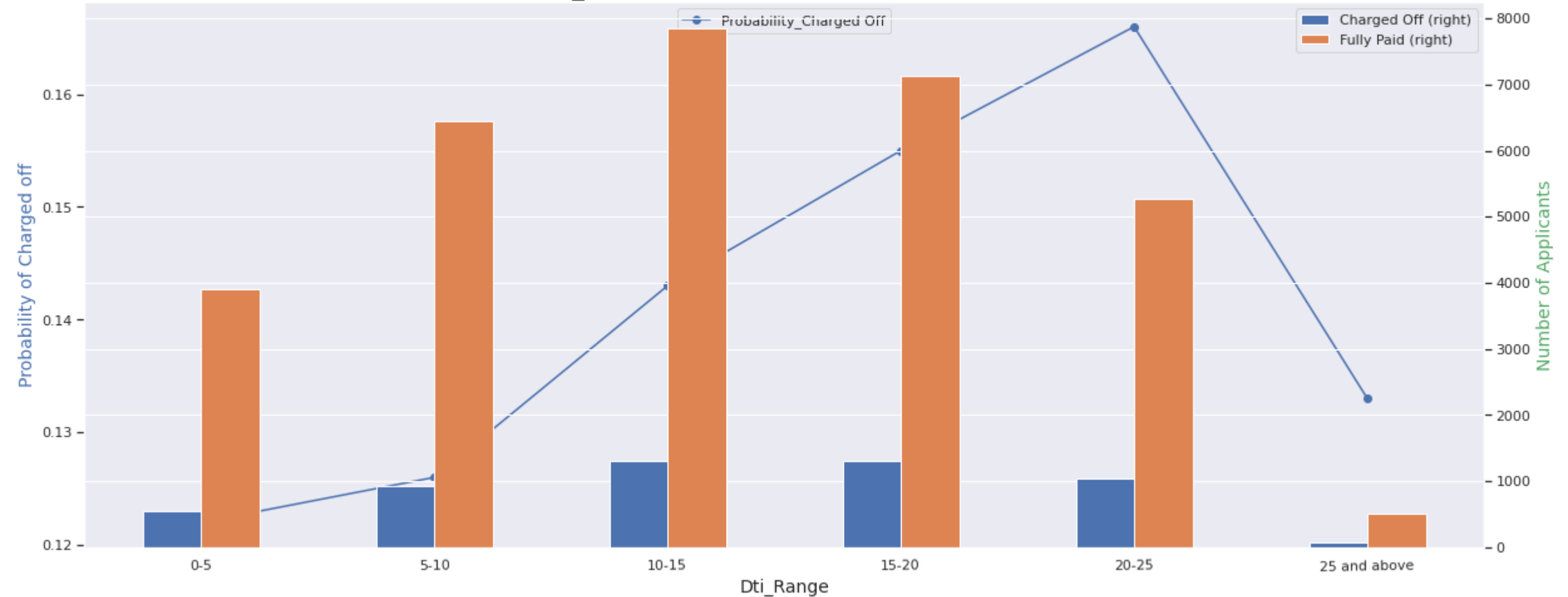
1. The probability of an applicant being a defaulter increases with increase in DTI upto 25 and then decreases
2. Maximum probability of an applicant being defaulter is for one with DTI in range 20-25

RECOMMENDATIONS

The driver variables which can be used to as key indicators to identify defaulters are :-

1. Term
2. Interest Rate
3. Grade/Sub-grade
4. Annual Income
5. Purpose
6. Loan Amount
7. Home Ownership
8. Employment Length
9. Verification Status
10. Location
11. Issue Month

Following parameters increase the probability of loan being charged off :-

1. The loan term is of 60 months
2. The interest rate is 15% or more
3. The grade is of lower category like F or G
4. The subgrade is either of F3, G3 or G5
5. The annual income of the applicant is less than 25k
6. The purpose of loan is small business
7. The loan amount is equal to or greater than 25k
8. The applicant has home ownership as rent or other
9. The employment length of the applicant is < 1 year or 10+ years
10. The income is verified and loan tenure is 60 months
11. The address state of the applicant is either NV, TN or AK.
12. The loan is issued in the month of Dec, May or Sept