Lending Club Case Study

SUBMISSION

Group Members:

➢ Dr Renu L Rajani (Group Facilitator)

➢ Vijaykumar Kamma

Lending Club Case Study

A consumer finance company which specialises in lending various types of loans to urban customers. **The company wants to understand the driving factors (or driver variables) behind loan default**, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

When a person applies for a loan, there are **two types of decisions** that could be taken by the company **Loan Accepted** and **Loan Rejected**. In loan Accepted we have 3 scenarios **Fully paid** (Applicant has fully paid the loan), **Current** (Applicant is in the process of paying the instalments), **Charged-off** (Applicant has not paid the instalments in due time for a long period of time).

When company wants to take a decision on loan application whether to approve or reject loan, there are two types of risks associated with it,

- 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

Business Objective:

Identify driving factors which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. The company can utilise this knowledge for its portfolio and risk assessment.

Goals:

Perform Univariate and Bi-Variate Analysis on the given data, find the driving factors (strong indicators) behind loan default.

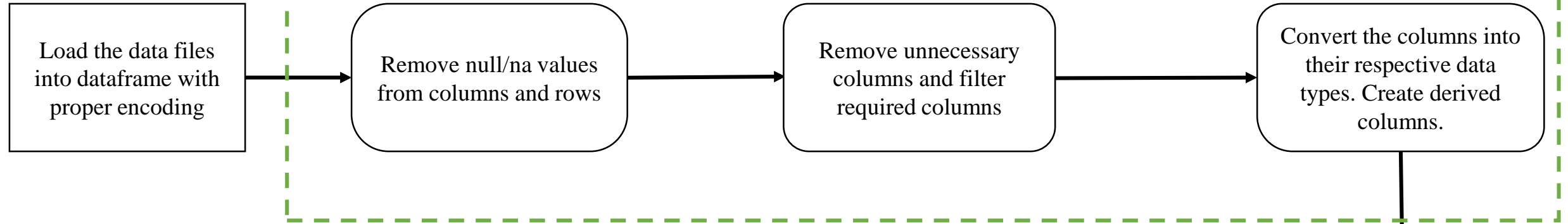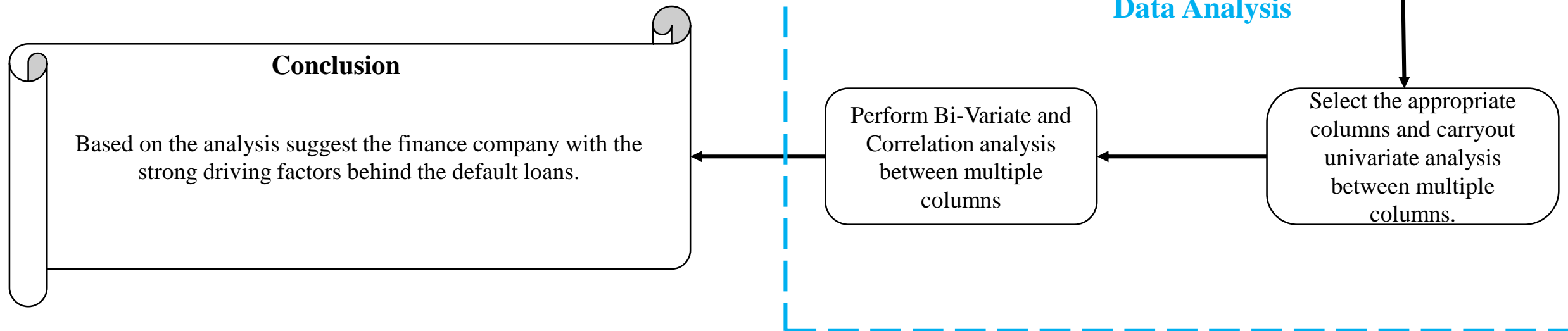# Data:

We have 2 csv files,

Loan Data Set: It contains the complete loan data for all loans issued through the time period 2007 t0 2011.

**Data Dictionary:** which describes the meaning of these variables in Loan Data Set.

Problem Solving Flow Chart

Data Cleaning

Load the data files into dataframe with proper encoding

→

Remove null/na values from columns and rows

→

Remove unnecessary columns and filter required columns

→

Convert the columns into their respective data types. Create derived columns.

Data Analysis

Select the appropriate columns and carryout univariate analysis between multiple columns.

←

Perform Bi-Variate and Correlation analysis between multiple columns

←

Conclusion

Based on the analysis suggest the finance company with the strong driving factors behind the default loans.

Data Cleaning

- Removed the columns having null/na values more then 30%. Rows and Columns in Loan Data frame after removing (39717, 53).

- Dropped 9 columns having only one value. Converted the columns into their respective data types after removing special characters and strings.

- Maximum annual income is way above the 75th percentile. We decide to limit the records with annual income values within 95th percentile. Removed the extreme values in the annual income column.

- In home ownership column Rent, Own, Mortgage are valid ownership types, it is difficult to understand and analyze on "Other", "None", removed the remove the rows with Other/None values.

- We have added 6 derived fields in loan dataframe -loan_income_ratio, issue_month, issue_year, loan_amnt_range, annual_inc_range, int_rate_range. Total number of columns is 46.

- We have dropped the loans having status as Current because, these candidates are not labelled as 'defaulted'.

