LENDING CLUB CASE STUDY

Jay Jannan

OBJECTIVE:

The objective of this case study is to conduct a comprehensive EDA on Loan Lending Company's real data to find the driving factors which corresponds to the customer's loan defaults.

Steps undertaken to achieve the objective:

- Data Understanding: Refer the documentation (Data_Dictionary.xlsx) that explains the structure, types, and meaning of each data variables.
- Data Cleaning: Handling missing values, Checking and Removing duplicates(If present), Fixing or removing incorrect data points, identify and managing outliers.
- Data Transformation: Creating new fields from existing variables for better research, convert numerical variable into categorical variables like grouping the segments of distribution into labels.
- Data Visualization: Univariate, Bivariate and Multivariate Analysis.
- Data Interpretation: Draw insights from the visualizations and identify patterns in the data.
- Reporting: Preparing a presentation to communicate the findings.

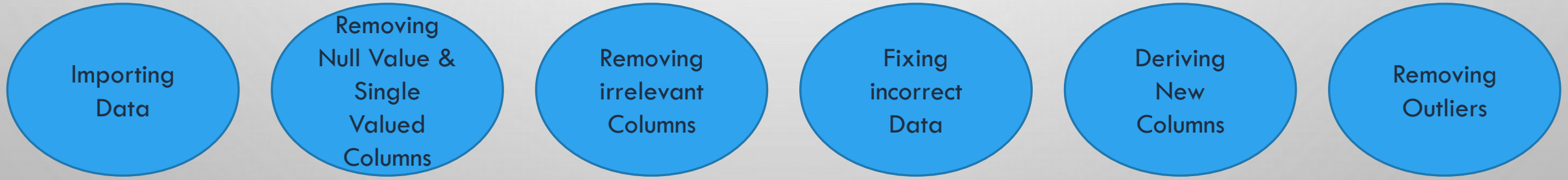
Business Context:

The business context is to take informed decisions on whether to approve or reject the loan applications based on the driving factors(fields) from the given dataset.

Dataset Info:

The given dataset contains 3.9 lakh+ disbursed loan applications and their status such as whether the applicant had fully paid or current(ongoing) or Charged off(Default).

Data Cleaning and Transformation:



Univariate:

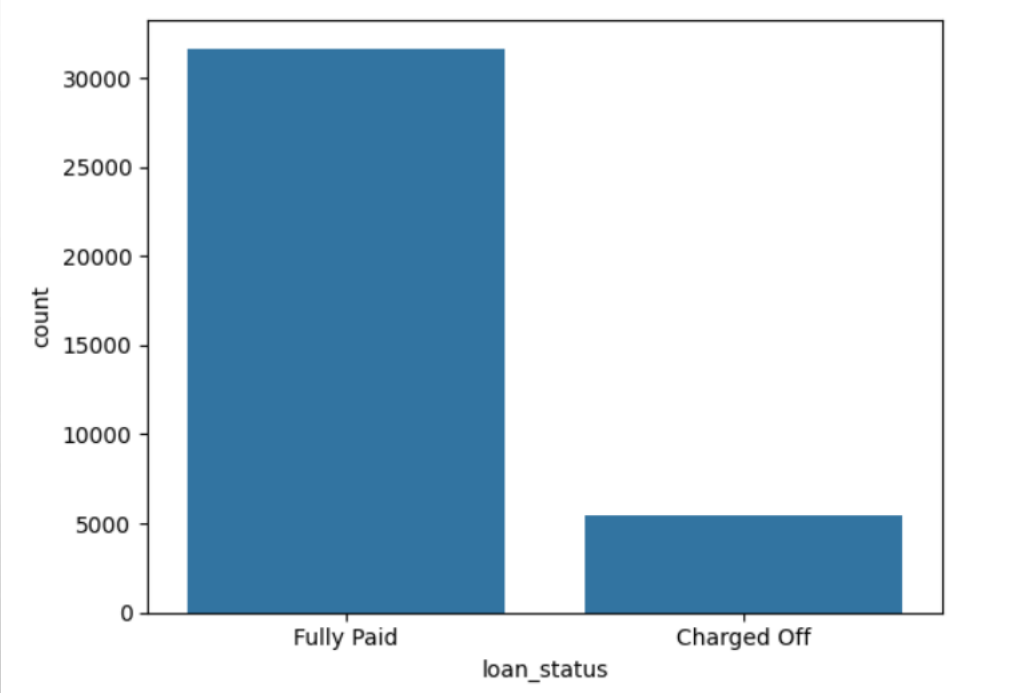
It is a statistical method to analyze and summarize data sets based one variable. Univariate analysis can be performed on both categorical and quantitative variables. It deals with the analysis of single variable (factor) rather than multiple factors and to understand the distibutions.

Categorical Variables: It can be of two type, ordered and unordered.
Ordered -> grade, (sub grade) sub_grade, term, (Employee experience) emp_length, issue_y (issue date)
Unordered -> home_ownership, loan_status, purpose

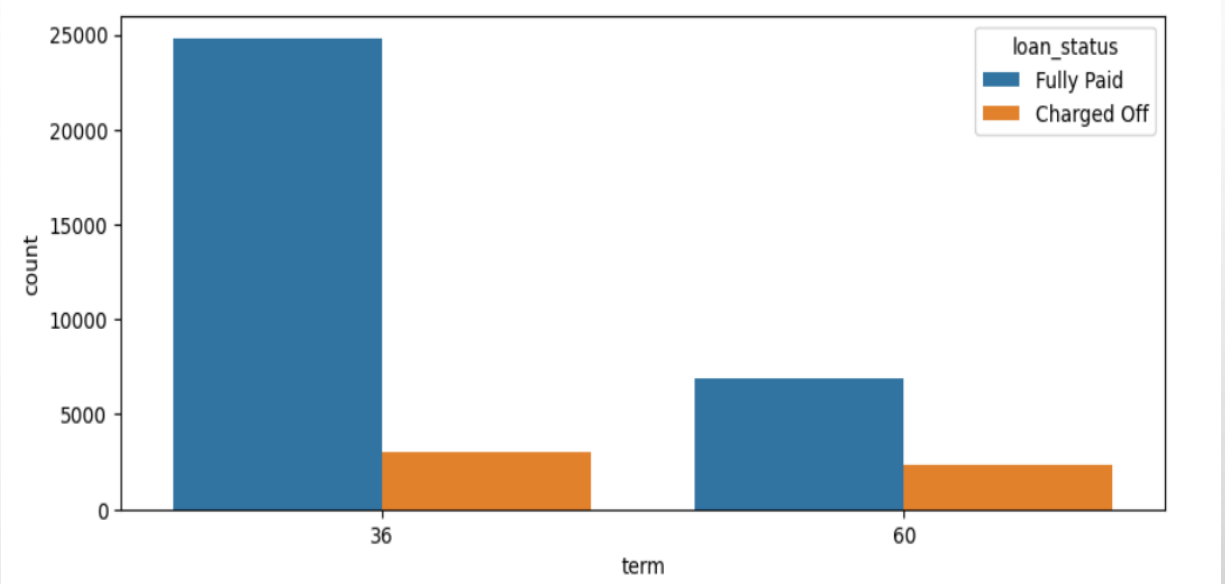
Quantitative variables:

- 'int_rate_groups' (Interest rate groups)
- 'loan_amount_groups' (Loan amount groups)
- 'funded_amount_inv_groups' (Founded amount investor groups)
- 'funded_amnt_groups' (Funded amount groups)
- 'annual_inc_groups' (Annual income groups)
- 'dti_groups' (Debt to income ratio groups)

Loan Status and Term:

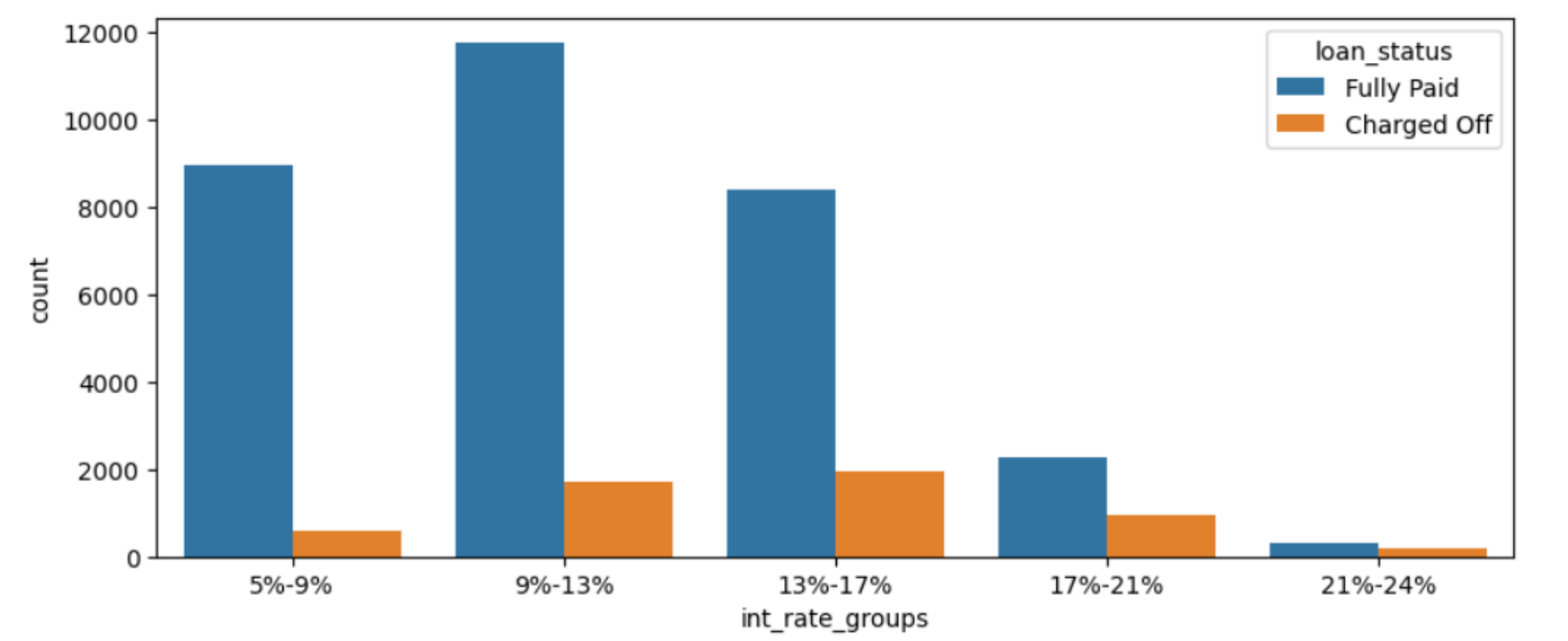


Number of Charged off loan is very much smaller compared to Fully paid loans, Charged off loan is only 17% compared Fully Paid loan count.



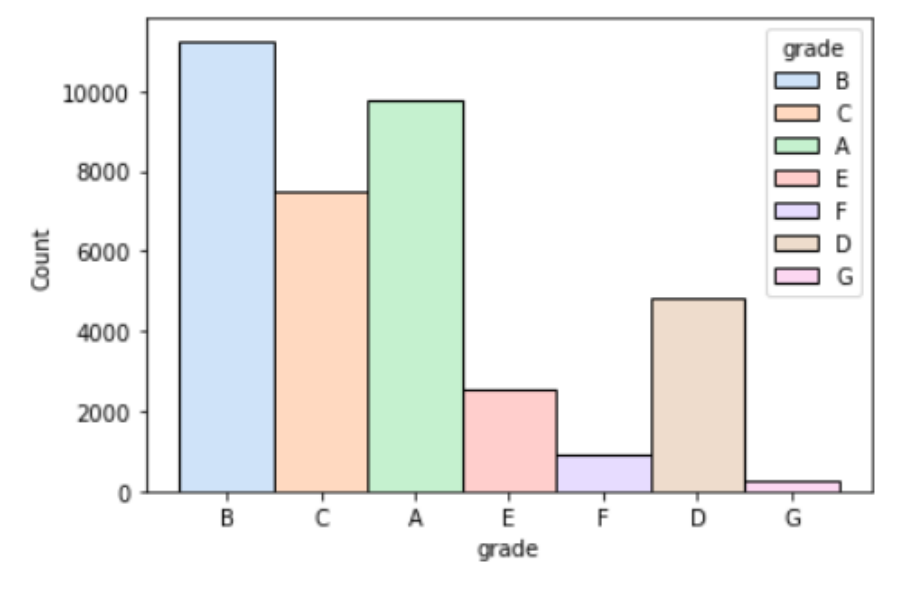
Most loan applications are for 36-month installments, but the number of charged-off loans is slightly higher for 36-month and has higher chance of default.

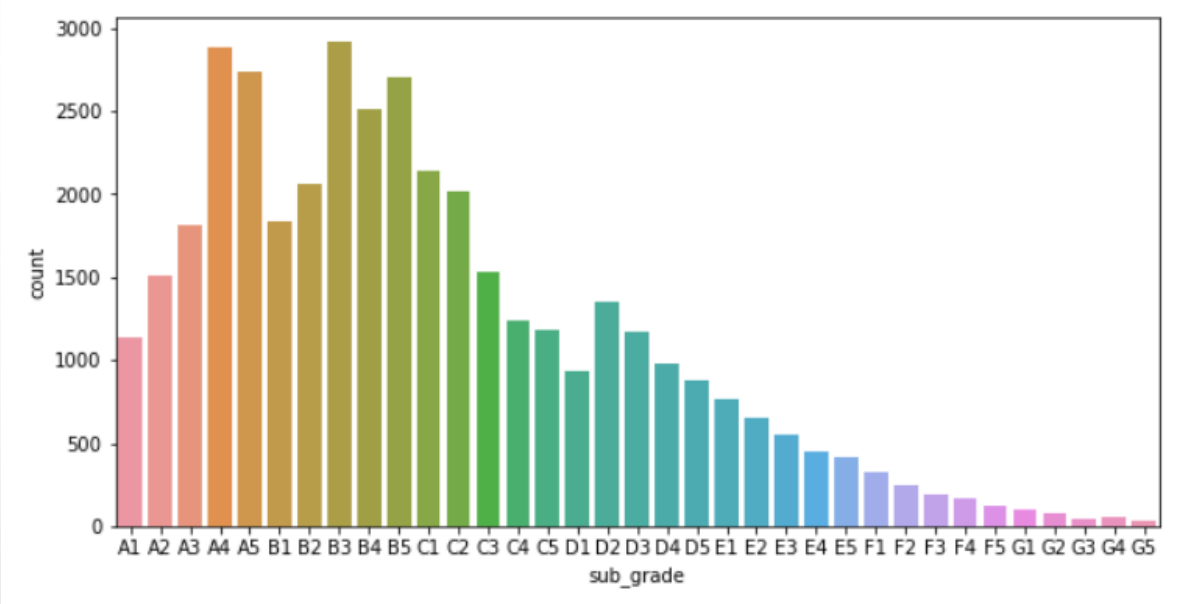
Loan Interest Rate:



Number of Loan applications gradually decreasing with higher interest rates and in Charged off loan applications segment the loan applications are gradually increasing with higher interest rate which shows high chance of loan default at medium to higher interest rates.

