

# LENDING CLUB CASE STUDY

## SUBMISSION

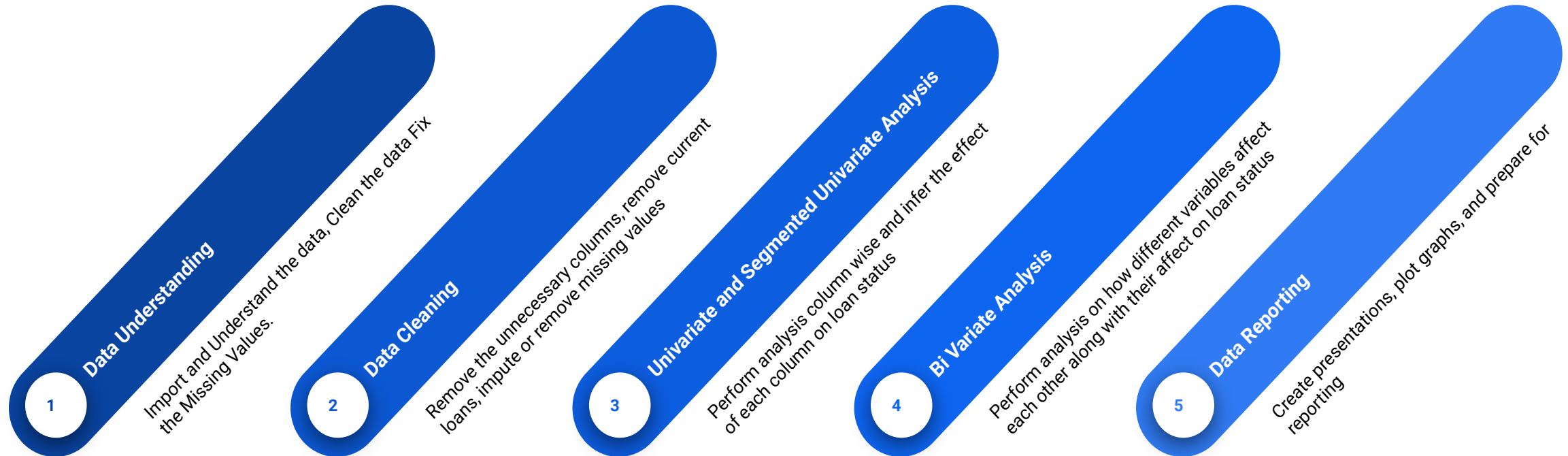
**Kotapati Venkata Vineel Reddy**

# Objectives

The main aim is to identify predictors of default so that at the time of loan application, we can use those variables for approval/rejection of the loan.

- We need to minimise two types of business risks for the lending company
  - 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
- Use EDA to understand how consumer attributes and loan attributes influence the tendency of default.
- Loan Lending Organizations may choose to utilize this knowledge for its portfolio and risk assessment of new loan applicants