Univariate Segmented Univariate Analysis

Analysis by Annual Income:
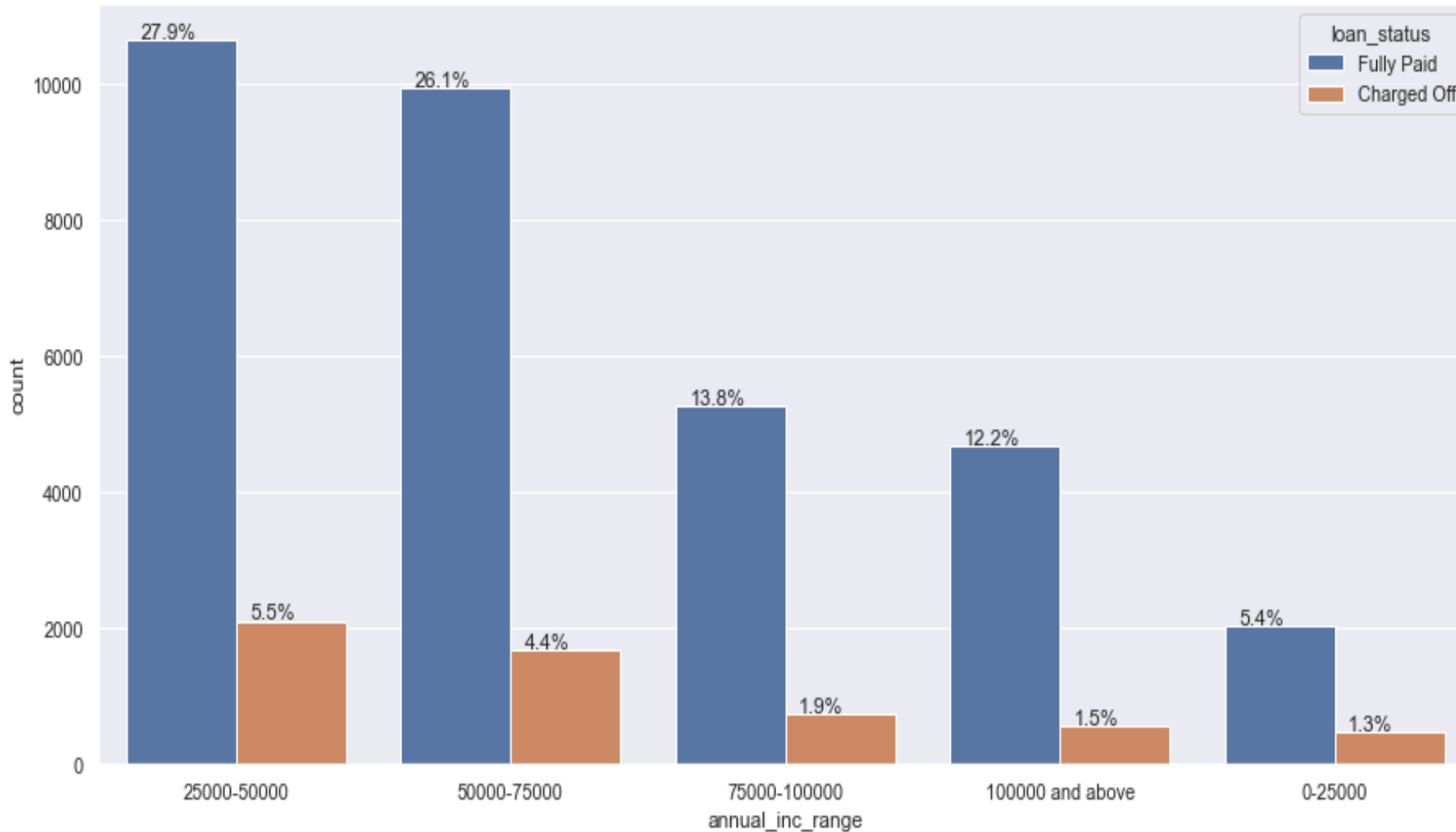
- The largest number of loans (12761 out of total 38131 loans, 33.3%) are extended to people with USD25000-50000 annual income bracket.

- The largest % defaulters 5.5% also belong to this bracket

Analysis by Interest Rate:
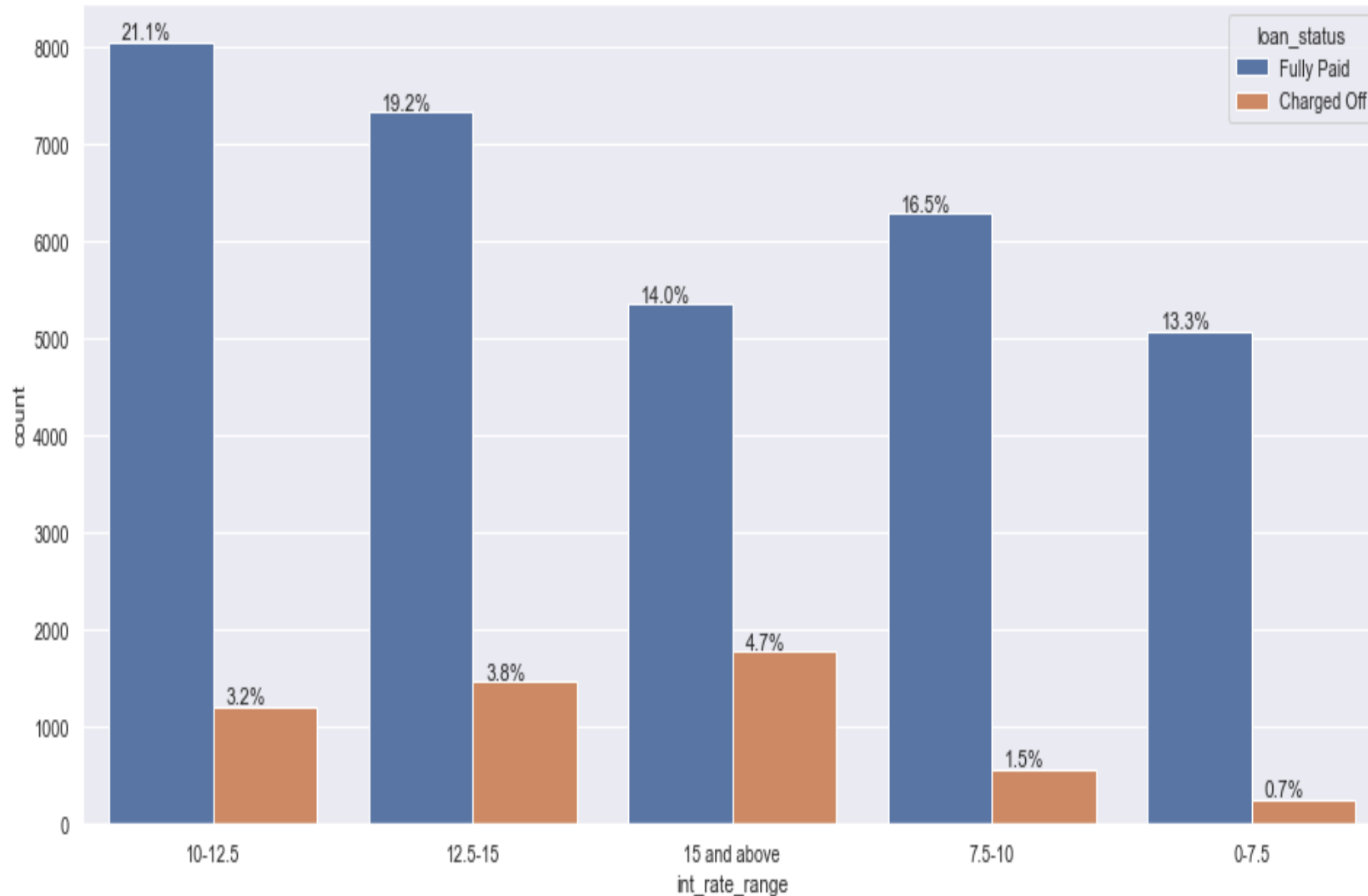
- The largest number of loans (9265 out of total 38131 loans, 24.3%) are in the interest rate bracket of 10-12.5 percent.

- The percentage defaulters are the highest in interest rate range of 15% and above.