Loan Grade:

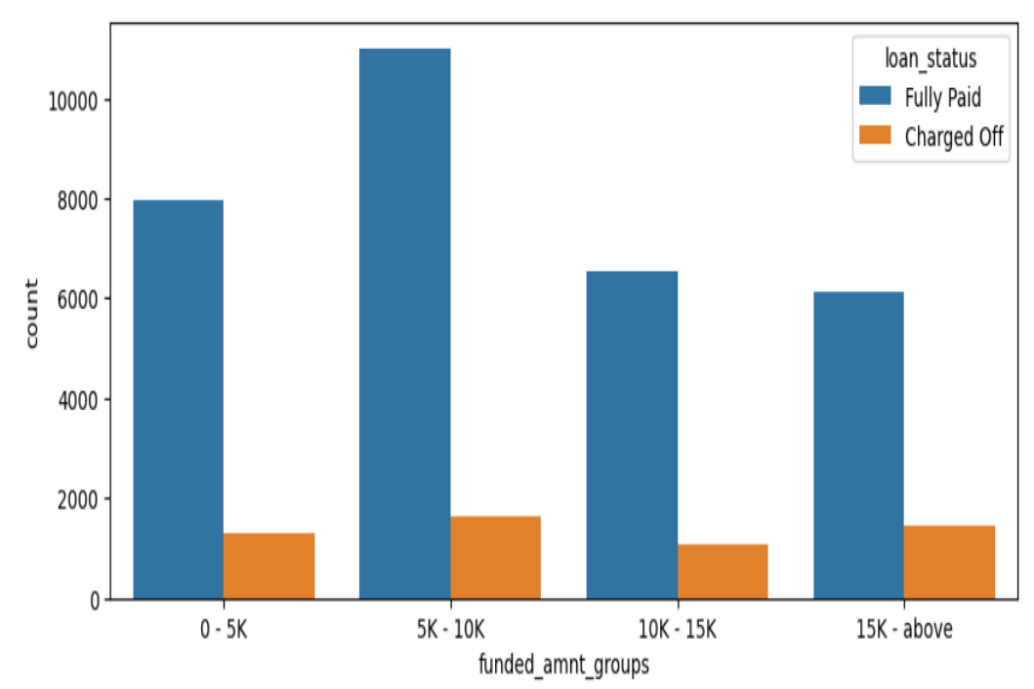




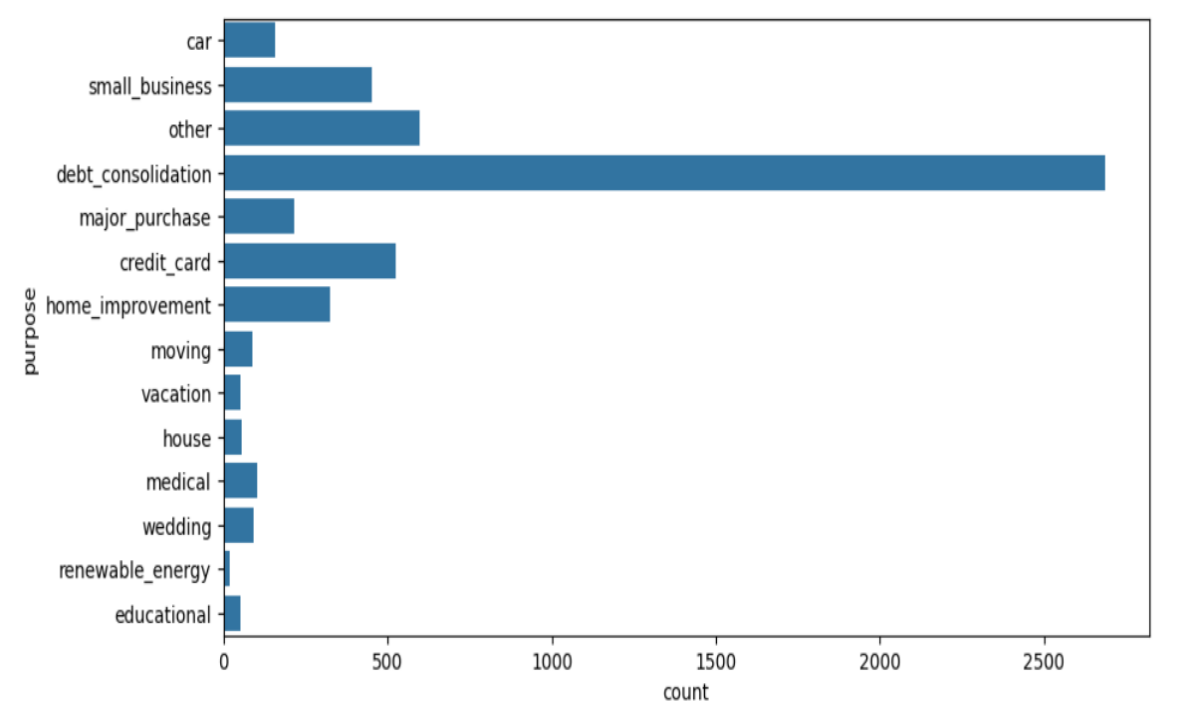
Large number of loans applications are with higher grade A,B,C thus we can assume high grade has less chance of defaulting.

This plot clearly explains there is hike in sub grades like A4, A5, B3 and B5.

Loan Amount & Purpose:

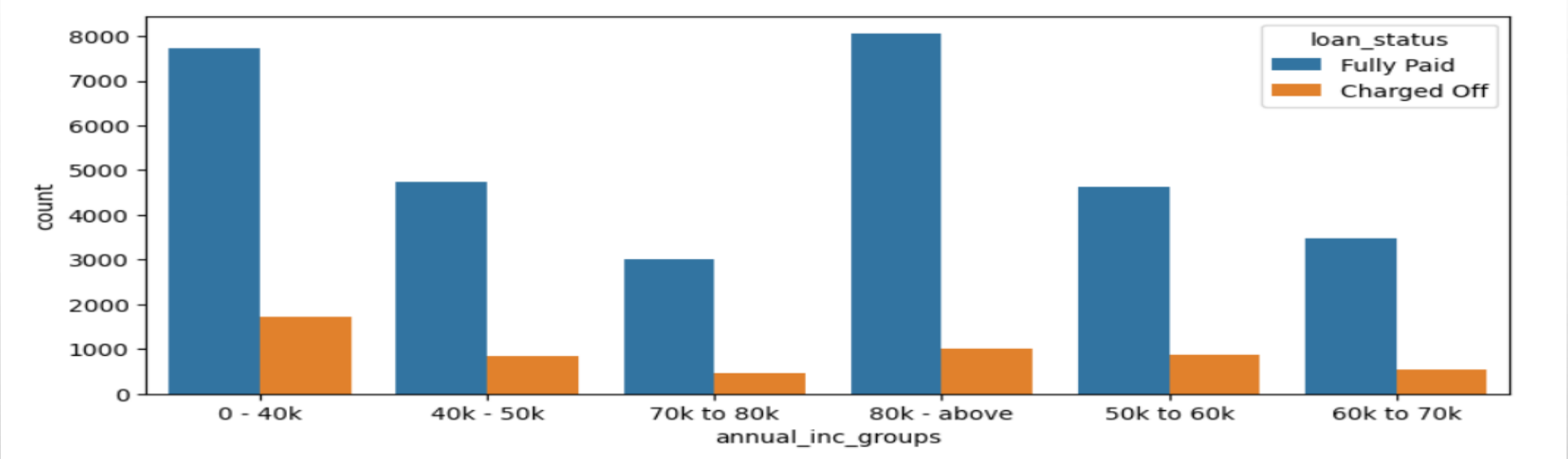


Large number loan defaults are from loan applications with 5k – 10k loan amount



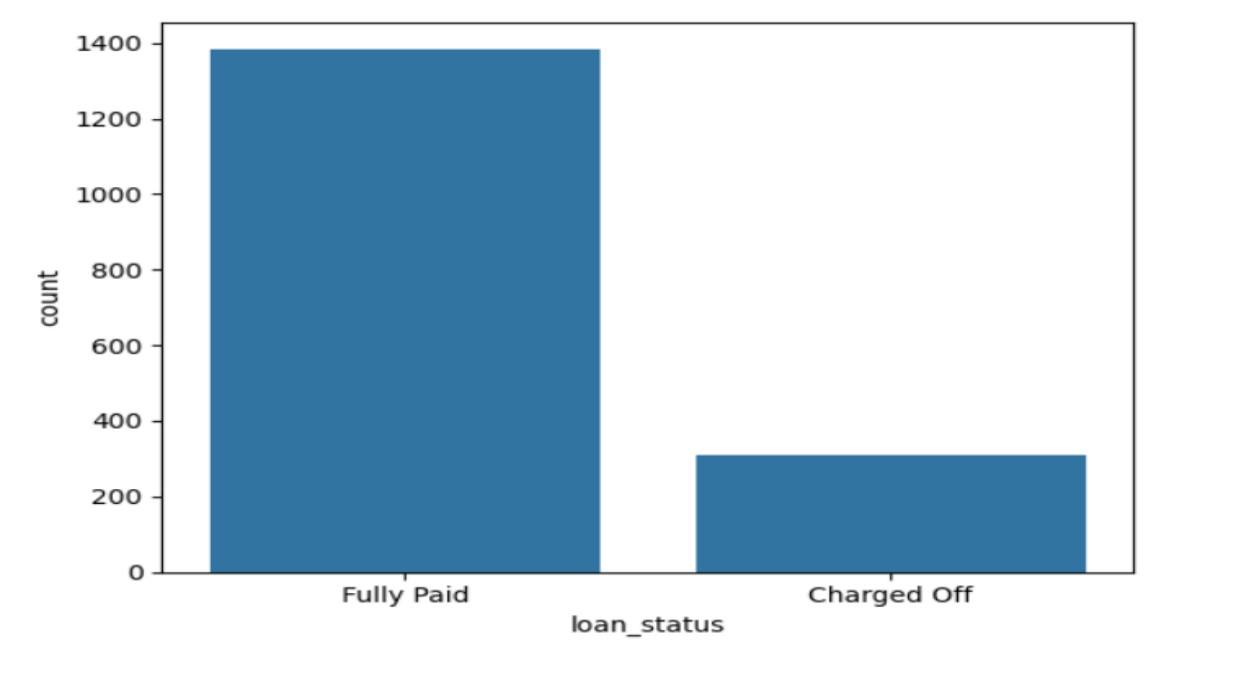
This plot shows the count of default loan application at purpose category level. Which shows debt consolidation has higher number of defaults.

Annual Income:



Large number of default loan applications are from applicants with income range less than 40K.

Loan Scaling:



Loans which were scaled are largely fully paid. Scaling down the Loan amount shows a low number of defaults. So whatever considerations made in planning on scaling have done well in reducing defaults.

Loan scaling is adjusting the principal amount of a loan based on certain criteria or conditions. This adjustment involves majorly decreasing or increasing the loan amount to better match with the borrower's financial situation.

Here Loan Scaling(Yes/No) checks whether the loan amount(Borrower requested) and funded amount(loan disbursed) are varying.

Note: All the loans which were scaled are only decreased. There are no cases of scaling upwards.

Bivariate:

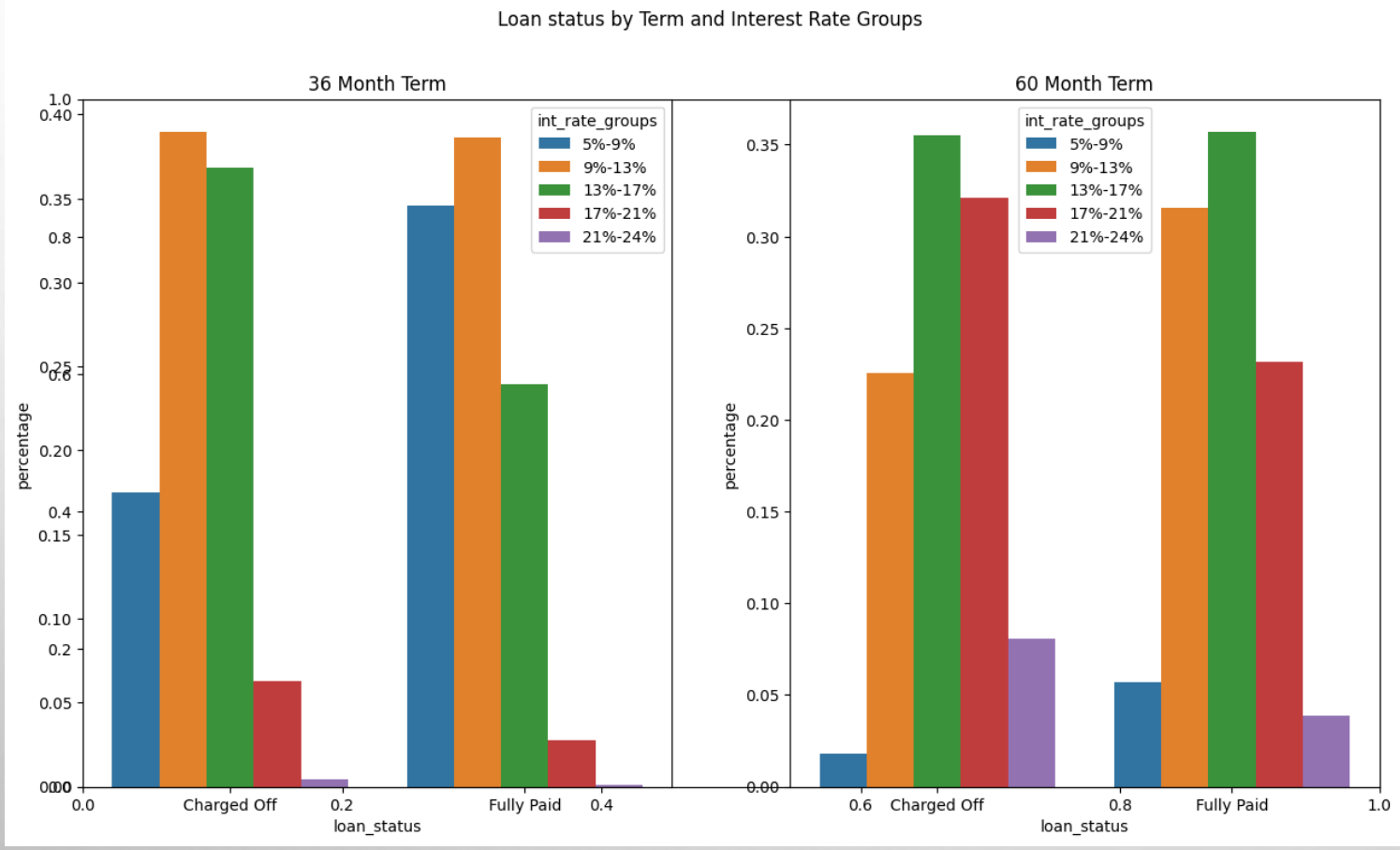
It is a statistical method to analyze and summarize data sets based on two variable, to determine the relationship and identify patterns between them. Bivariate analysis can be performed on both categorical and quantitative variables.

Categorical Variables: It can be of two type, ordered and unordered.
Ordered -> grade, (sub grade) sub_grade, term, (Employee experience) emp_length, (issue date)issue_y
Unordered -> home_ownership, loan_status, purpose

Quantitative variables:

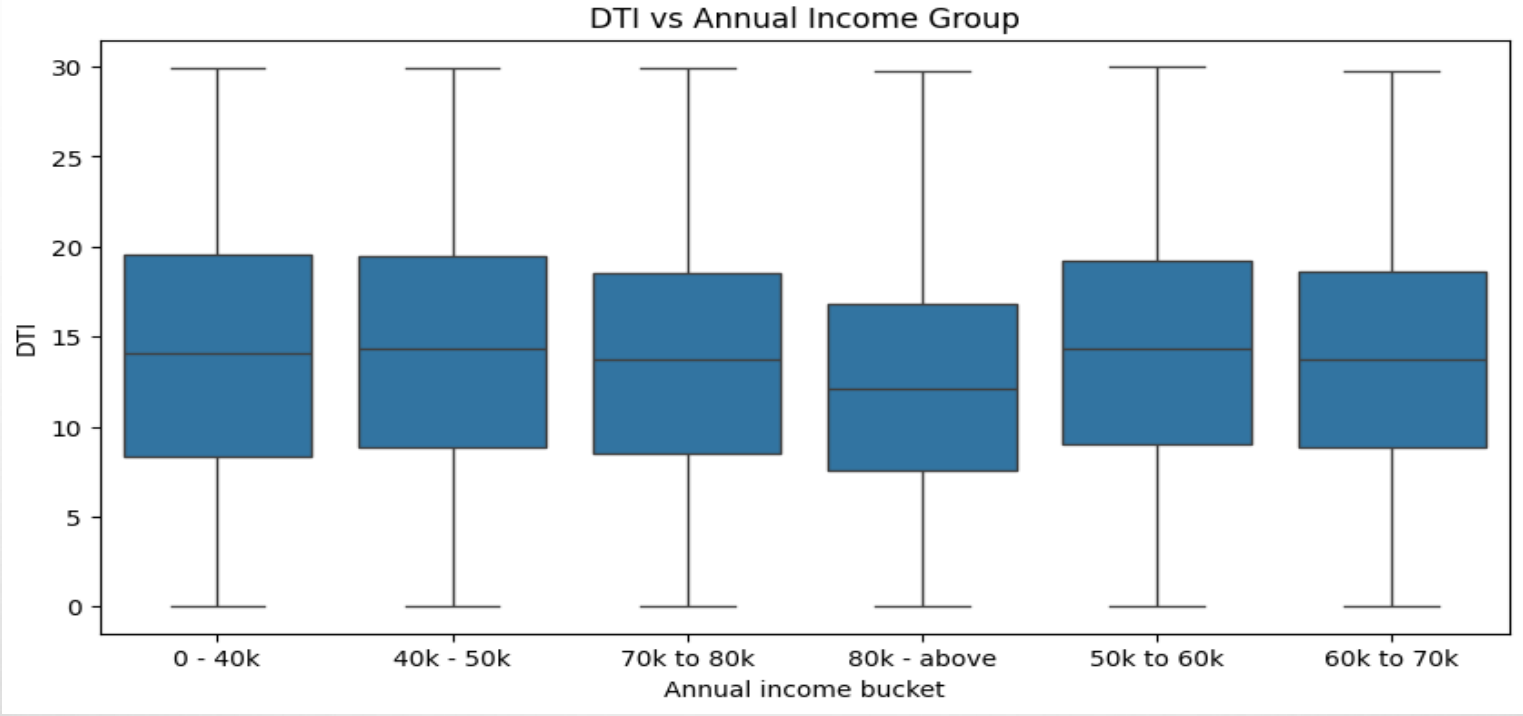
- 'int_rate_groups' (Interest rate groups)
- 'loan_amount_groups' (Loan amount groups)
- 'funded_amount_inv_groups' (Founded amount investor groups)
- 'funded_amnt_groups' (Funded amount groups)
- 'annual_inc_groups' (Annual income groups)
- 'dti_groups' (Debt to income ratio groups)

Compare Loan status by Term and Interest Rate Groups



Graph clearly indicates a significant increase in the percentage of defaults compared to fully paid loans, as follows:
Higher Default Probability:
- 13% - 17% Interest Rate Group with 36 Months Term: There is a high probability of defaulting (charged-off loans) compared to fully paid loans within this category.
- 17% - 21% Interest Rate Group with 60 Months Term: There is also a high probability of defaulting within this category.

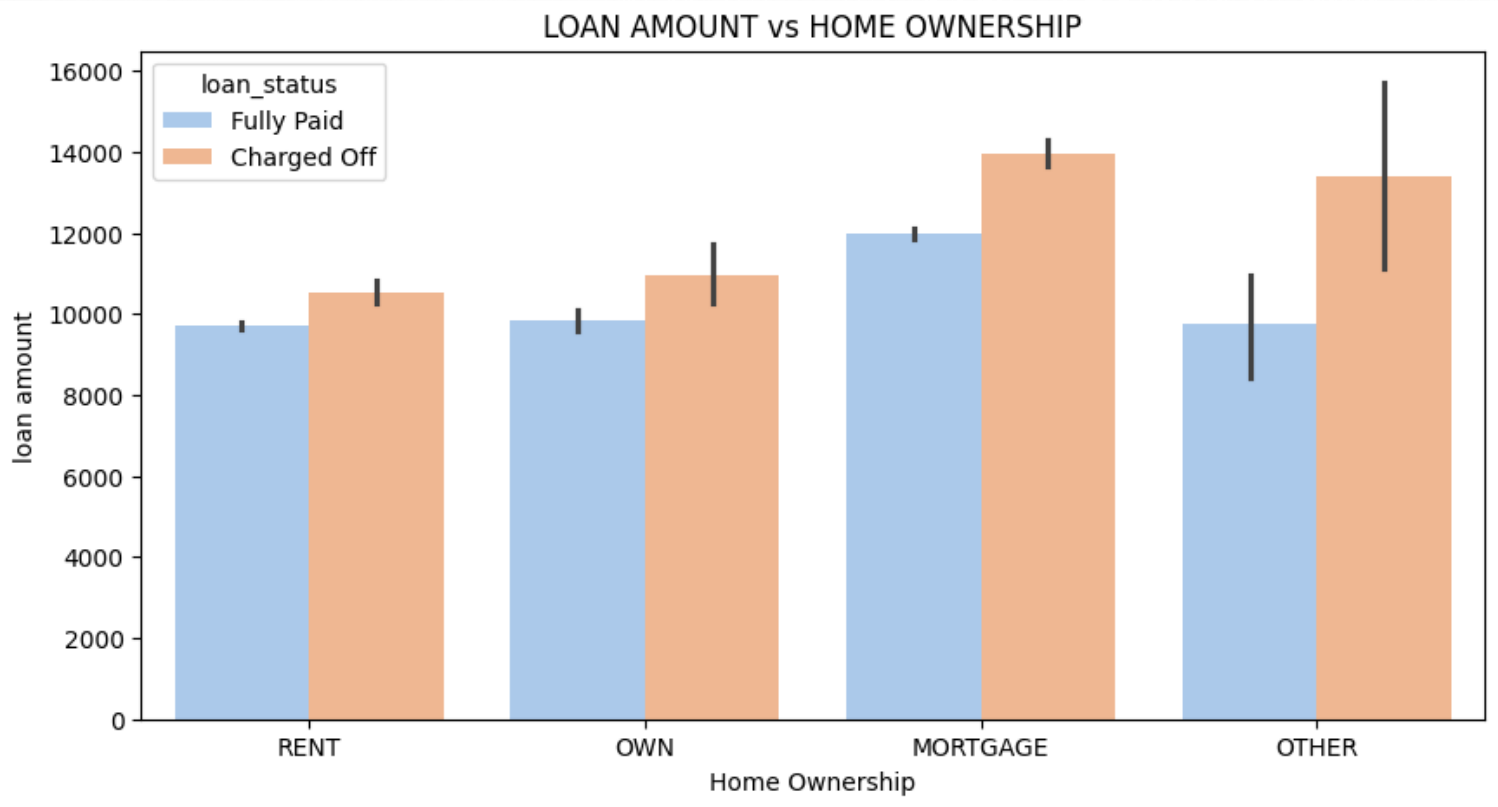
Annual Income Group vs DTI:



loan_status	annual_inc_groups	Charged Off	Fully Paid	total	charge_off_ratio
0	0 - 40k	1705	7734	9439	18.0
1	40k - 50k	847	4744	5591	15.0
2	50k to 60k	854	4634	5488	16.0
3	60k to 70k	543	3484	4027	13.0
4	70k to 80k	451	3011	3462	13.0
5	80k - above	1016	8051	9067	11.0

- People with higher income and in range of 80K+ has low charge offs.

