

Lending Club Case Study

Srujana Bhamidipalli

Business Understanding

You work for a consumer finance company which specialises in lending various types of loans to urban customers. 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

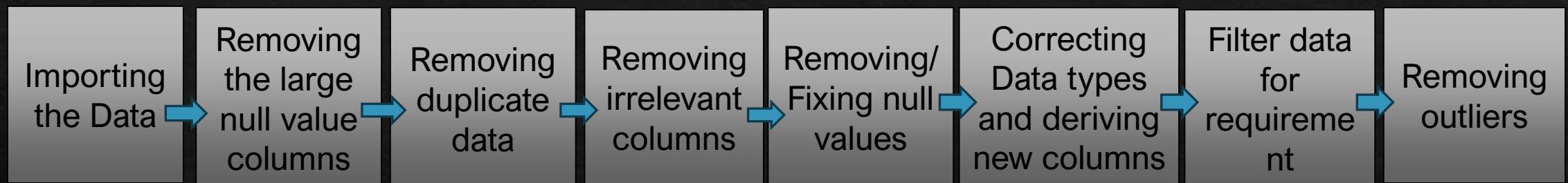
The data given below contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

General Information

This project can be seen as a data science and machine learning case study aimed at understanding loan defaults, performing risk assessment, and providing business insights.

- The main goal of the project is to analyze Lending Club loan data to build predictive models that can assess the risk of loan defaults.
- Python scripts for cleaning and manipulating the data.
- Visualizations and statistical analysis to provide a detailed understanding of the data.
- A trained machine learning model capable of predicting loan defaults.
- A detailed report on the model's performance, including evaluation metrics and potential areas for improvement.
- A summary of the key insights derived from the data and how they can be applied in a business context.

Data cleanup and manipulation process



Outlier Treatment

Clearly indicating the presence of outliers.

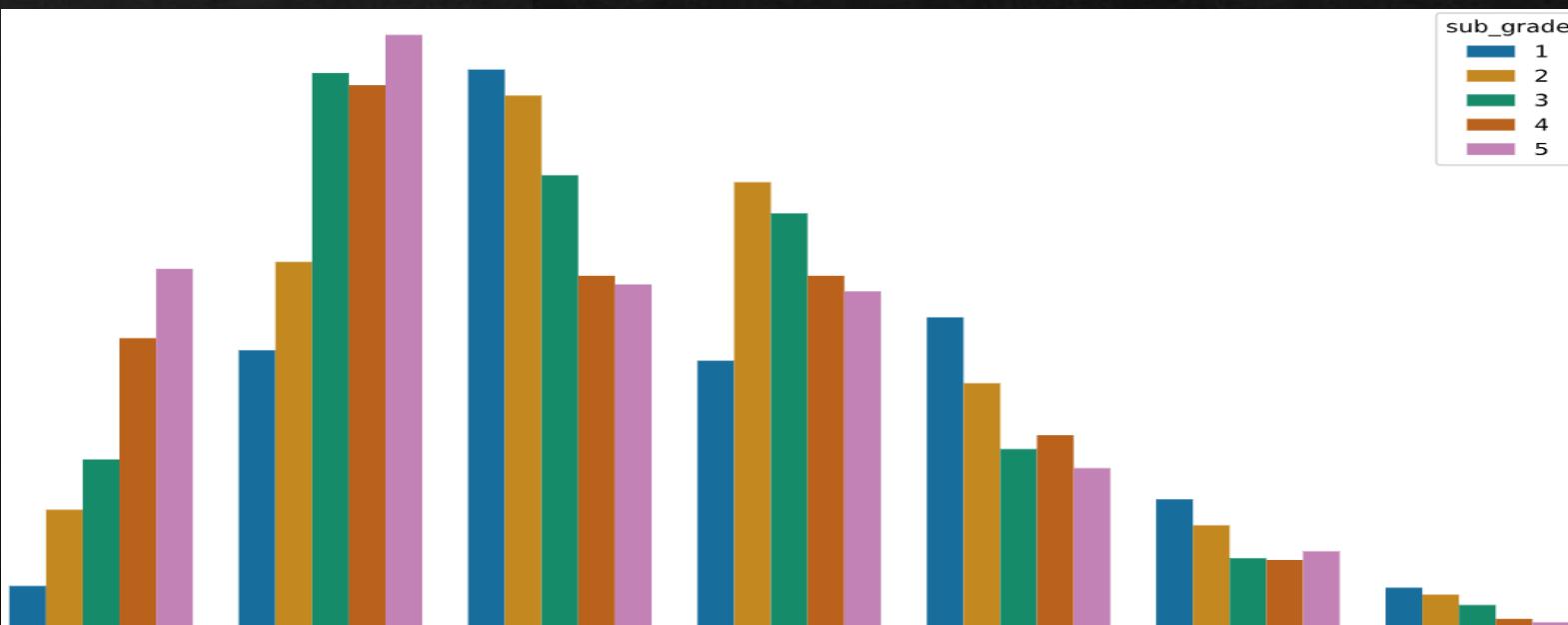
- So, Removing them.
- Let's see the quantile info and take an appropriate action.
- The values after 95 percentile seems to be disconnected from the general distribution and also there is huge increase in the value for small quantile variation.
- So, considering threshold for removing outliers as 0.95



