



LENDING CLUB CASE STUDY

N Vinay Kumar & Maka Sitaram

PROBLEM STATEMENT

Lending Club a renowned finance company which lends loans to customers has to go through few steps before approving a loan. When the applicant drafts his or her willingness for loan, the company must check the eligibility of each customer before processing a loan. There are two types of risks 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.

OBJECTIVE

As a data scientist working for Lending Club analyze the dataset containing information about past loan applicants using EDA to understand how consumer attributes and loan attributes influence the tendency of default.

The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with money owed, this would result in financial loss to the bank, the customers who do this are labelled as 'charged-off or the 'defaulter'.

The Dataset contains loans issued between 2007 to 2011, and as a Data Scientist analyzing the whole dataset involves:

1. Identification of loan applicant traits that tend to "Default" paying back.
2. Understand the "Deriving Factors" or "Derived Variables" behind loan default.
3. Scrutinize the new loan applicant's portfolio and risk assignment based on historic data.

STEPS INVOLVED

DATA EXPLORING

1. No. Of Columns
2. Understand the columns and their data types
3. With the help of data dictionary understand the meaning of each column

DATA CLEANING

1. Identify the columns & row with nulls, missing & unique values
2. Remove the Nulls, high missing percentage
3. Create new columns if required for identification of new metric

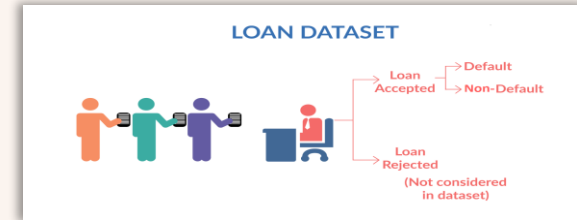
EXPLORATORY DATA ANALYSIS

1. Univariate
2. Univariate Segmented
3. Multivariate

DATA VISUALIZATION

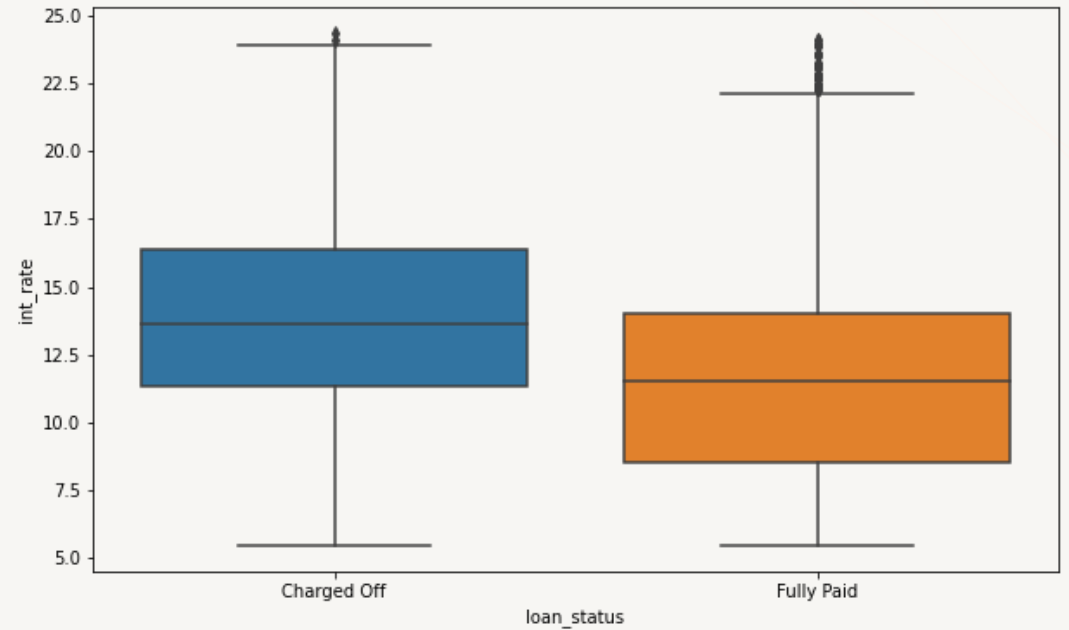
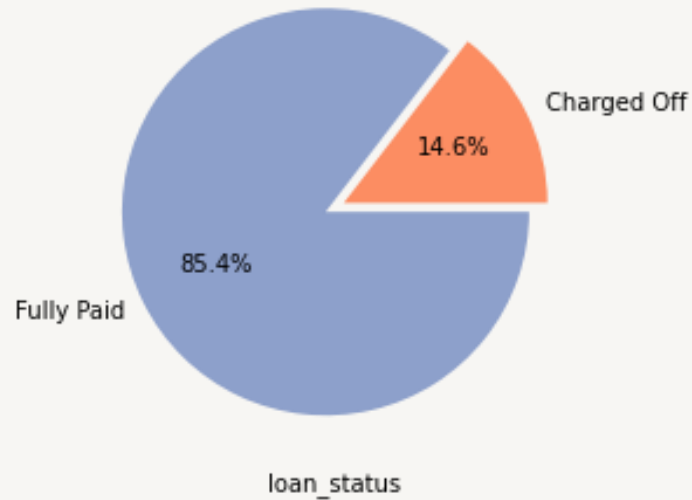
Using Seaborn's boxplot, dist plot, bar plot, pie charts visualize the output and provide insights

VARIABLES INVOLVED IN ANALYSIS



Variable Type	Desscription
loan_status	Current status of the loan
purpose	Pupose of the loan
addr_state	The state provided by the borrower in the loan application
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified
grade	LC assigned loan grade
sub_grade	LC assigned loan subgrade
home_ownership	The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
int_rate	Interest Rate on the loan
dti	debt to income
annual_inc	The self-reported annual income provided by the borrower during registration.
issue_d	Interest Rate on the loan

LOAN STATUS ANALYSIS



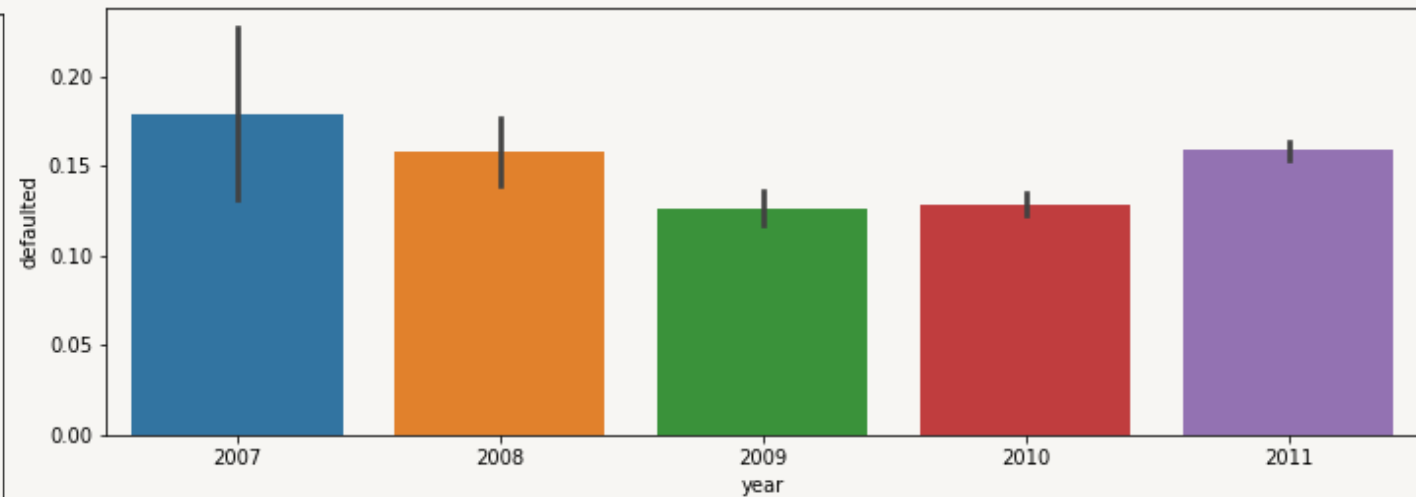
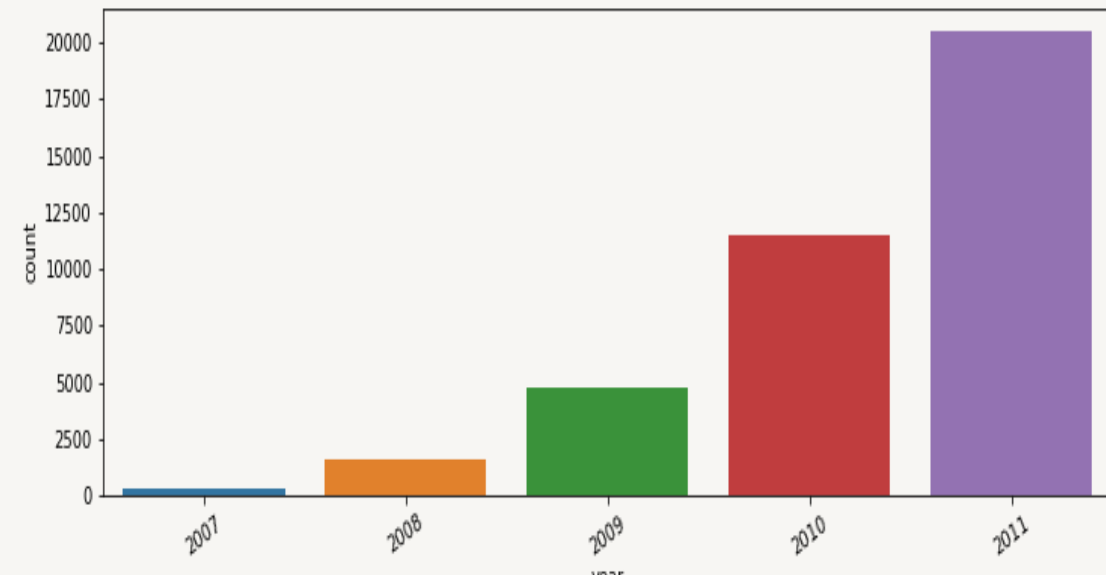
The above Pie Chart depicts the following insights:

- ❖ Average Default rate is **14.6%**
- ❖ Defaulters **have higher interest rate** than Fully paid

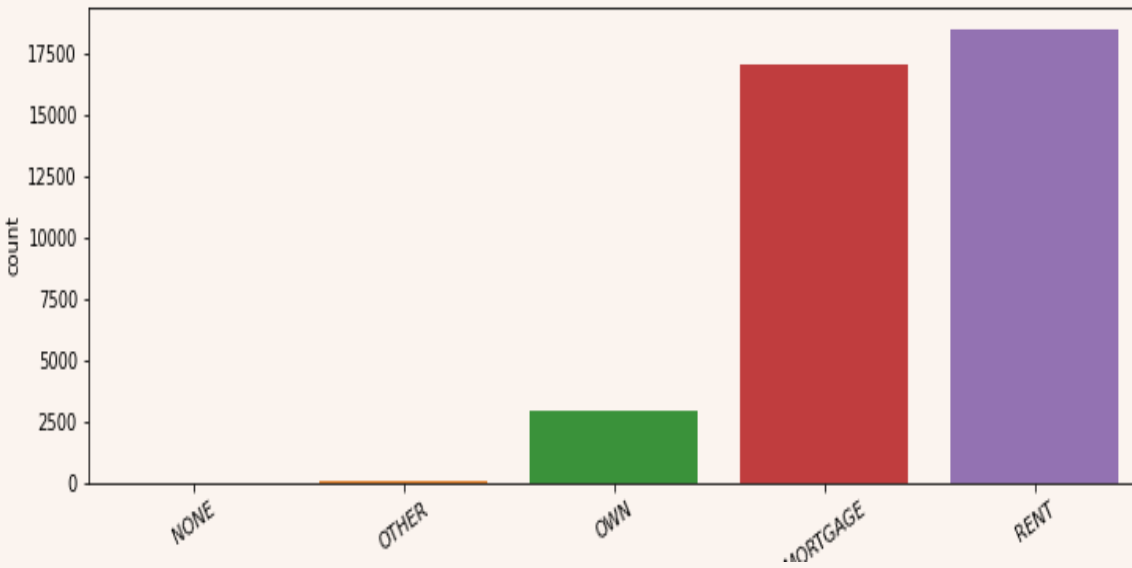
The variance between the fully paid customers and defaulters is high, however if the bank continues to give loan to defaulters, then there is a high risk to the bank.

LOAN ISSUE YEAR ANALYSIS

- The outreach of loans grew substantially over the years this could either be the result of market expansion of bank. As the graph depicts, customers opted for loans in the year 2007 are comparatively very less to customers who took loan in 2011. There is a **sudden spike of loans from 2010 to 2011**
- Though the number of loans approved in the year 2007 are very less, but the defaulters list is the maximum among all the years.

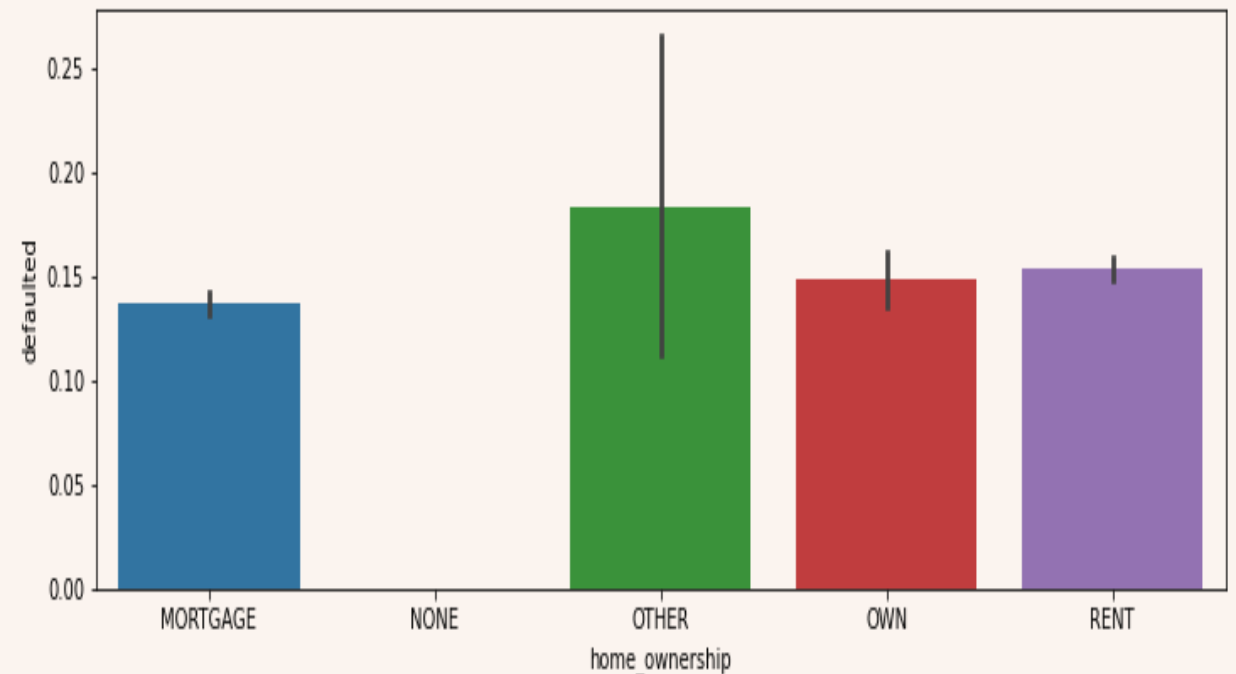


HOME OWNERSHIP ANALYSIS

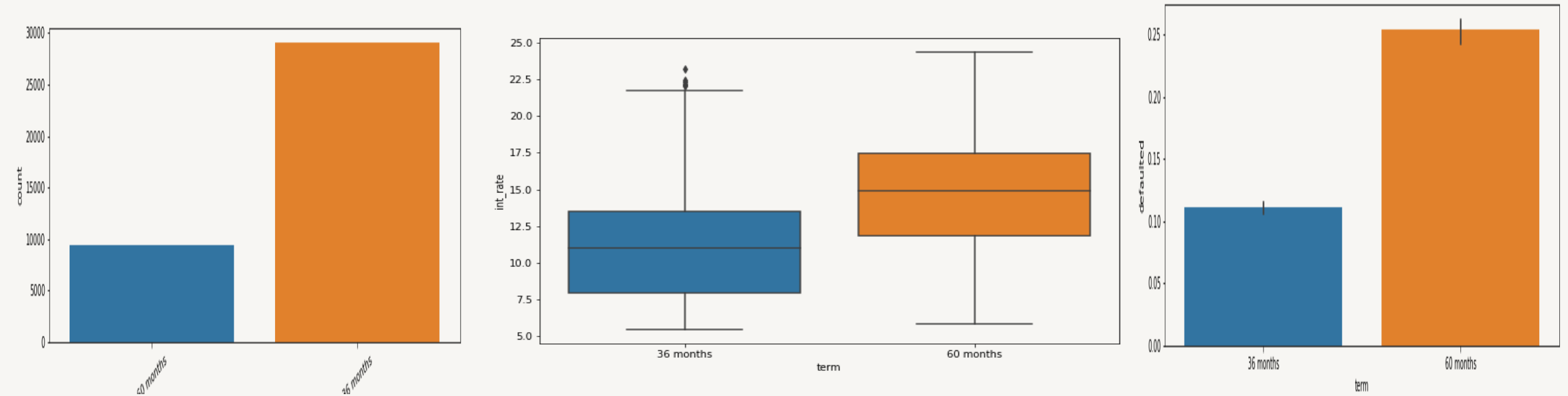


- ❖ Customers whose purpose of loan is stated as OTHER show high defaulter's rate

- ❖ Customers who live on rent apply for loans more comparatively, this could either be because the customer has taken a loan for himself to construct a new house or for other purpose