Analysis by Loan Amount:
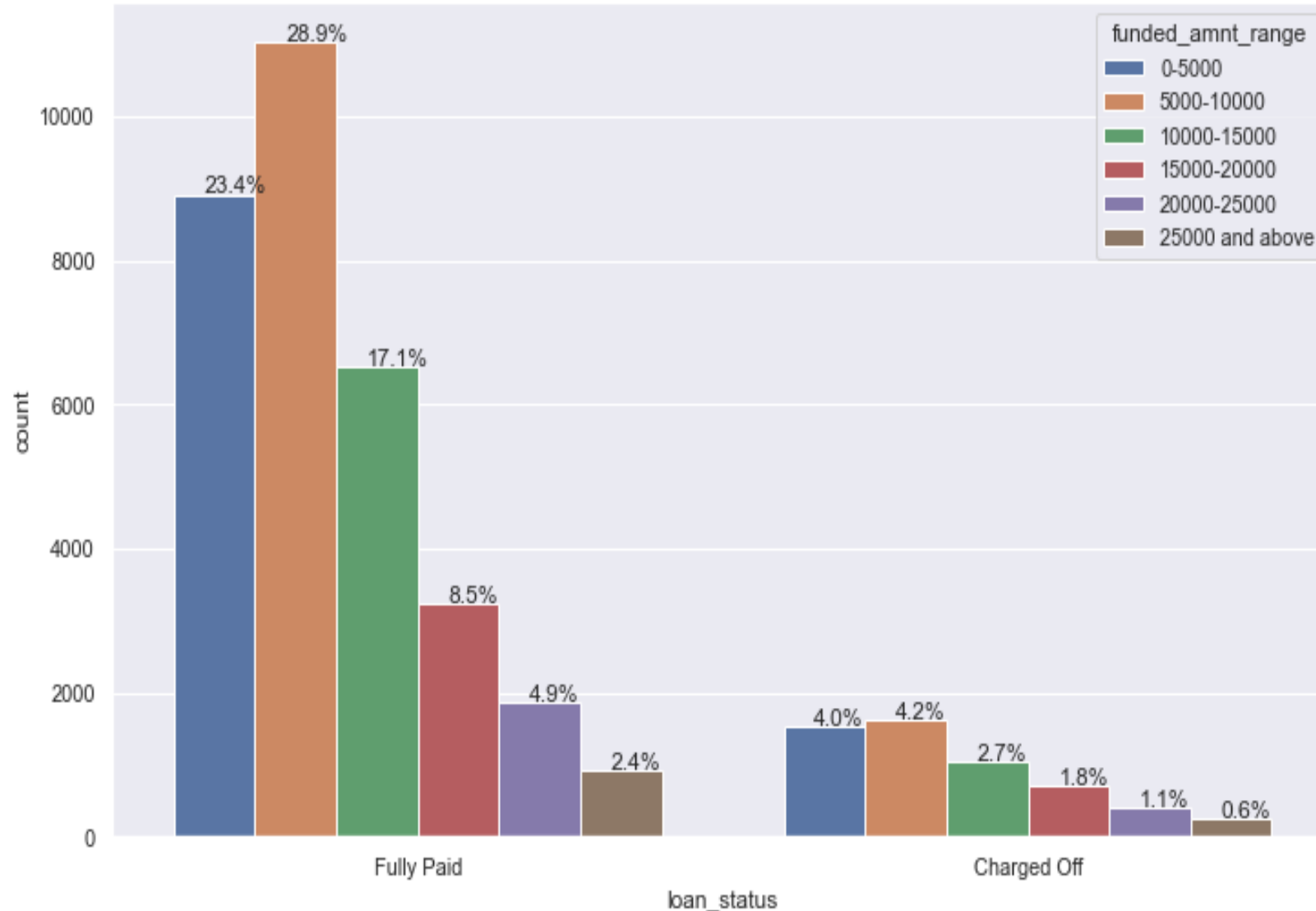
- Largest category of loans are between USD5000-10000 Loan Amount range, and the highest incidence of default is also among this group (4.2% of overall loans)