Visualizing Categorical Data

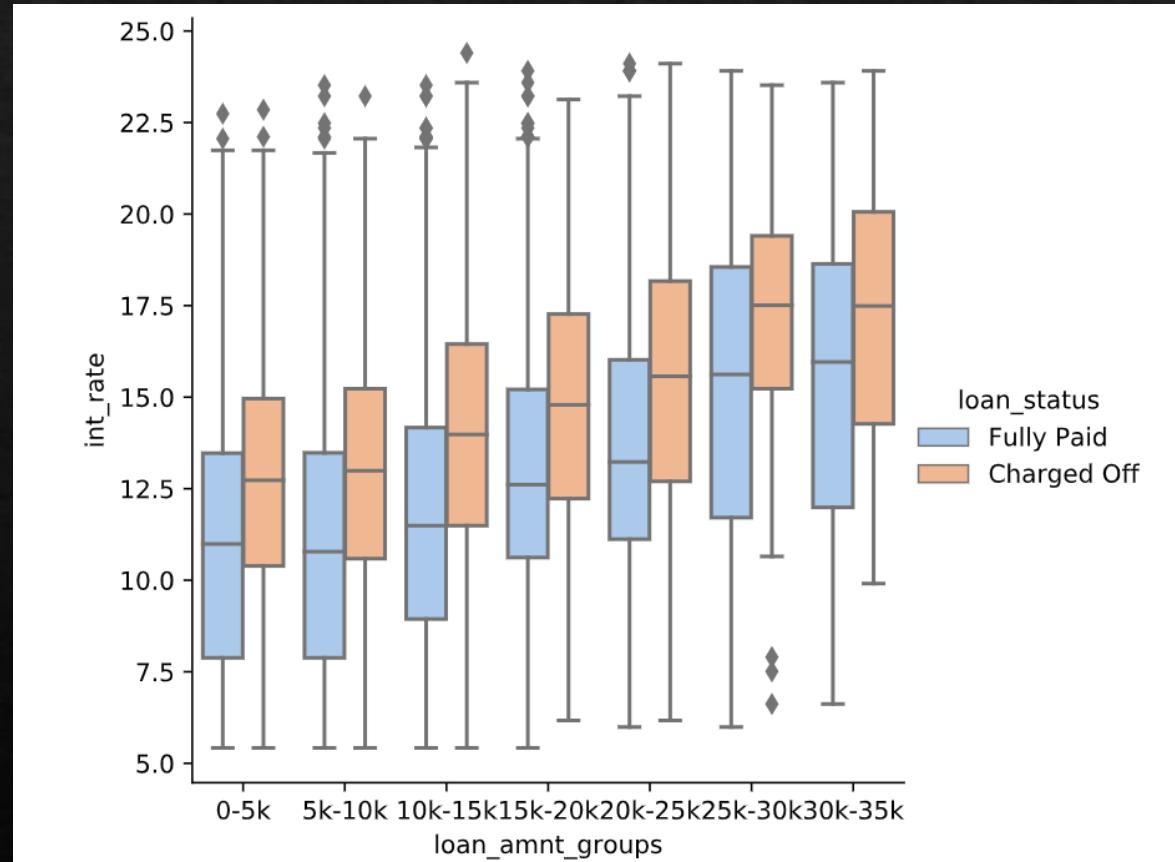
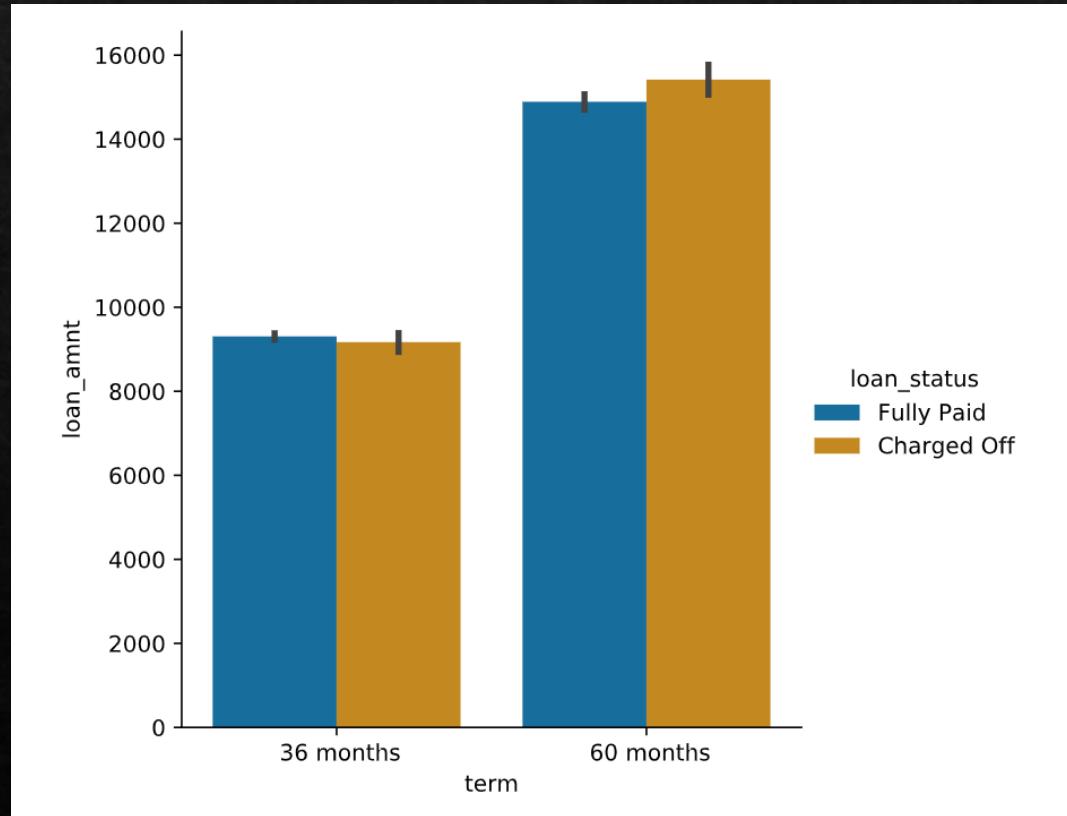
As we already have grade column, extracting only subgrade (int level value) from the sub_grade variable

We are analyzing and visualizing only the defaulter data. So subsetting the data while plotting only for 'Charged Off' loan_status for below plots



The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan_amount groups.