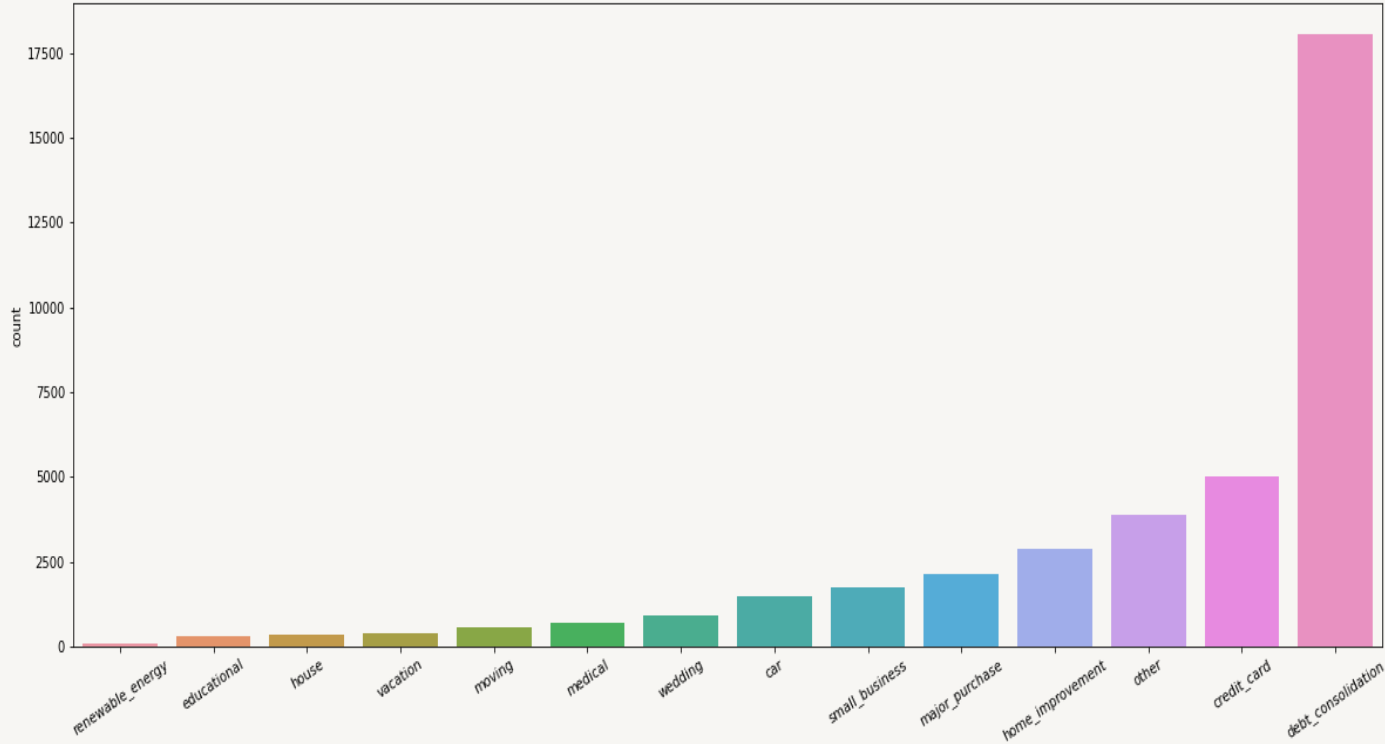


TERM-ANALYSIS

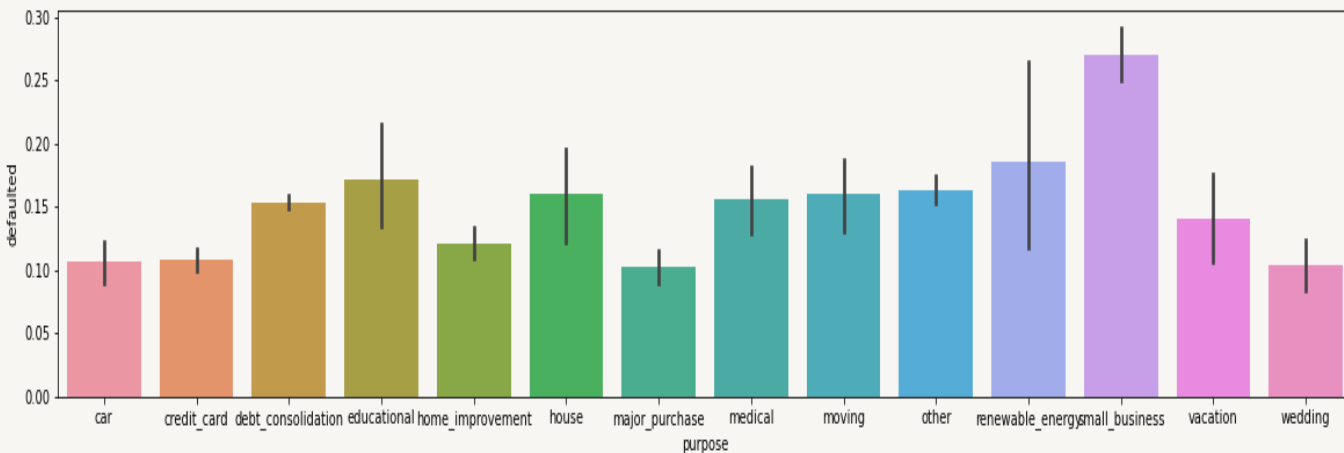


- ❖ Most people who have taken loan preferred 36 months tenure over 60 months tenure
- ❖ People who have chosen 60 months as their tenure period are more prone to become defaulters, this is due to the high interest they end up paying.