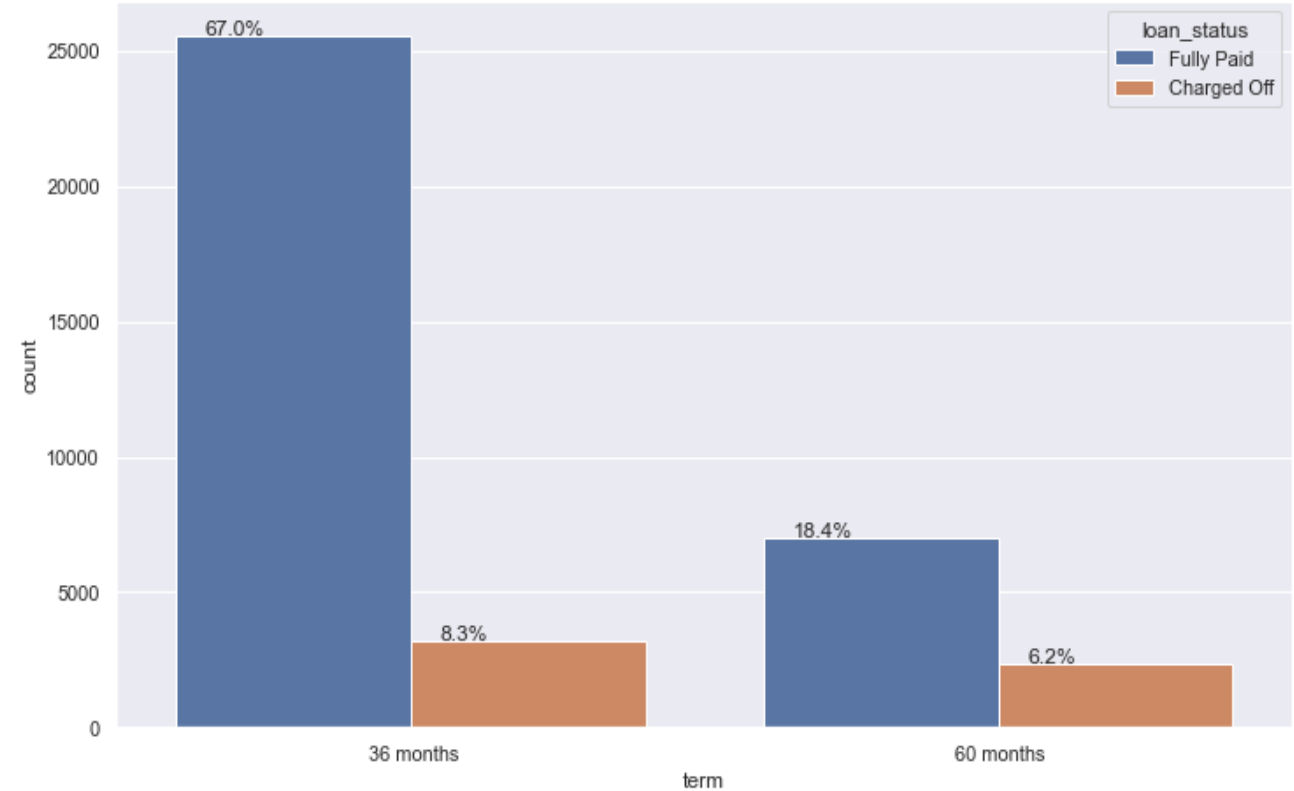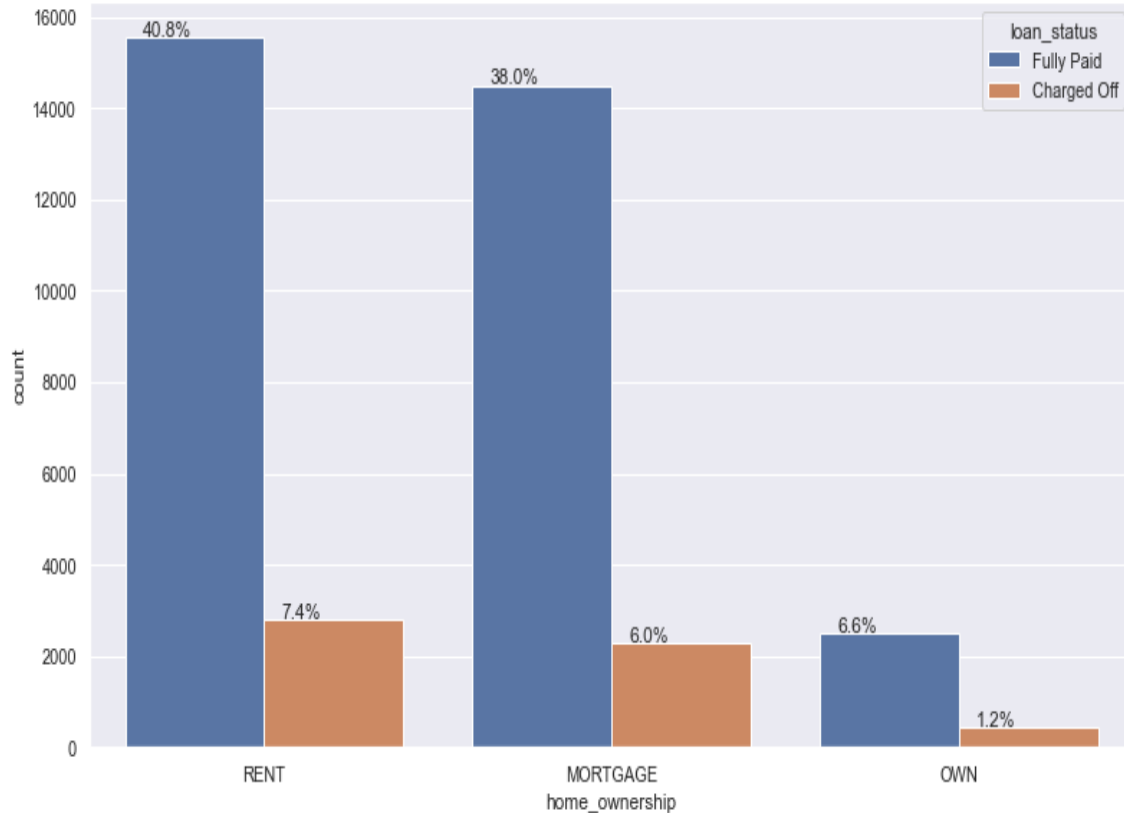
Analysis by Loan Purpose:
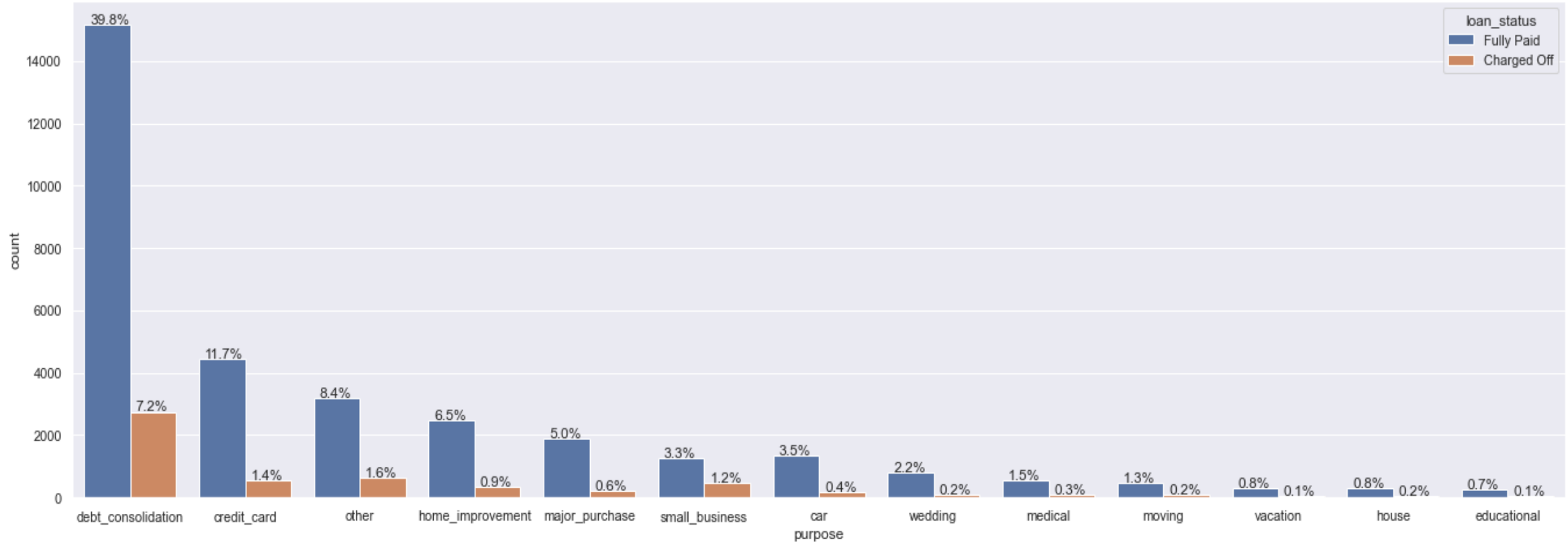
- Debt Consolidation emerges as top most purpose for taking loan and appears are topmost category for Charged Off loans.