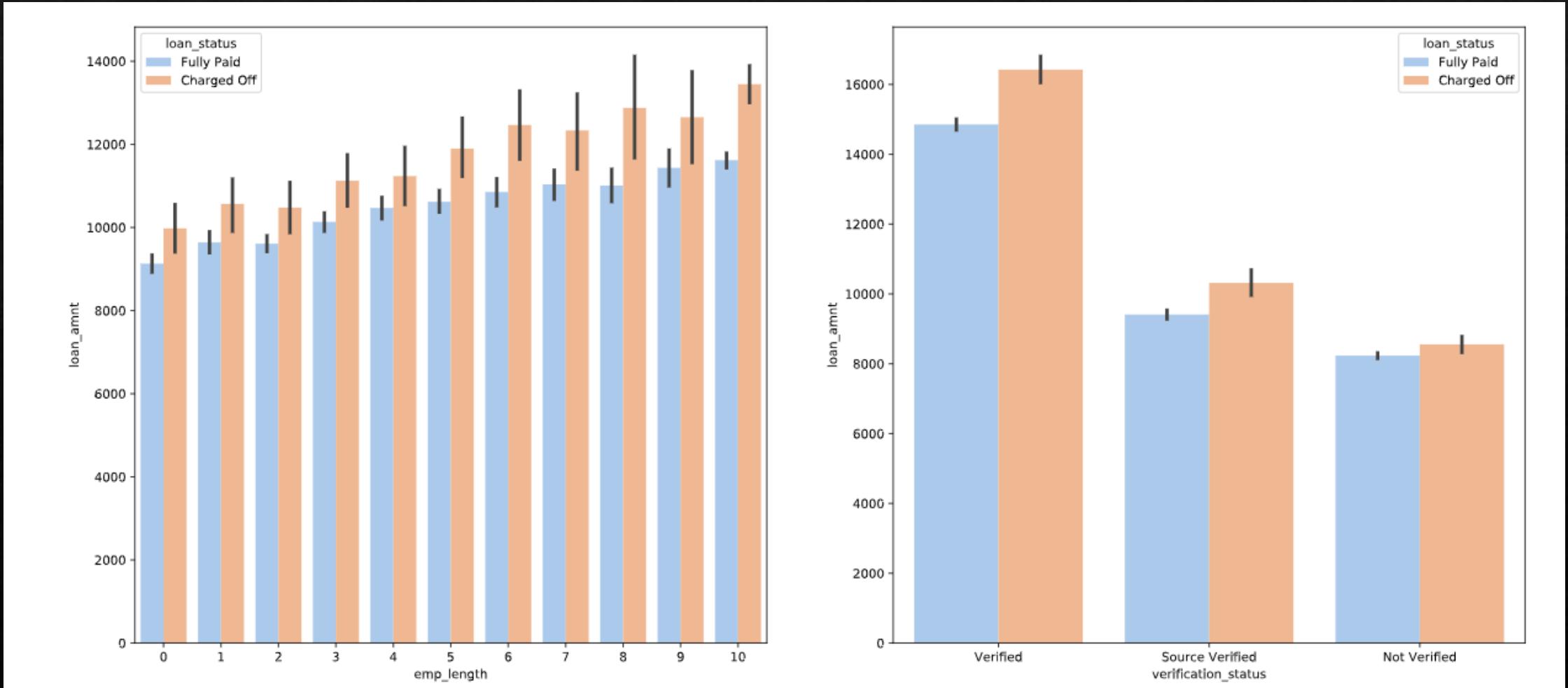
- This can be a pretty strong driving factor for loan defaulting.