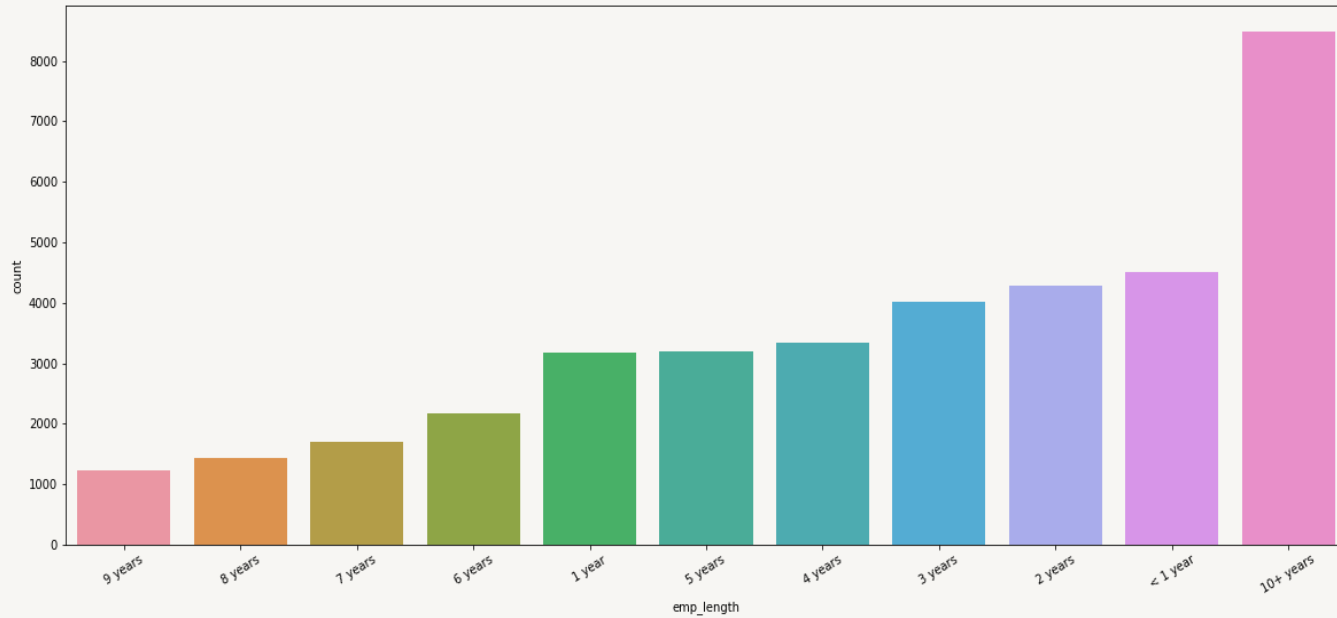
PURPOSE-ANALYSIS



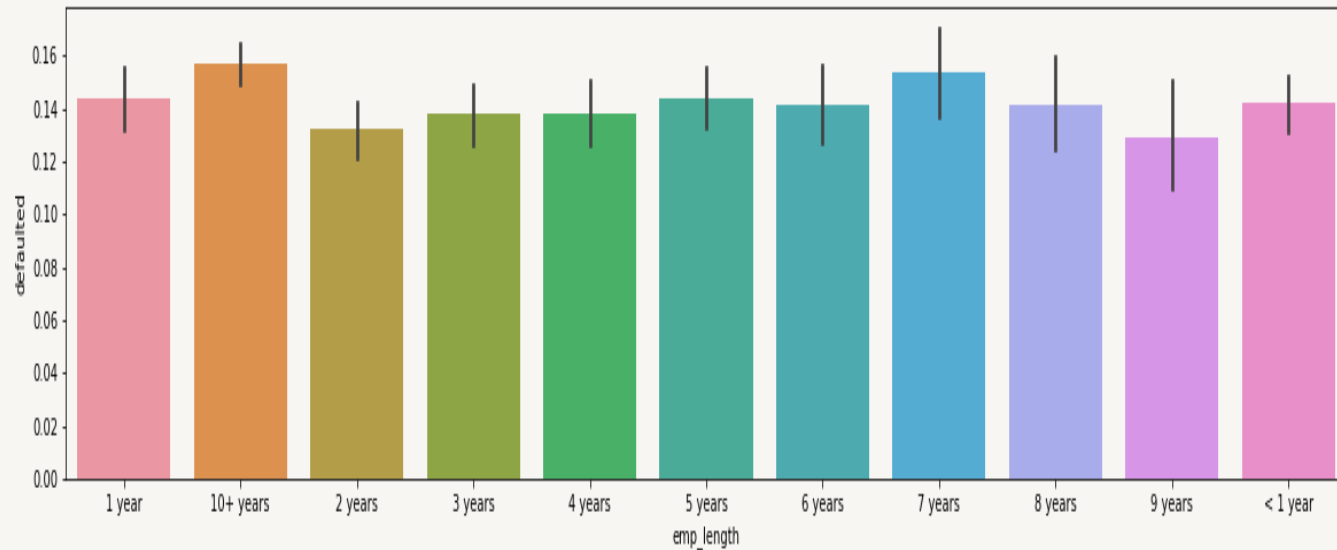
- ❖ Purpose of loan for most of the customer is stated as debt consolidation, meaning customers prefer taking loans to clear out their past loans
- ❖ Customers who have taken loan for small business purpose end up becoming defaulters, this could be due to loss incurred in the business
- ❖ Bank should carefully examine the idea behind the business before lending out loans.



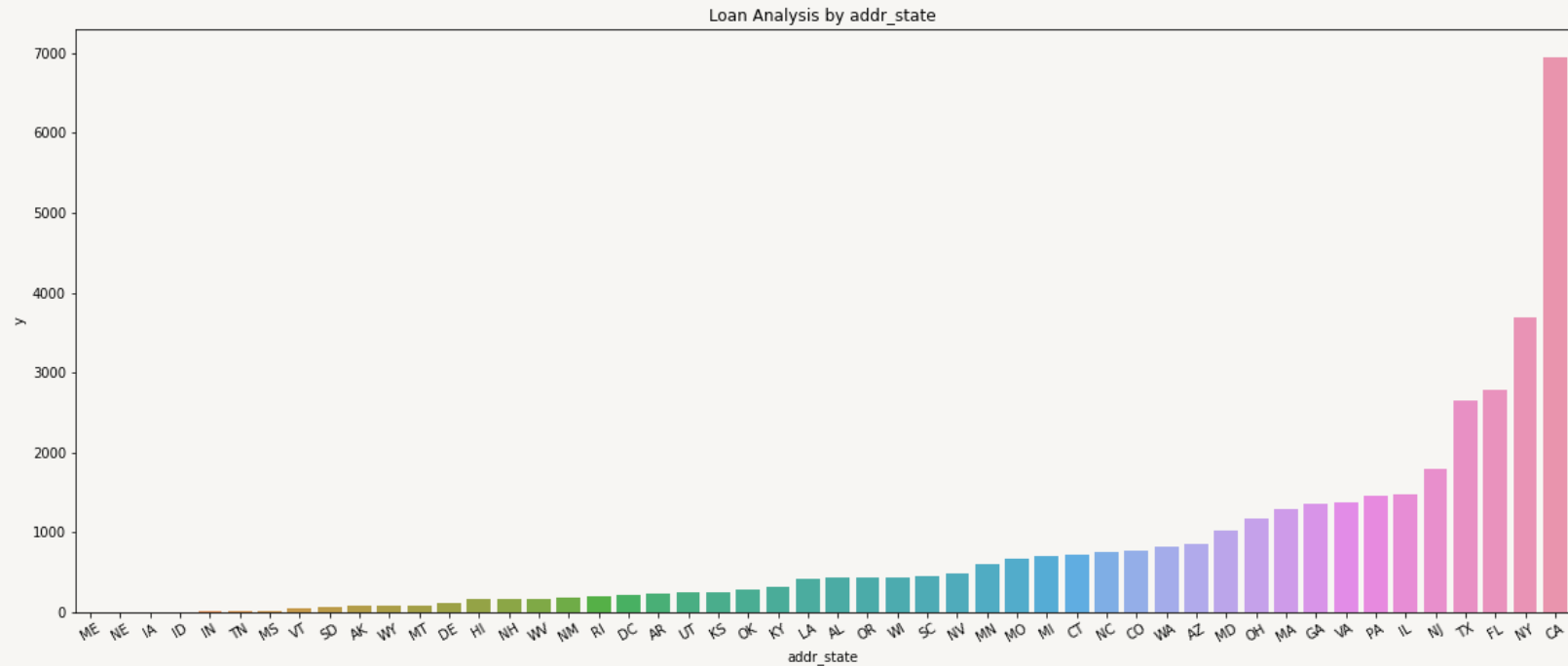
EMPLOYEE LENGTH ANALYSIS



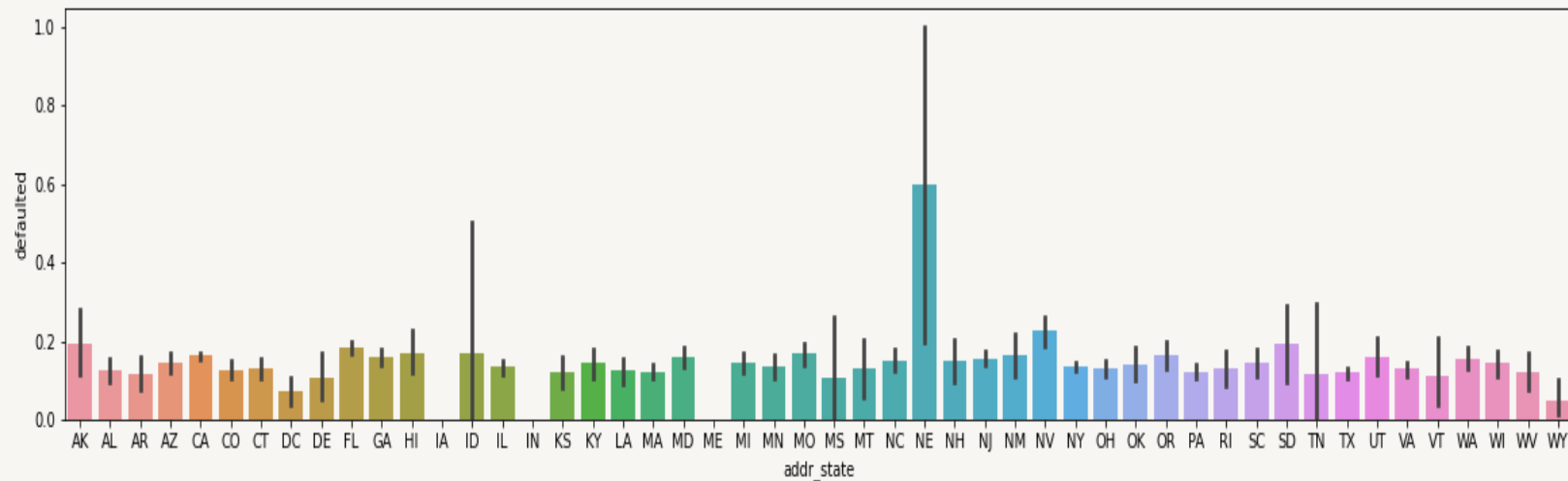
- ❖ Customers with higher experience have taken more loans when compared with customers having less than 10 years experience.
- ❖ The defaulters among different levels of experienced customers is consistent.