# Problem solving methodology



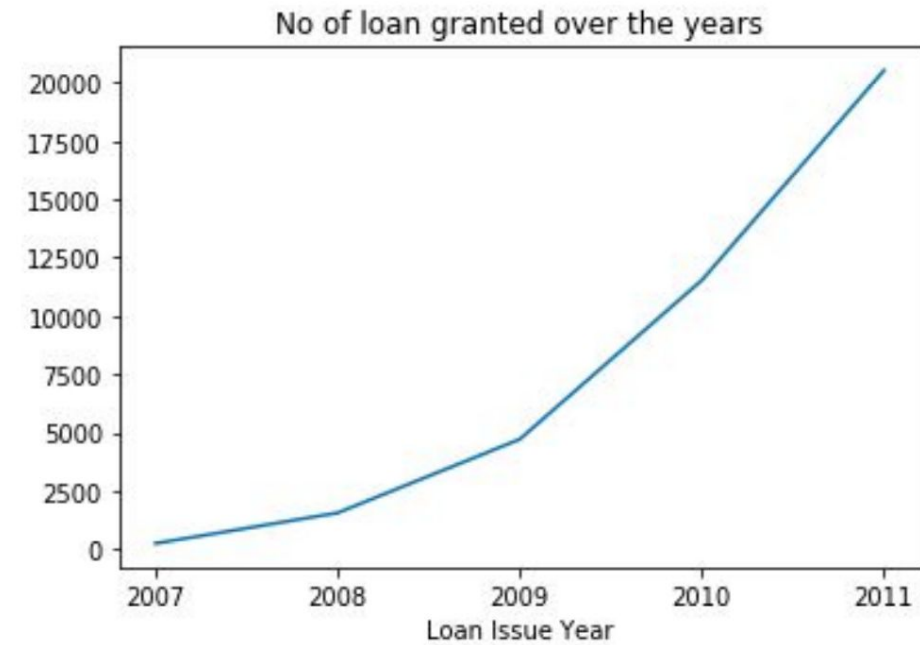
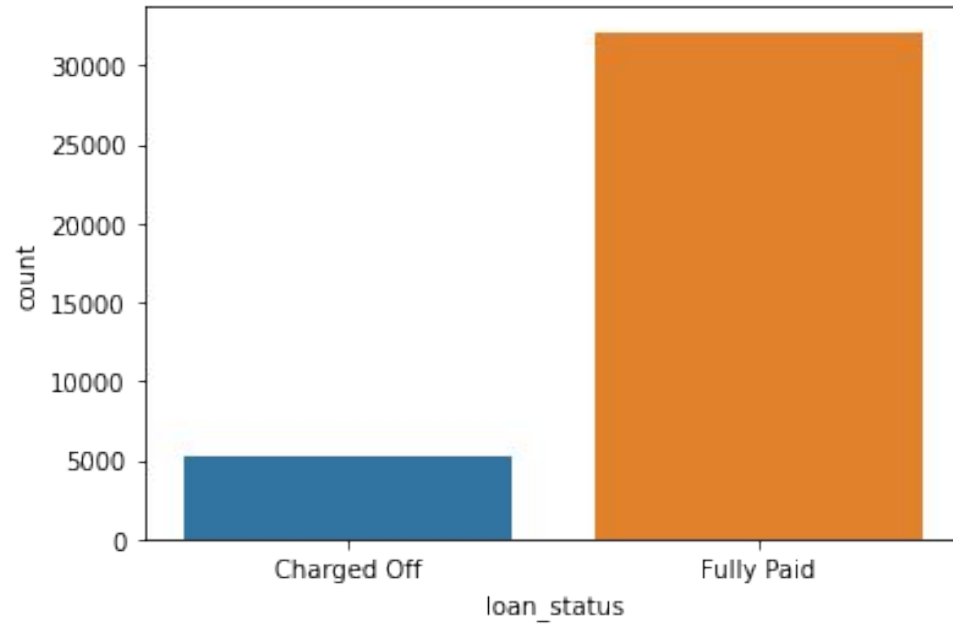
# Data Understanding And Data Cleaning

- Delete columns with a lot of missing values, only one unique value.
- Treat missing values with appropriate approach (Removal, Imputation) .
- Remove Rows where the loan status is current as they are not needed.
- Filter the columns to remove customer behaviour variables which we might not have at the time of loan approval
- Remove Columns where each column is unique(id,desc etc), change columns to appropriate types
- Create additional columns for use in segmented univariate analysis, for date time columns
- There are broadly three types of variables –
  1. those which are related to the applicant (demographic variables such as age, occupation, employment details etc.),
  2. Loan characteristics (amount of loan, interest rate, purpose of loan etc.) and
  3. Customer behavior variables (those which are generated after the loan is approved such as delinquent 2 years, revolving balance, next payment date etc.).