Analysis by Home Ownership and Term:

- The highest instances of loans issued (18798 out of 39265, 47.9%) are to people with home ownership as "Rent". The default (7.2% of all loans) is also the highest in this group.
- 73.2% loans are extended for 36 months. However the charged-off proportion is higher for 60 month tenure.

Correlation Analysis
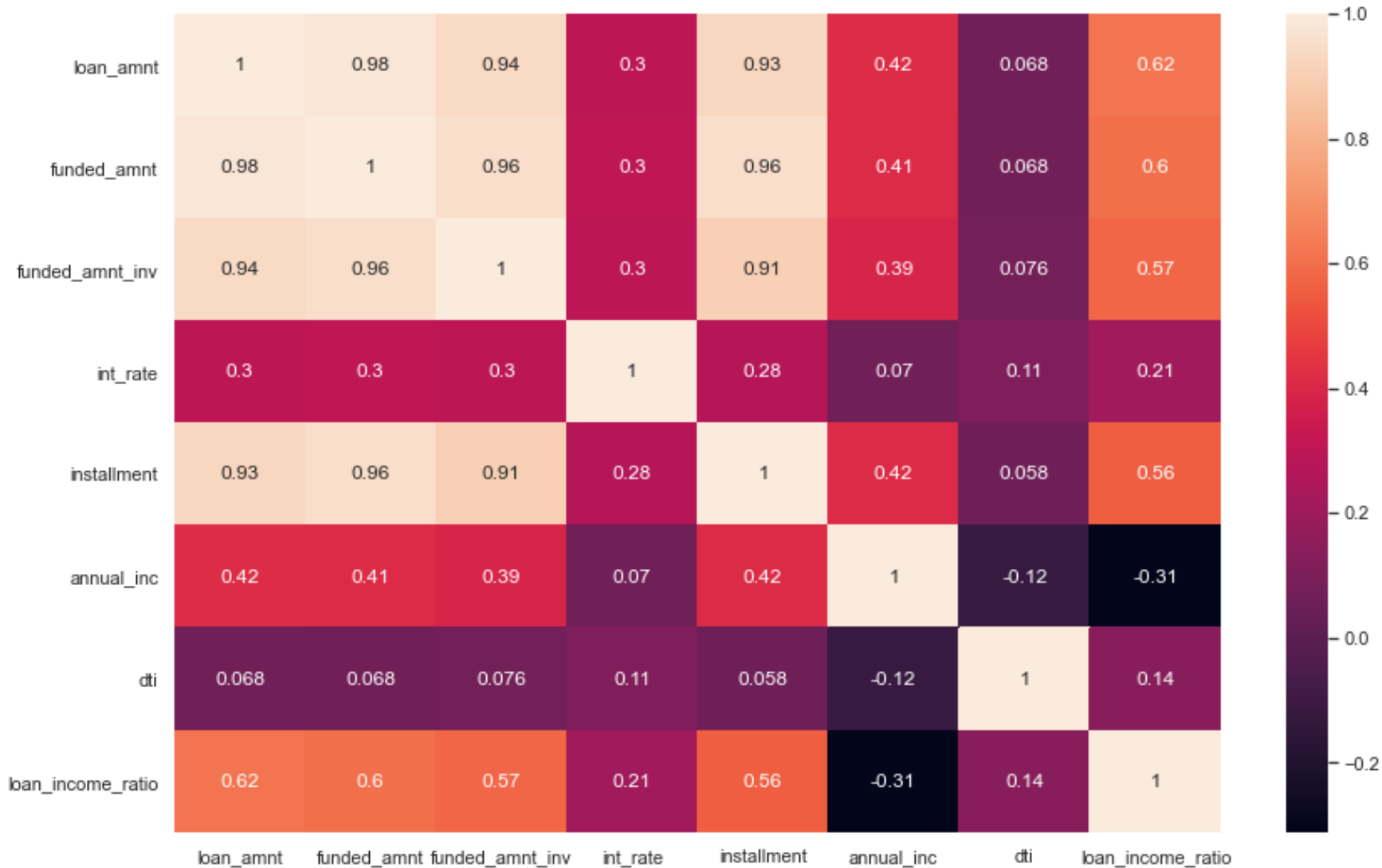
- Loan_Amnt, Funded_amnt, funded_amnt_inv, and installment amount are highly correlated.

- We keep our analysis to Loan amount and not for other variables therefore.

- annual_income and loan_income_ratio are negatively correlated.

Bi-Variate Analysis

Analysis by Emploment Length VS Loan Amount:

Employment Length vs Loan Amount by Loan Purpose
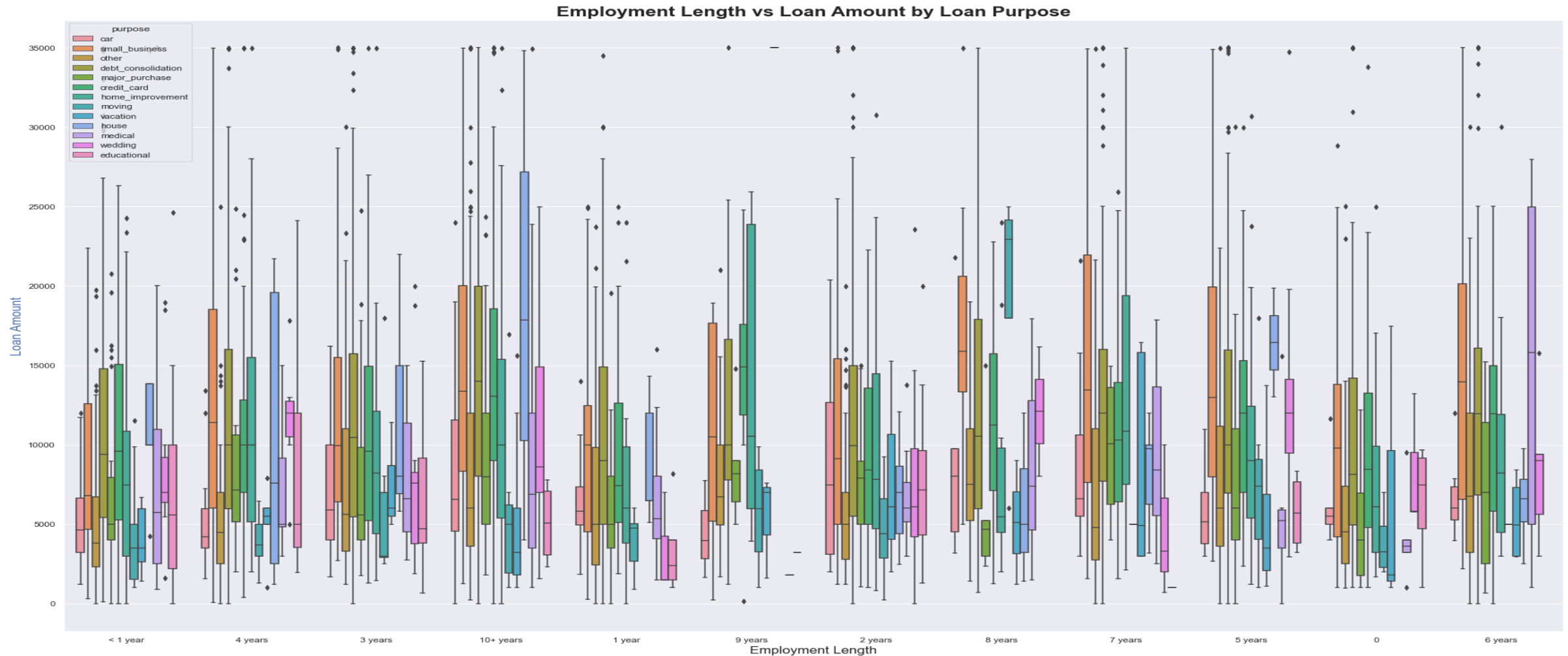
- 10+ years Employment length category have the highest variance of loan amount extended against the mean loan amount in the box plot.