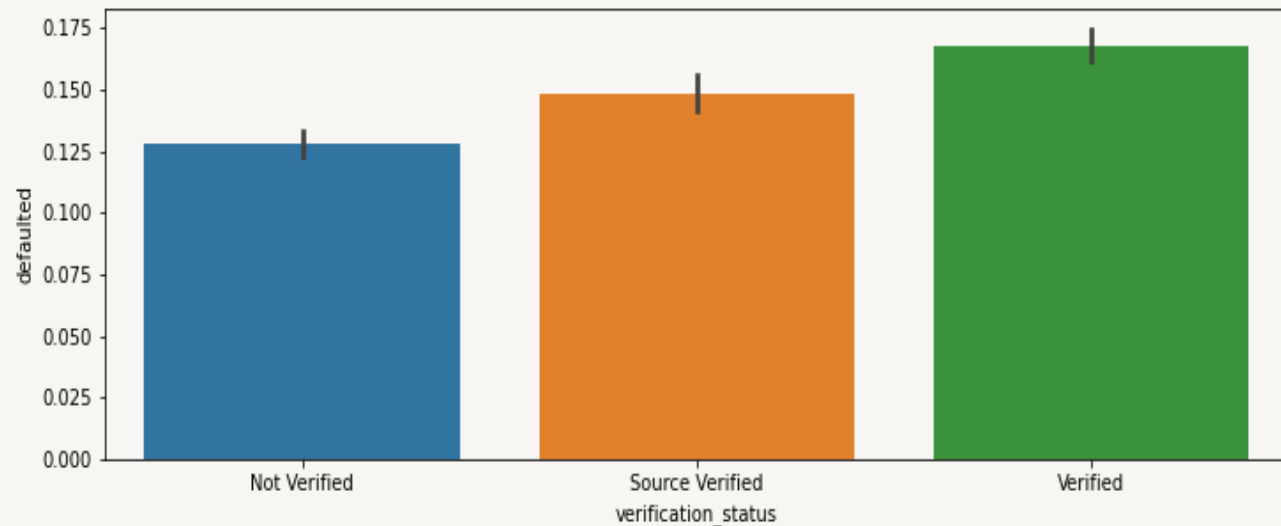
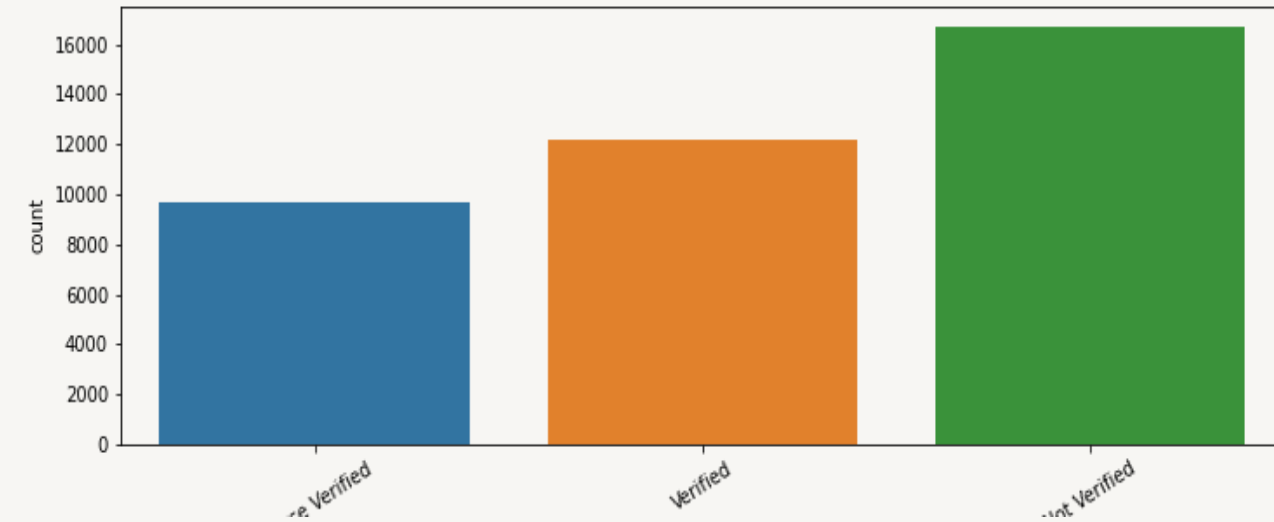
ADDR_STATE-ANALYSIS



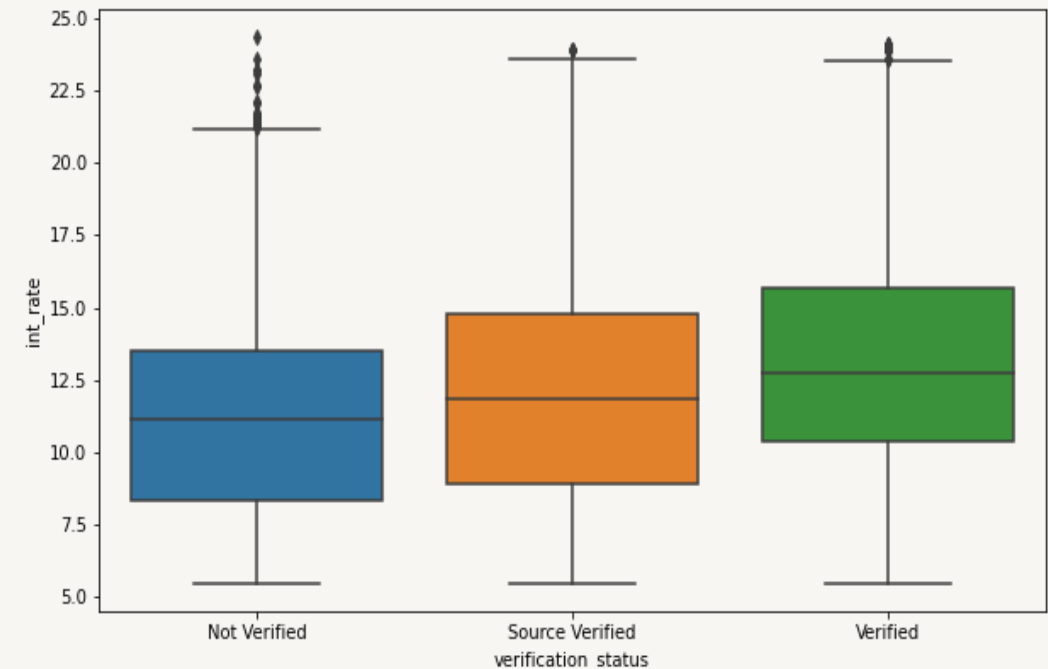
- ❖ Most of the loan borrowers are coming from **California (CA)** and second least from **Nebraska (NE)**
- ❖ Most of the defaulters are coming from **Nebraska (NE)**, more than half the total borrowers haven't paid loan back