# Univariate, Segmented Univariate and Bivariate Analysis

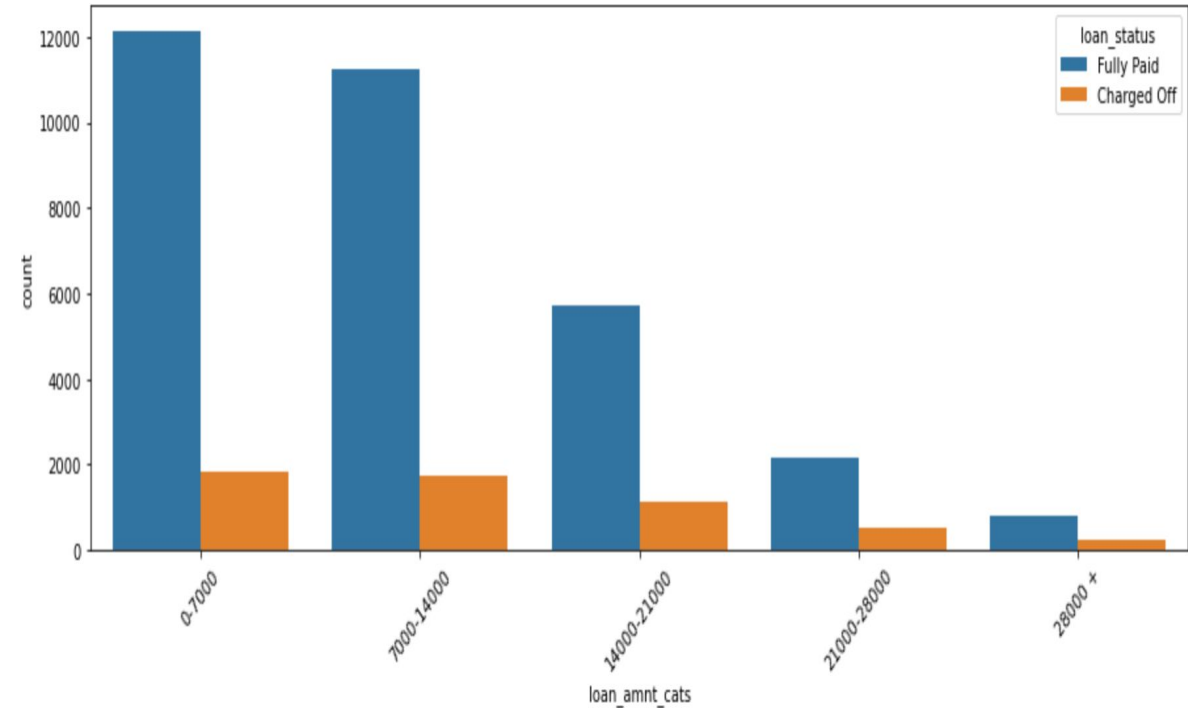
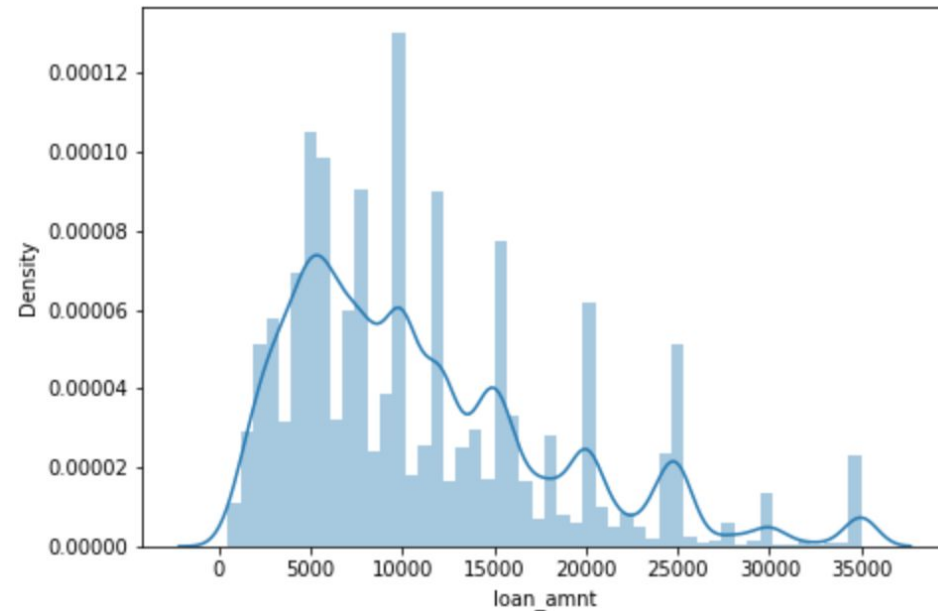
- The customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval.
- In Univariate Analysis : we have checked for the default rate across various categorical features,
- We have also performed binning on continuous features and done univariate analysis on those.
- We took the Loan status as the target variable and have checked various other features against it for analysis of defaulters.
- We have only considered use the loan amount because , the loan\_amnt is the amount applied by potential borrowers, funded\_amnt is the amount recommended/approved by Lending Club, and the funded\_amnt\_inv is the amount funded by investors.
- In bivariate analysis: we have taken two variables and more to analyse the default variable
- Loans status, loan amount, term , purpose ,grade ,annual income etc were important factors which played a big role in analysis.
- We have also done multivariate analysis using all the important variables ,which gave the below results:
  - Higher the interest rate higher charged off ratio
  - Higher the annual income higher the loan amount slightly
  - Interest rate is increasing with loan amount increase this results in high charged off.