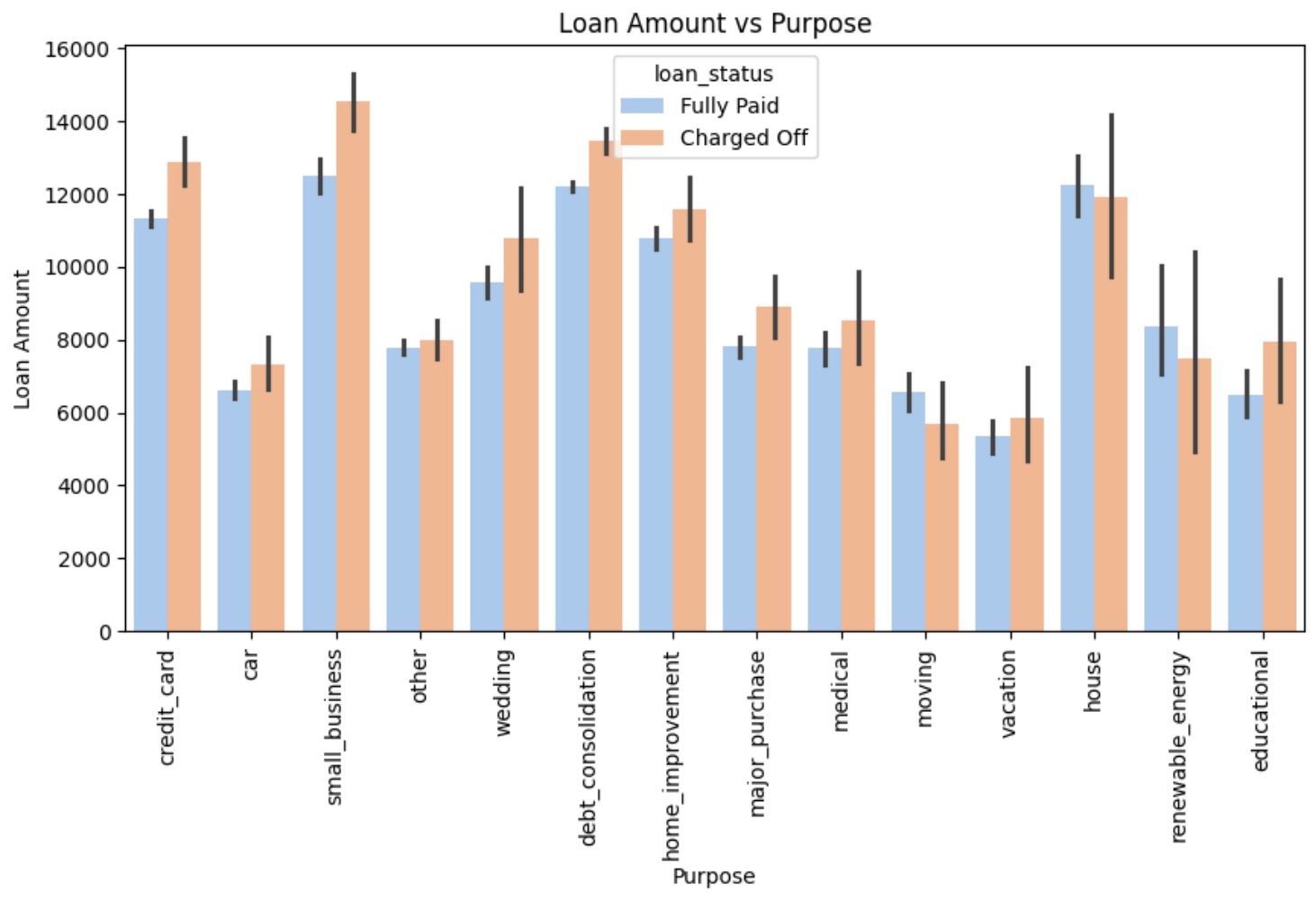
Loan Amount vs Home Ownership:



Applicants with home ownership as 'others' has high chance to default

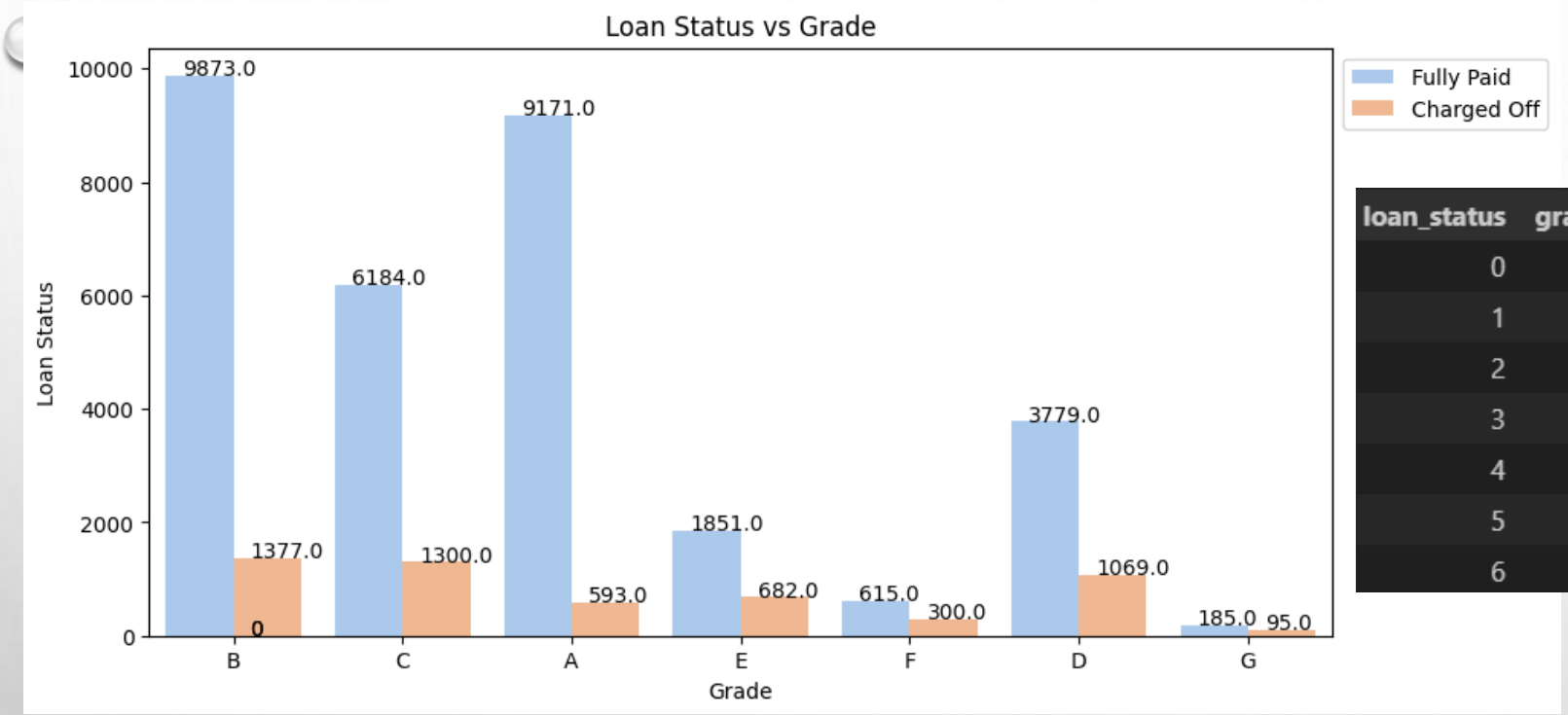
loan_status	home_ownership	Charged Off	Fully Paid	total	charge_off_ratio
0	MORTGAGE	2233	13977	16210	14.0
1	OTHER	17	76	93	18.0
2	OWN	424	2433	2857	15.0
3	RENT	2742	15172	17914	15.0

Loan Amount and Purpose:



loan in category of small business has higher chance to default.