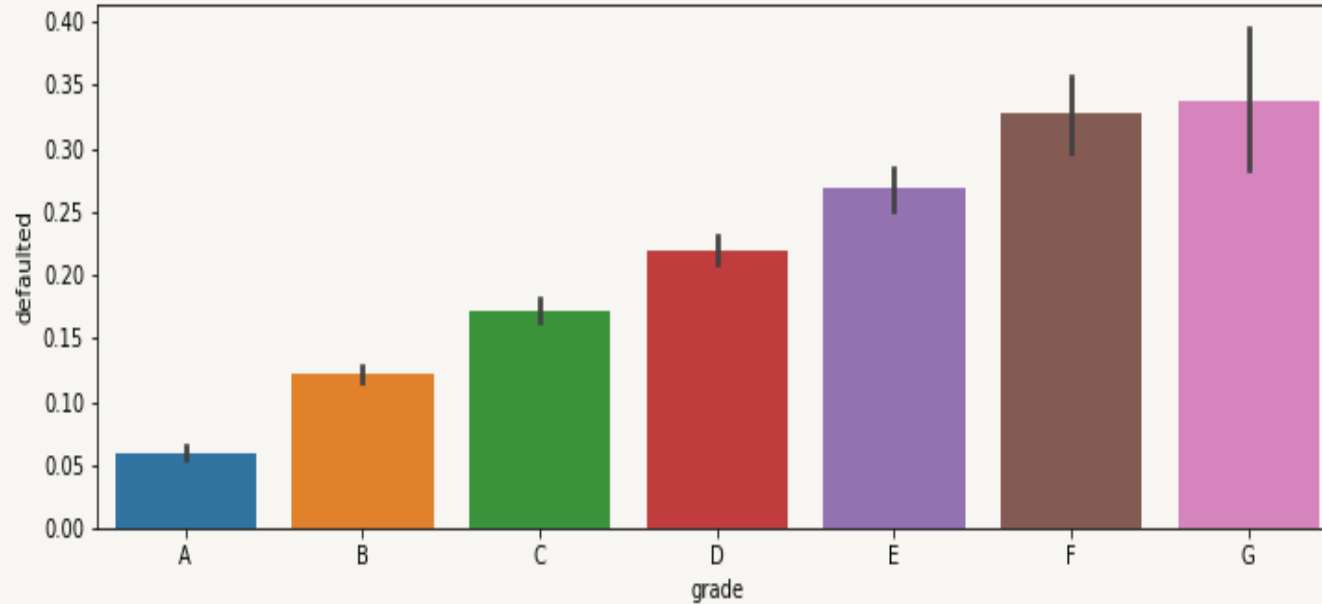
VERIFICATION_STATUS-ANALYSIS



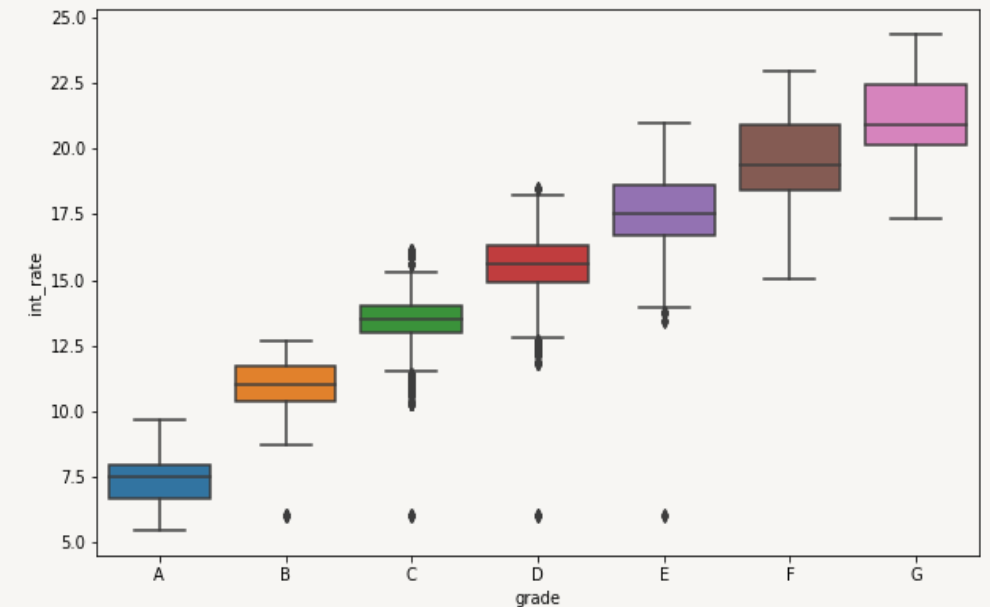
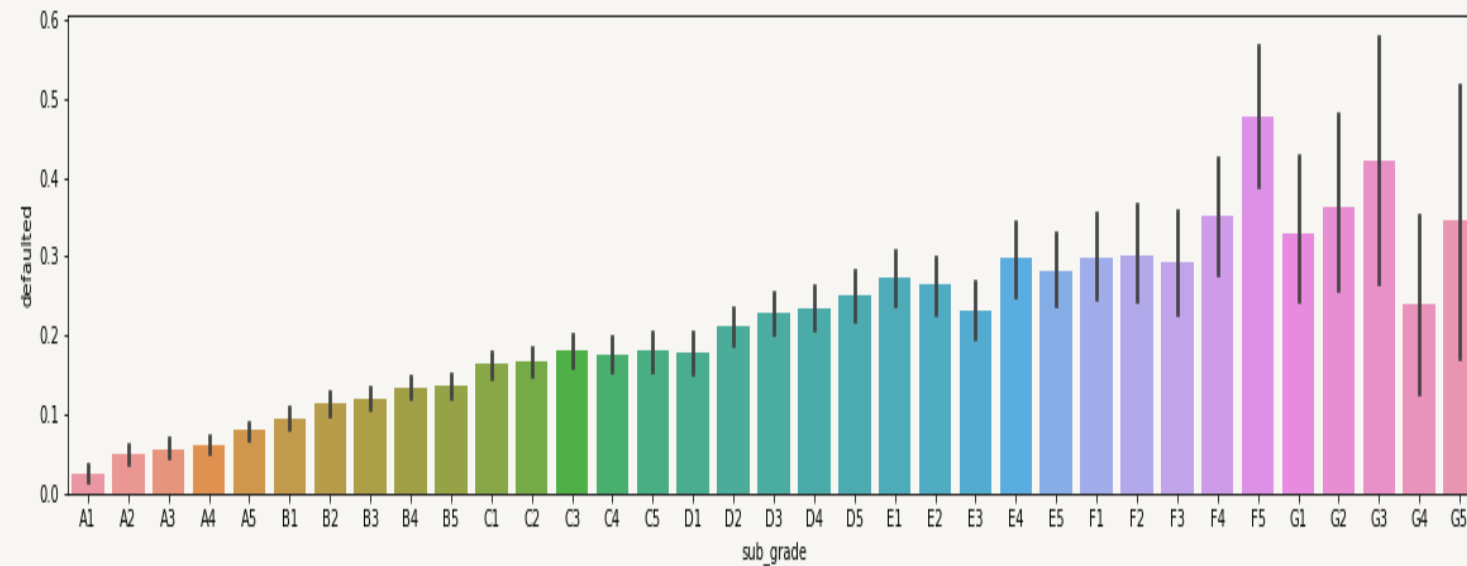
- ❖ Most of the loan borrowers are given loan by bank consultants without any background verification, and this is bad for business.
- ❖ Defaulters rate stand high for verified customers, this could be due to bad verification checks
- ❖ However the interest rate for Verified borrowers is higher than that of Non Verified Borrowers



GRADE-ANALYSIS

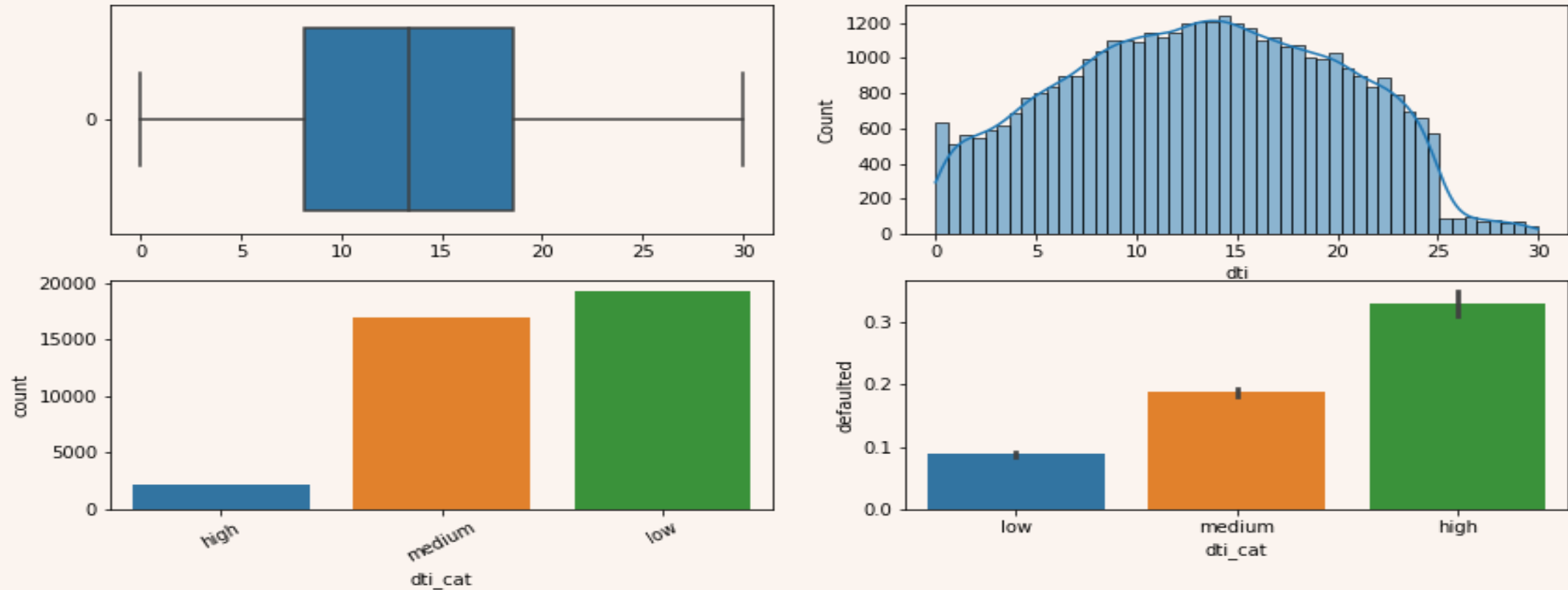


- ❖ Customers who have taken Grade G loans are highest defaulters, and Grade A loan borrowers are the lowest defaulters.
- ❖ From the analysis Grade A loans have low interest rates and hence the defaulters are the least for Grade A loans.