Analysis by Address State:


Addr_State vs Prob Charge Off
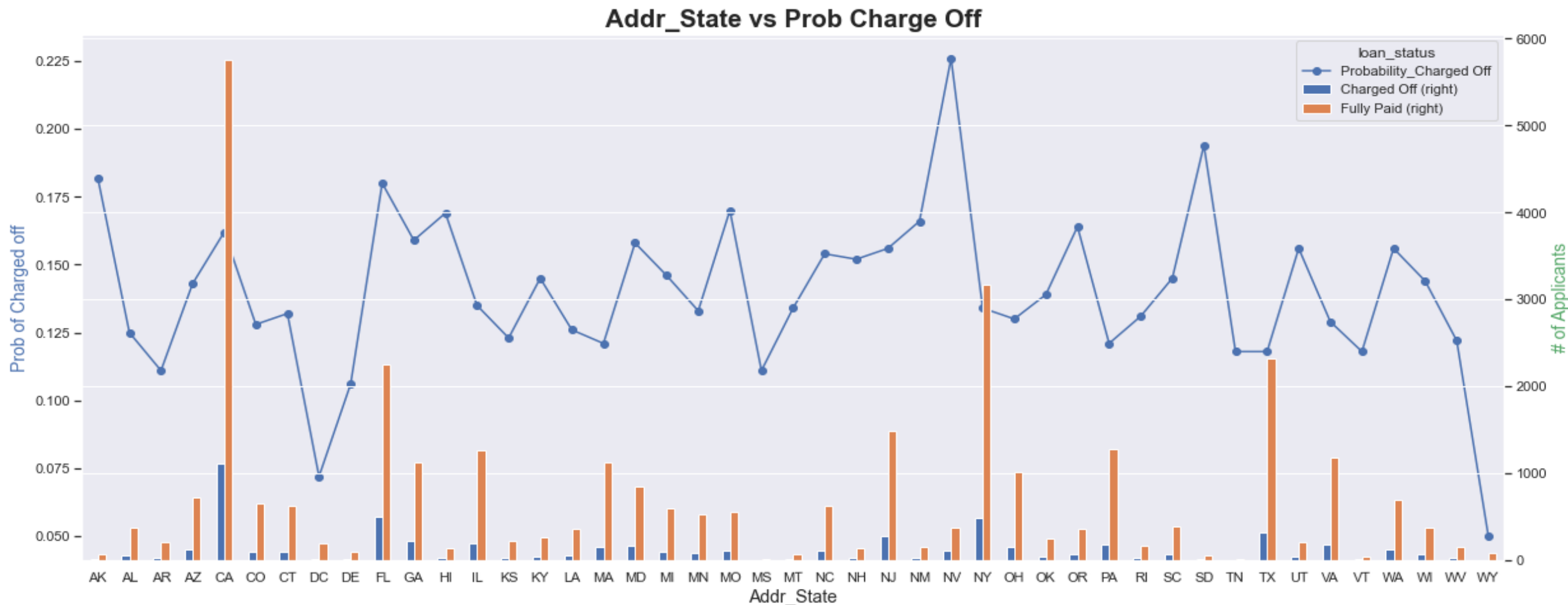
- Probability of Charged-off is the highest in state of Nevada (NV) and amount charged off is the highest in the state of California.

Analysis by Purpose:


Purpose vs Prob Charge Off
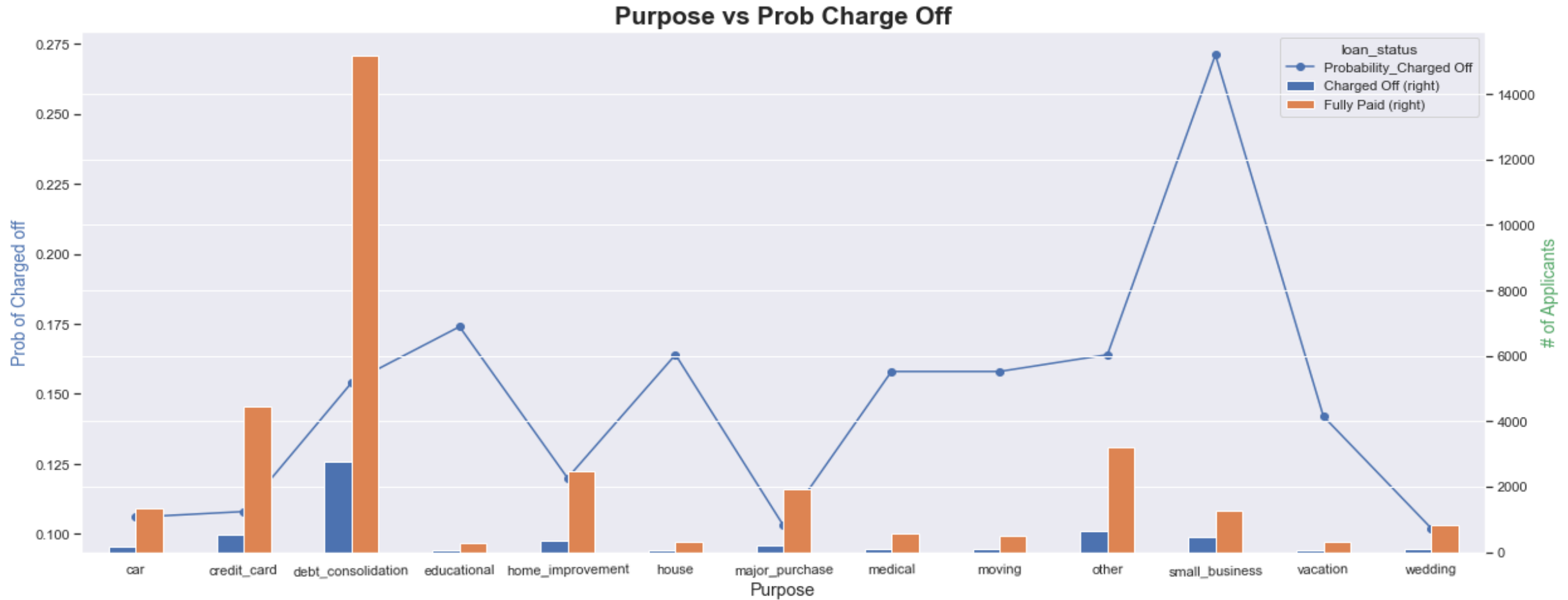
- The highest charged-off have happened for debt consolidation purpose loans, and highest probability of the charge-off is for loans taken for small business purpose.