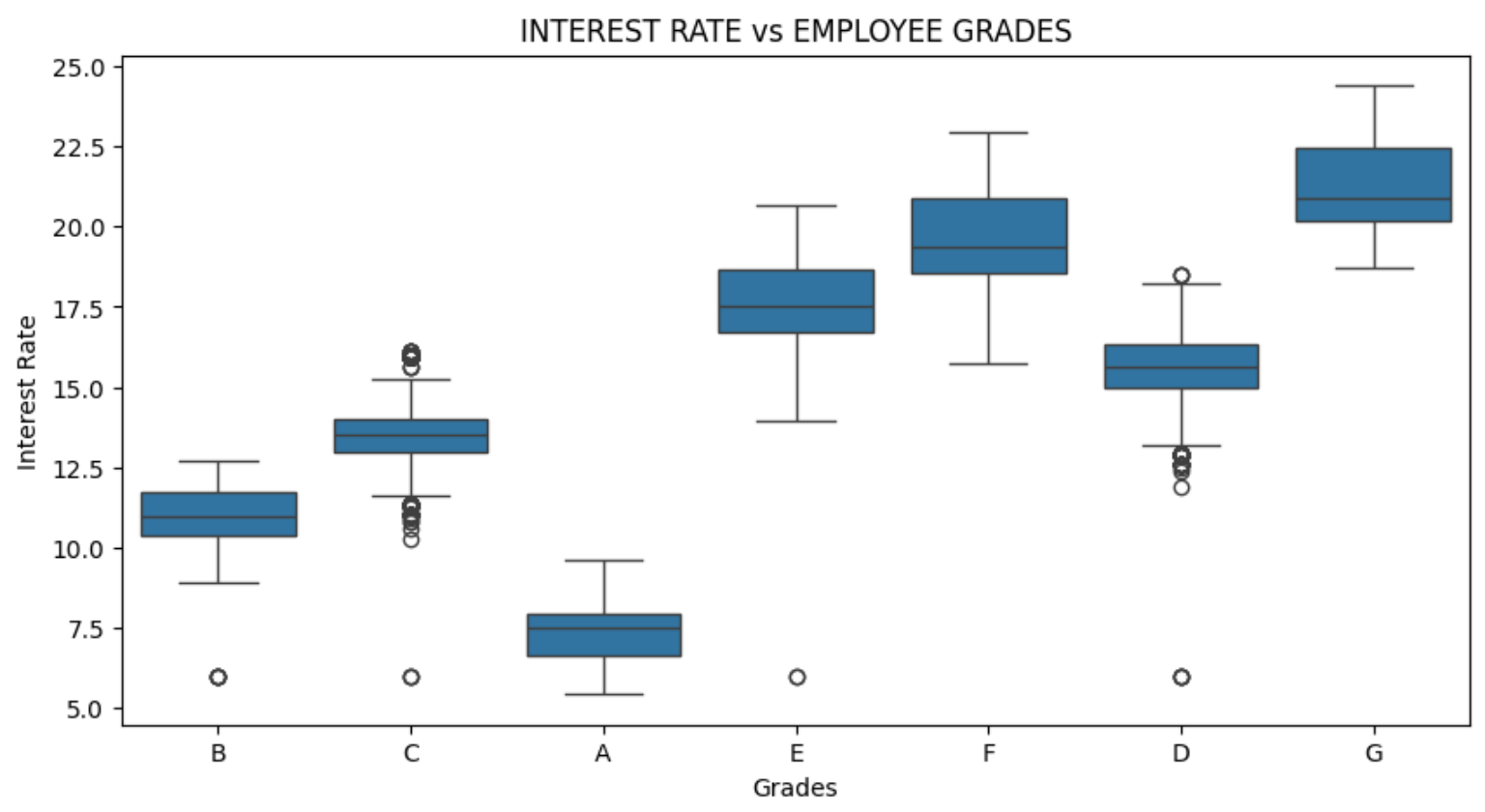
Loan state vs Grade:



loan_status	grade	Charged Off	Fully Paid	total	charge_off_ratio
0	A	593	9171	9764	6.0
1	B	1377	9873	11250	12.0
2	C	1300	6184	7484	17.0
3	D	1069	3779	4848	22.0
4	E	682	1851	2533	27.0
5	F	300	615	915	33.0
6	G	95	185	280	34.0

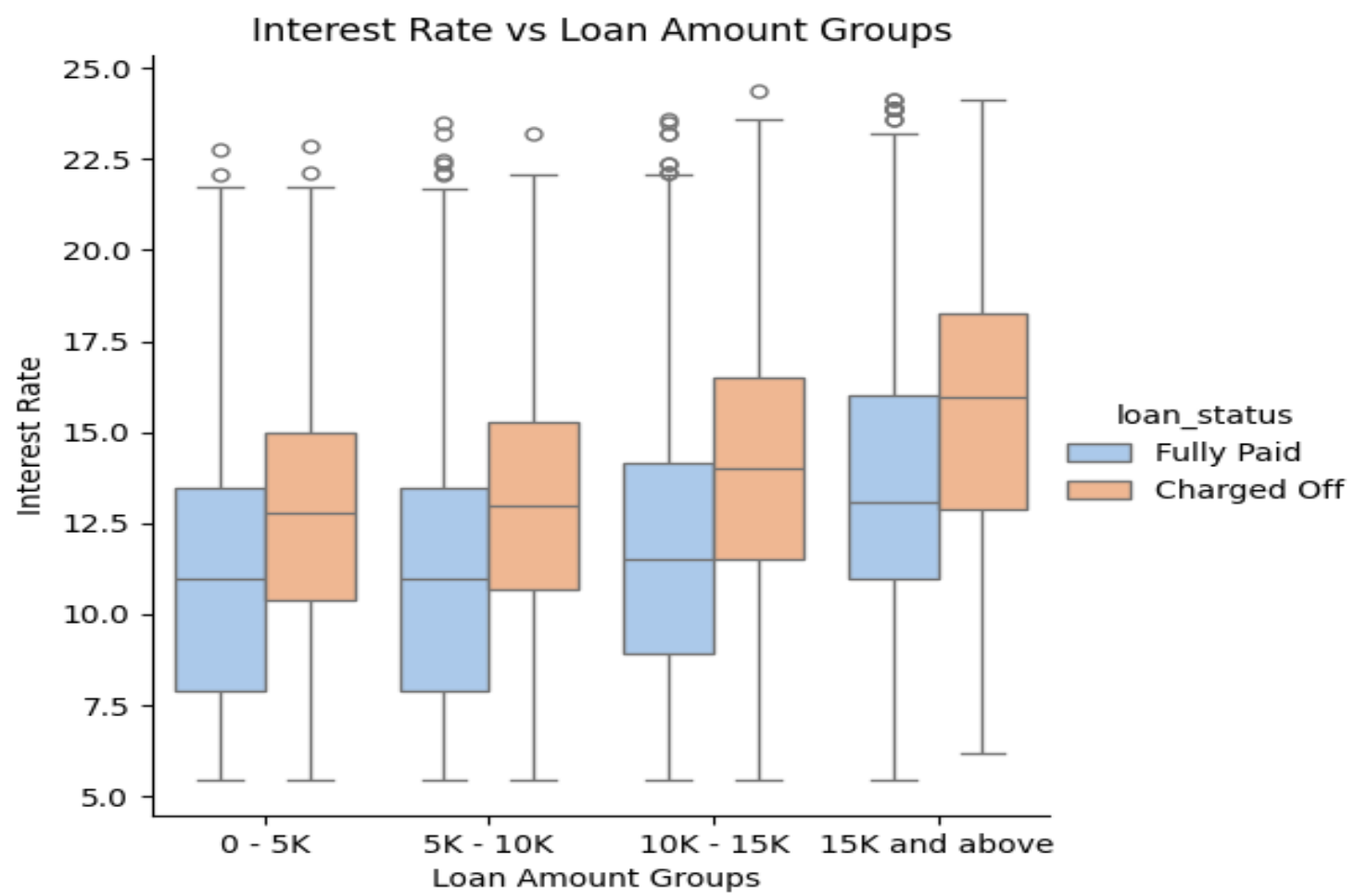
- Applicants belonging to Grade G, F, E, D, C, B has high charged of Ratio than A and increase the risk to default.