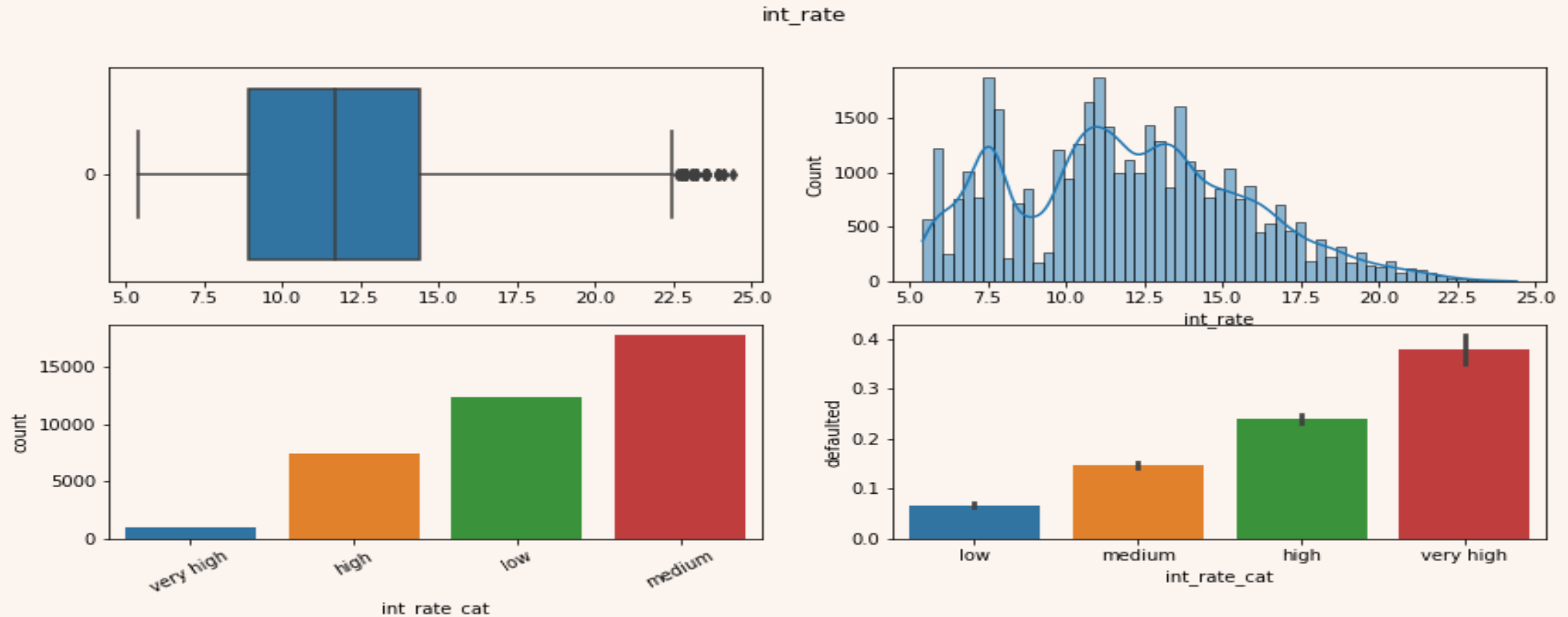
DTI-ANALYSIS

dti



- ❖ DTI (Debt to Income) average for the given dataset is 13.37
- ❖ In-case of low DTI, the borrower is performing as expected and the chances of customer being labelled as defaulter are too low
- ❖ DTI is inversely proportional to defaulters

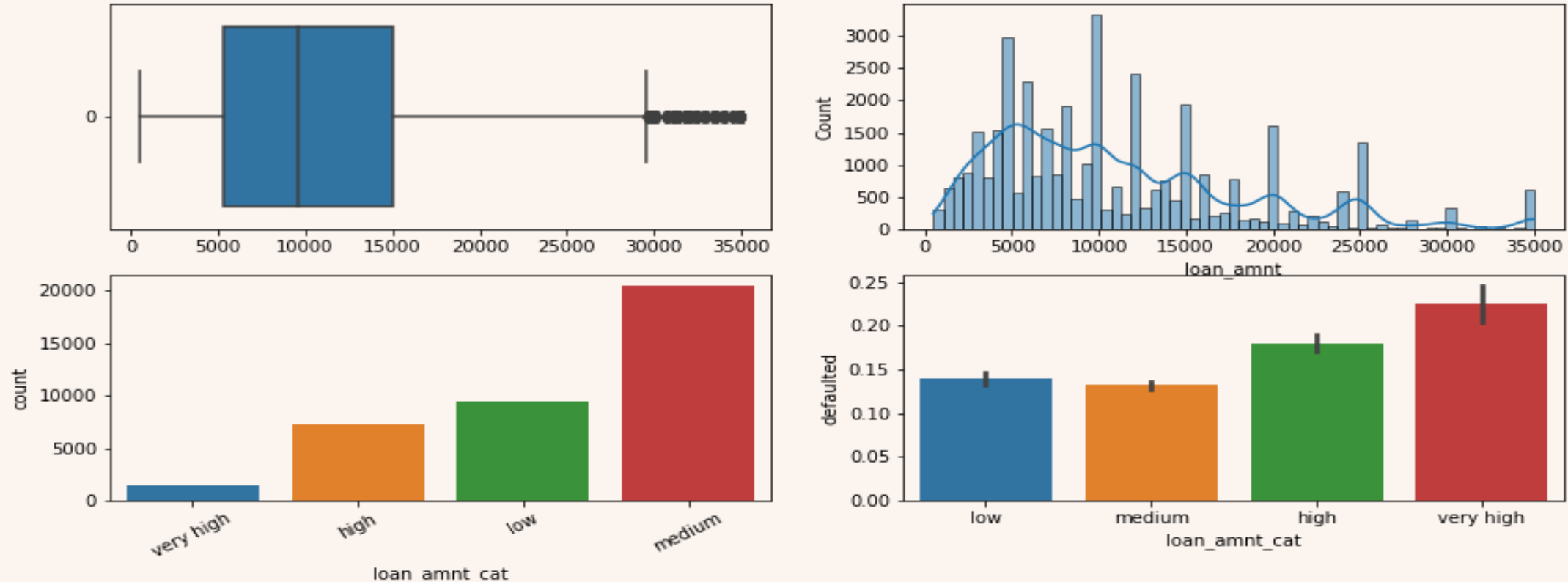
INT_RATE-ANALYSIS



- ❖ Though there are outliers seen, but outlier detection for int_rate would not make any sense, because the int_rate can vary from person to person, and from loan to loan
- ❖ If the interest is high the chances that the customer labelled as defaulter would also be high