# Loan Status



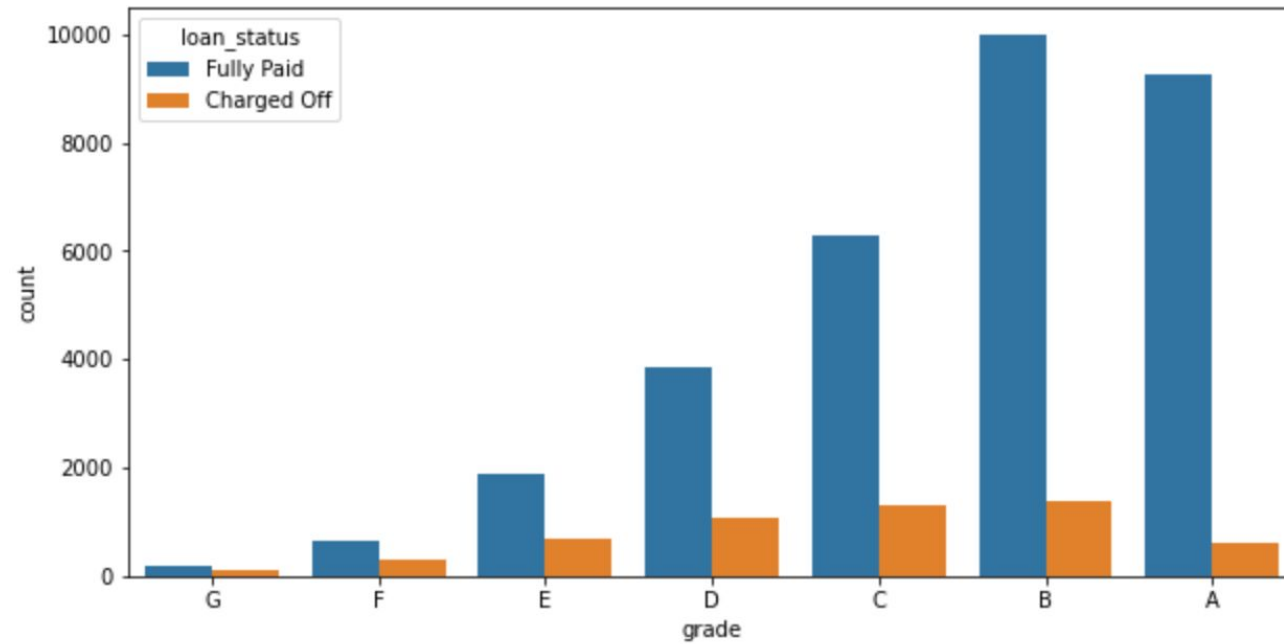
- About 14% of all the loans are Defaulted
- Number of loans granted increased over the years

# Loan Amount



- As the amount of loan increases the percentage of defaulted loans increases from 13% (0-7k) to 23% (28-35k)