Analysis by Grade:

Grade vs Prob Charge Off
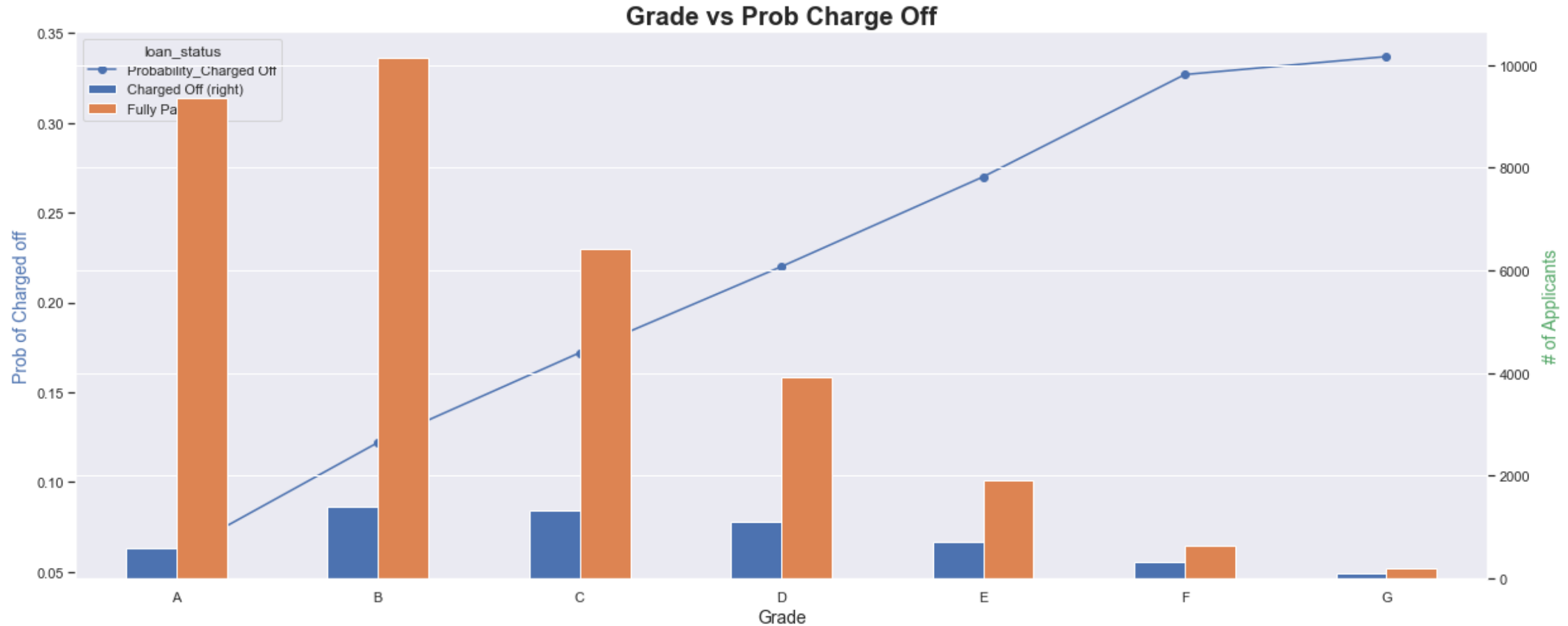
- Probability of charged-off is the highest for Grade G loans (31.8%). When analysed by sub-grade, the highest probability of charge-of is in the loans with grade F5.
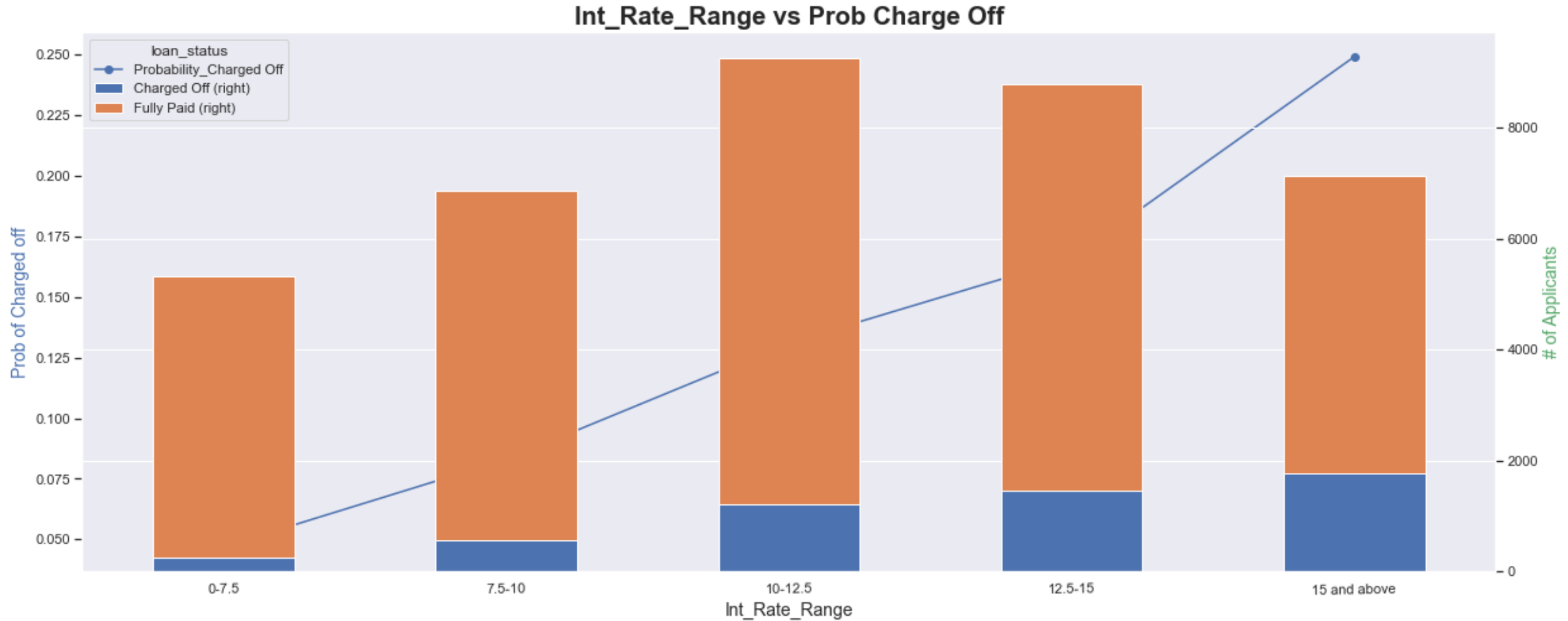
Analysis by Annual Income:

Annual_Inc_Range vs Prob Charge Off
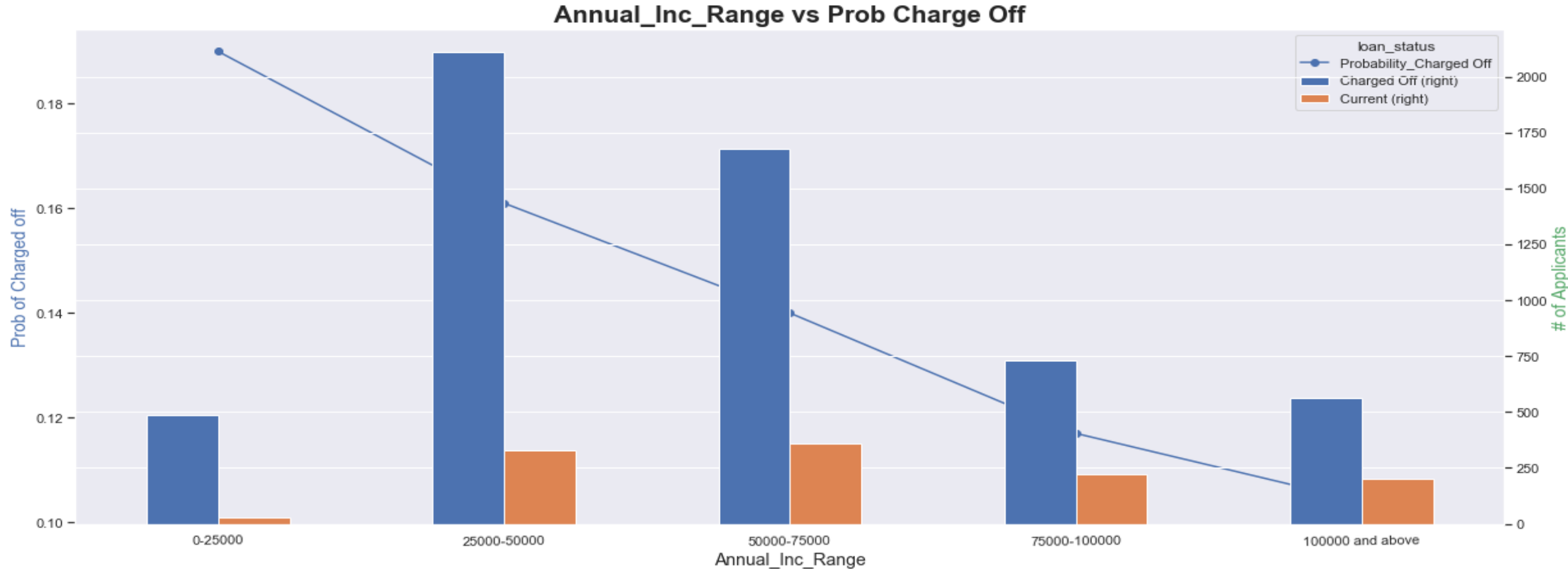
- The highest amount charged-off has been for the loans in the Annual Income Range of USD25000-50000. The highest probability of charge-off is for the loans in the category of Annual Income Range USD 0-25000 (19%).
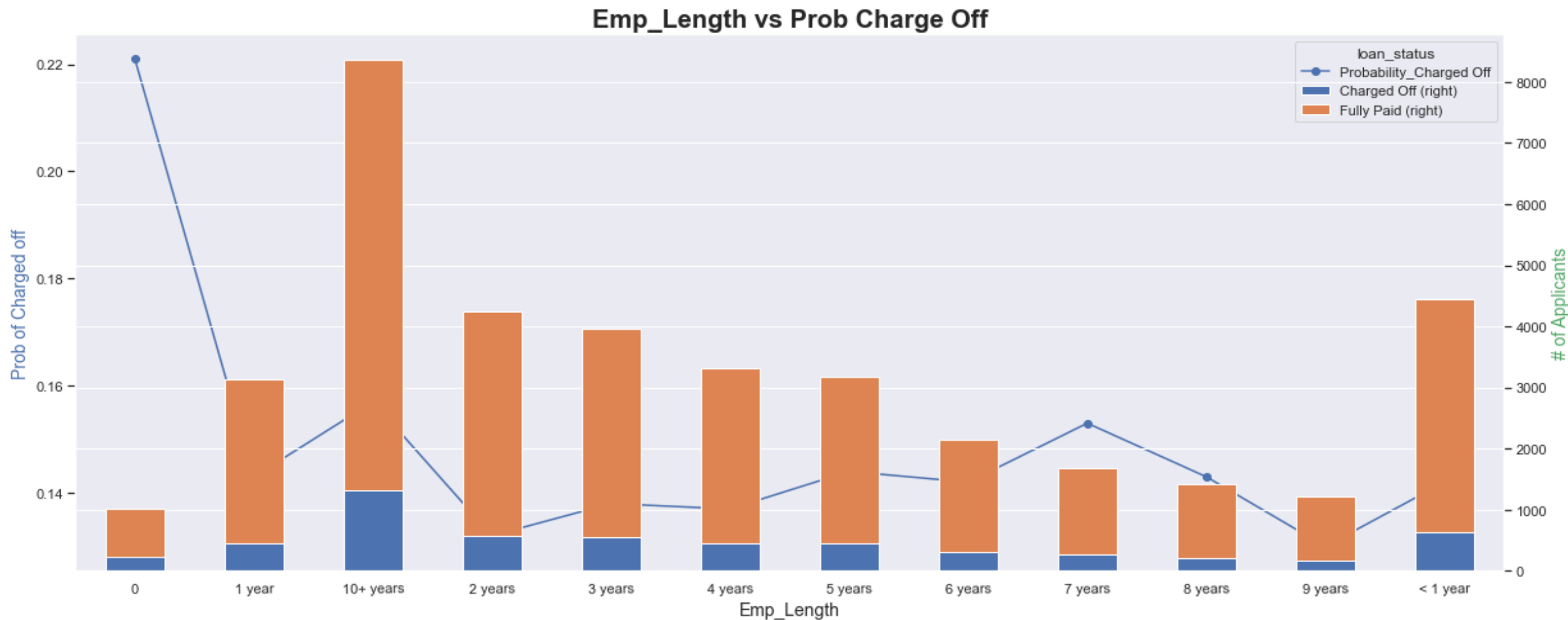
Analysis by Interest Rate:

Int_Rate_Range vs Prob Charge Off

- Amount Charged-off as well as probability of charged-off is the highest for the Interest Rate range of 15%

Analysis by Employment Length:

Emp_Length vs Prob Charge Off

- Probability of charged-off (22.1%) is the highest among low employment length.

Conclusion

Below are the driving factors behind default loans, company may follow these factors to avoid credit loss.

- Should minimise the giving loans for annual income range USD25000-50000, there is a 19% chance it will be default.

- Should avoid the approving the loans for low employment length, there is 21.2% chances it will be default loan.

- If interest rate increase the chances of default loans will also increase, minimise the approving the loans for high interest rate.

- The highest grades have the high chance of default, minimise the approving loans for highest grades.

- Reduce the approving the loans for small business purpose, there is 26% chance of defaulting the loan.