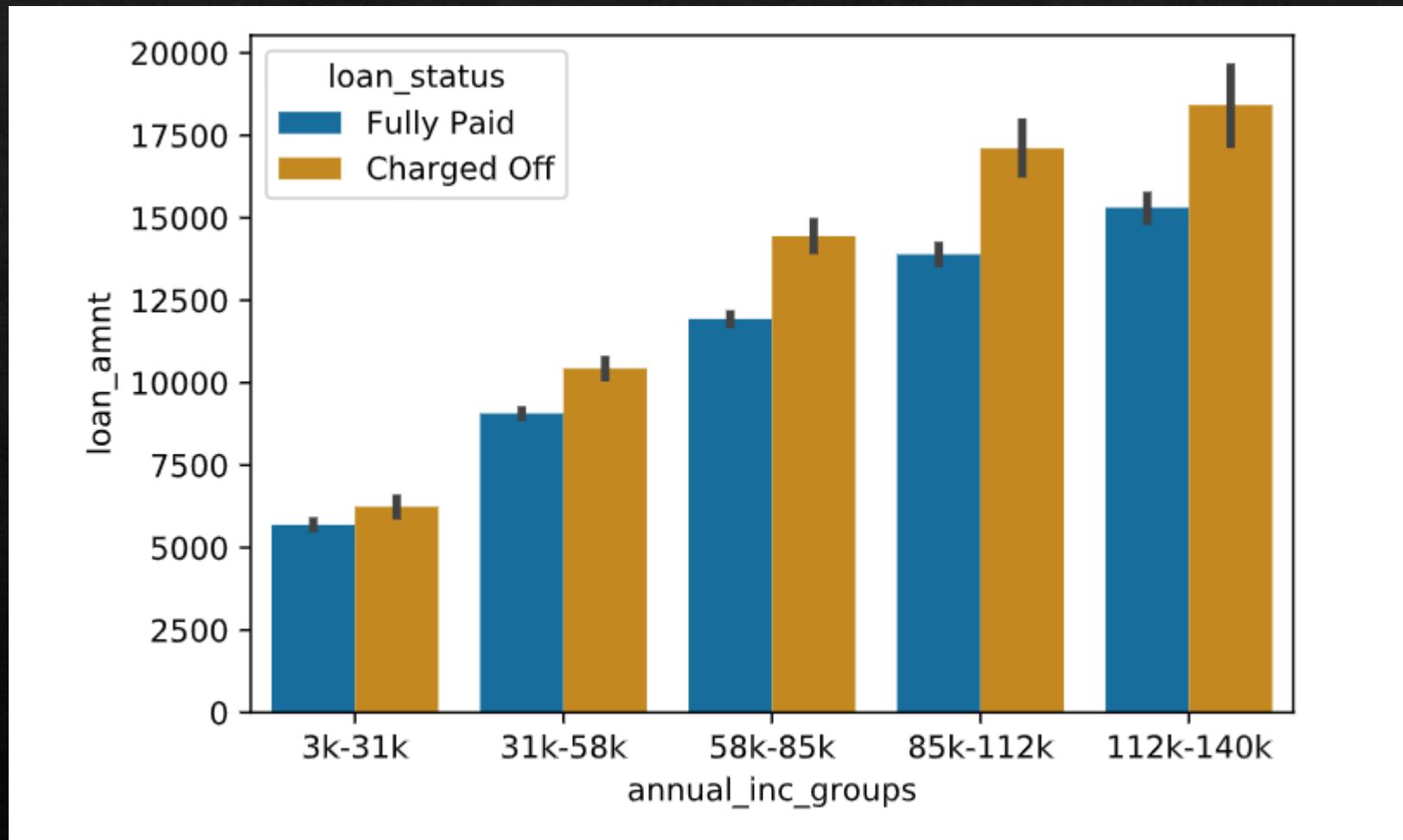
- ❖ **Observations**
- ❖ **The above analysis with respect to the charged off loans for each variable suggests the following. There is a more probability of defaulting when :**
 - Applicants having house_ownership as 'RENT'
 - Applicants who use the loan to clear other debts
 - Applicants who receive interest at the rate of 13-17%
 - Applicants who have an income of range 31201 - 58402
 - Applicants who have 20-37 open_acc
 - Applicants with employmenent length of 10
 - When funded amount by investor is between 5000-10000
 - Loan amount is between 5429 - 10357
 - Dti is between 12-18
 - When monthly installments are between 145-274
 - Term of 36 months
 - When the loan status is Not verified
 - When the no of enquiries in last 6 months is 0
 - When the number of derogatory public records is 0
 - When the purpose is 'debt_consolidation'
 - Grade is 'B'
 - And a total grade of 'B5' level.

Employees with longer working history got the loan approved for a higher amount.

- Looking at the verification status data, verified loan applications tend to have higher loan amount. Which might indicate that the firms are first verifying the loans with higher values.



- ❖ Across all the income groups, the loan_amount is higher for people who defaulted.



The above analysis with respect to the charged off loans. There is a more probability of defaulting when :

- Applicants taking loan for 'home improvement' and have income of 60k -70k
- Applicants whose home ownership is 'MORTGAGE and have income of 60-70k
- Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k
- Applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %
- Applicants who have taken a loan for small business and the loan amount is greater than 14k
- Applicants whose home ownership is 'MORTGAGE and have loan of 14-16k
- When grade is F and loan amount is between 15k-20k
- When employment length is 10yrs and loan amount is 12k-14k
- When the loan is verified, and loan amount is above 16k
- For grade G and interest rate above 20%