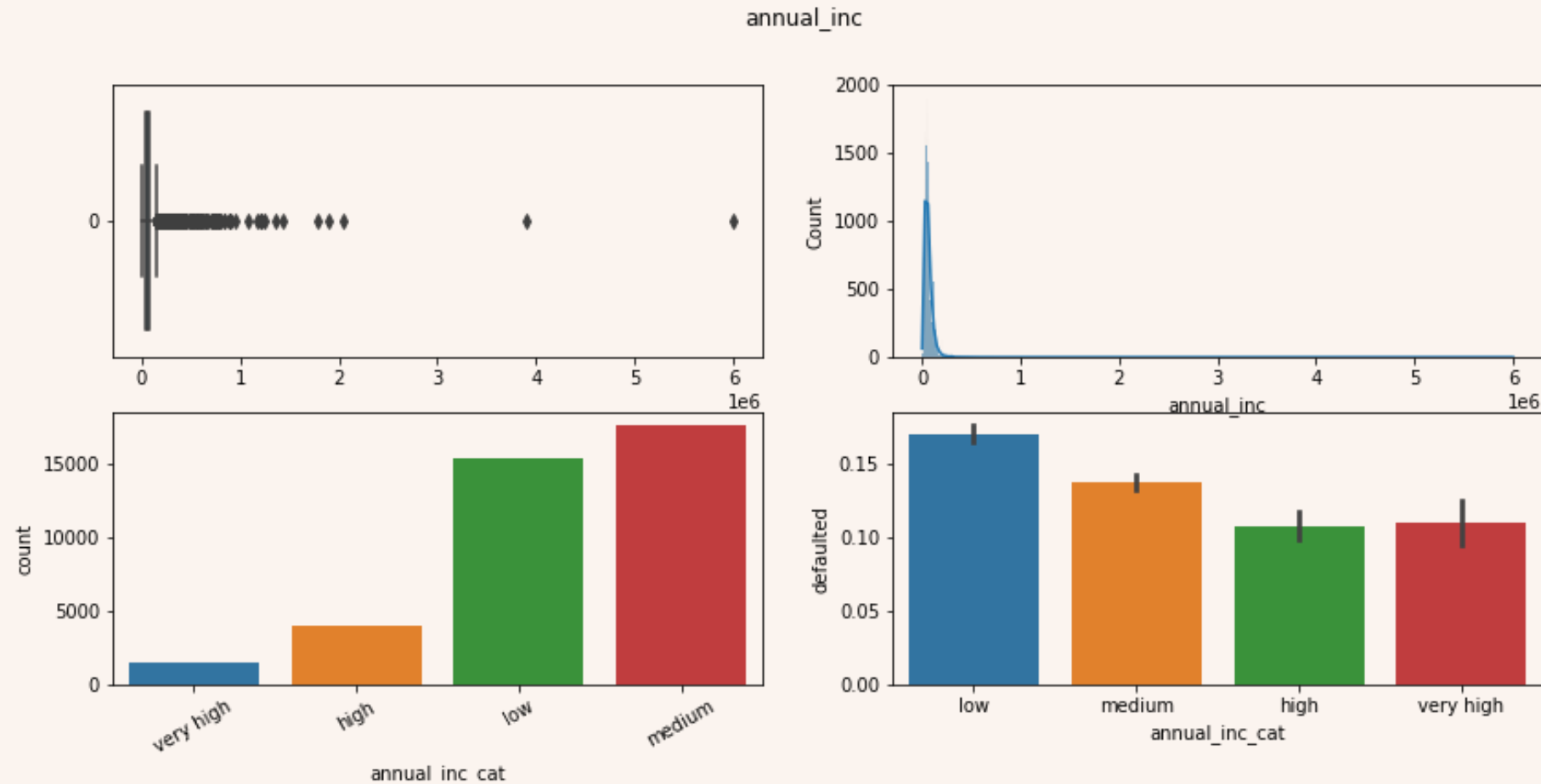
LOAN_AMNT-ANALYSIS

loan_amnt



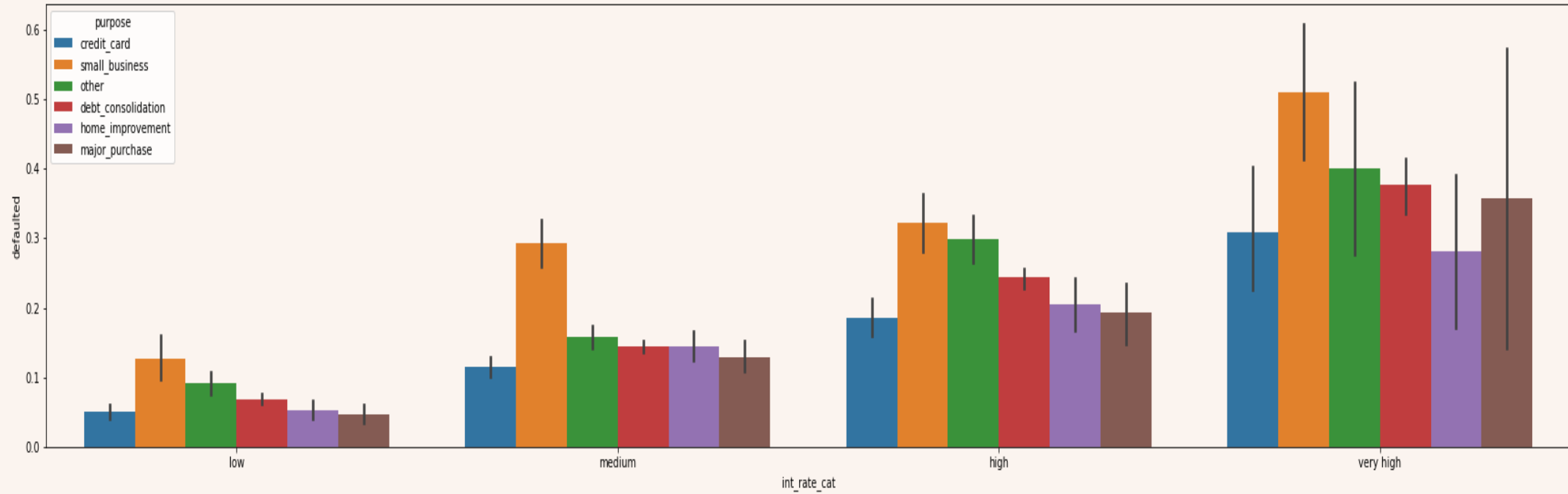
- ❖ For the given dataset the number of customers who borrowed loan amount between the range of 5000 - 15000 is high
- ❖ Borrowers who have taken high loan amount, i.e greater than 25000 have more chances of getting away without paying the loan

ANNUAL_INCOME-ANALYSIS



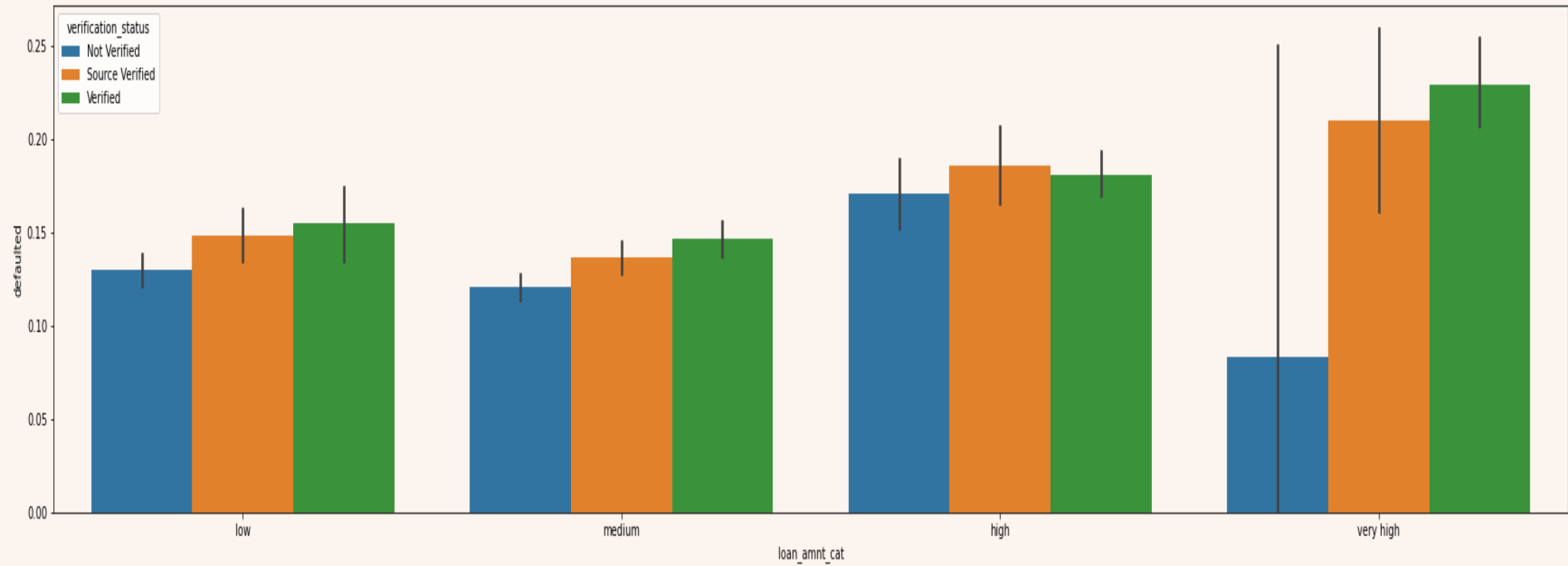
- ❖ For the given dataset the number of customers whose annual income is high and very high tend to not become defaulters.

INT RATE W.R.T PURPOSE



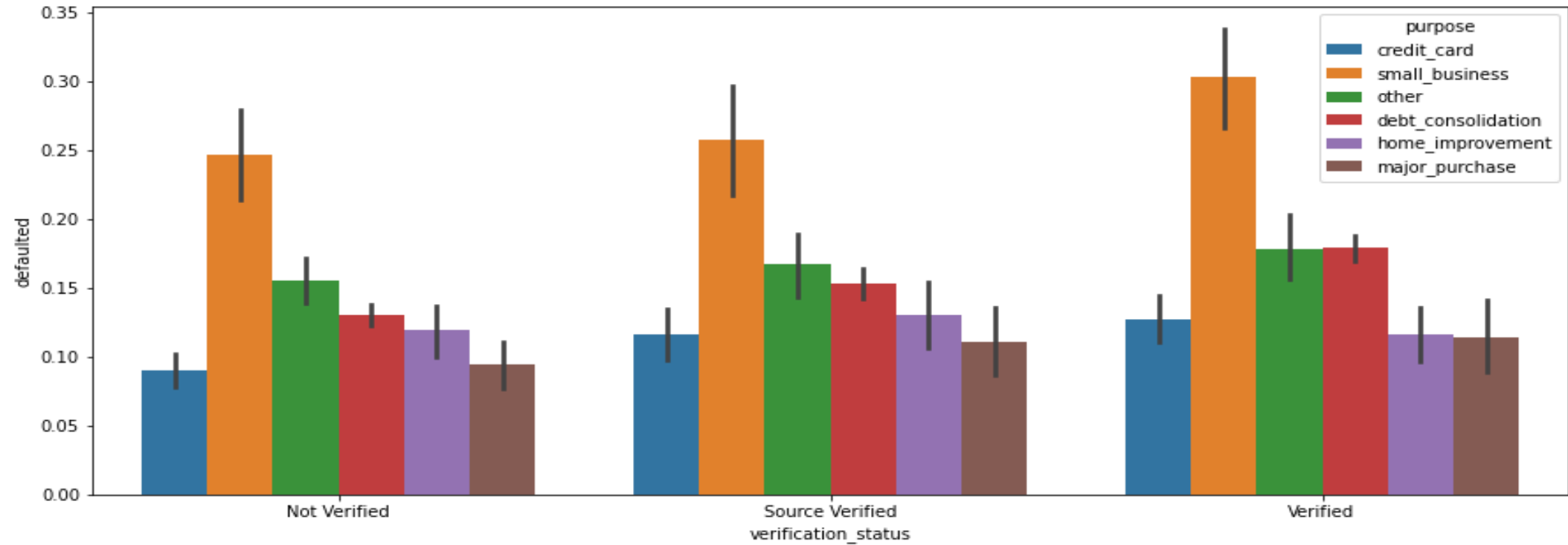
In every interest rate category, borrowers who have taken loan for small business tend to default

LOAN AMOUNT W.R.T VERIFICATION STATUS



In every loan amount segment, borrowers whose background verification is done tend to default.

VERIFICATION STATUS W.R.T PURPOSE



For any verification status, borrowers who took loan for small business tend to default.

RECOMMENDATIONS

- ❖ Before lending loans to small business purpose, LC needs to have additional checks to avoid financial loss
- ❖ Loan applicants verification needs to be thoroughly processed, as verified borrowers are more defaulters than other
- ❖ If DTI is available readily, then LC can consider it as a key driver
- ❖ In regions like Nebraska (NE) and other smaller states, LC needs to establish strong footprint



THANK YOU