Interest Rate vs Grade:



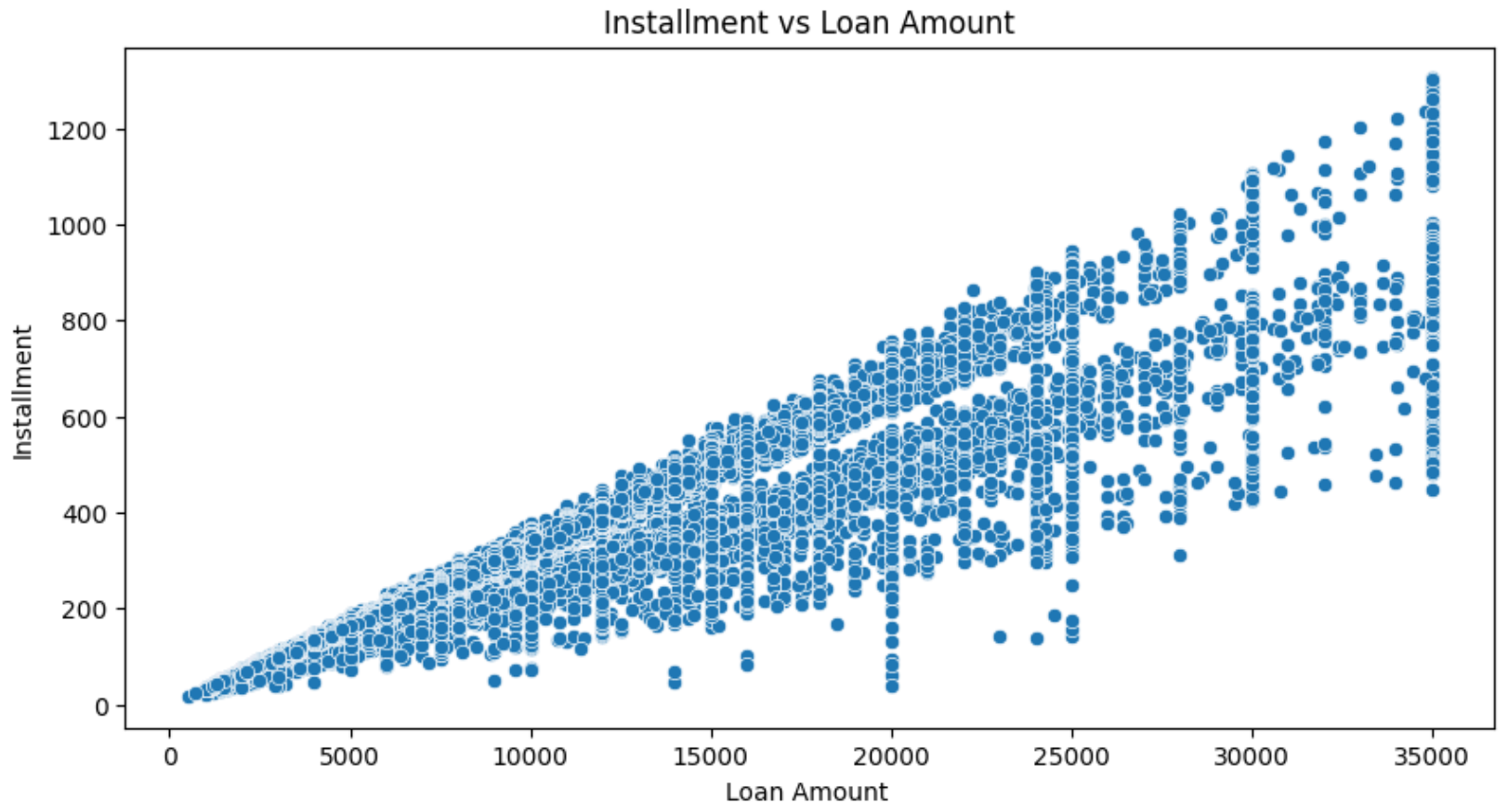
customer with Grade A has lowest interest rate while Grade G has highest rate of interest. Interest rates increases gradually moving form A to G.

Interest Rate vs Loan Amount Group:



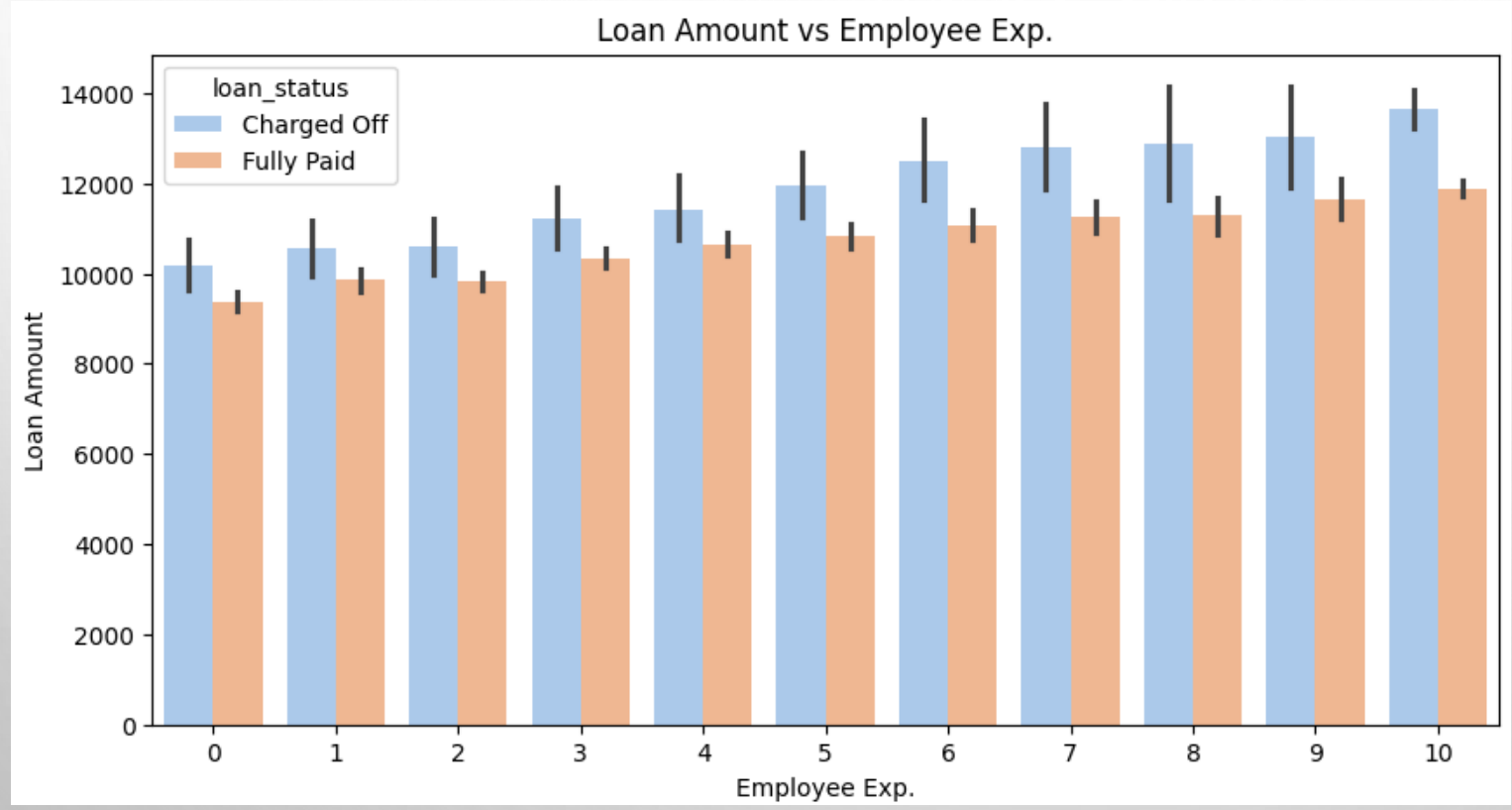
- High interest rate with loan amount15k+ above has more chance to default.

Installment vs Loan Amount:



- with increase in loan amount installment per month also increases linearly.

Loan Amount vs Employee Experience:



- People with higher experience has the higher chance to default.

THANK YOU