## Loan Grade



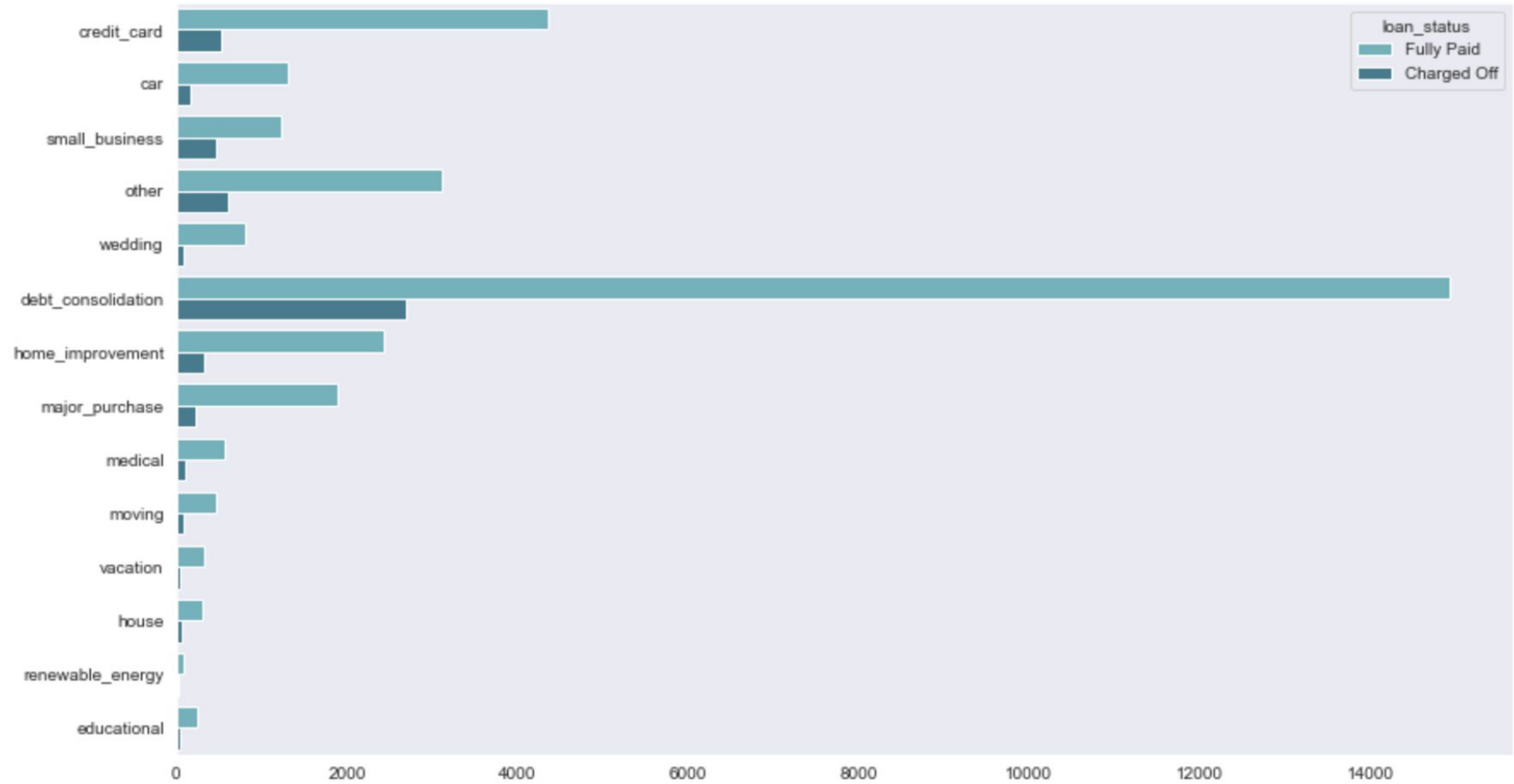
As the loan grade increases the percentage of default increases Which indicates the present grading system is functioning well



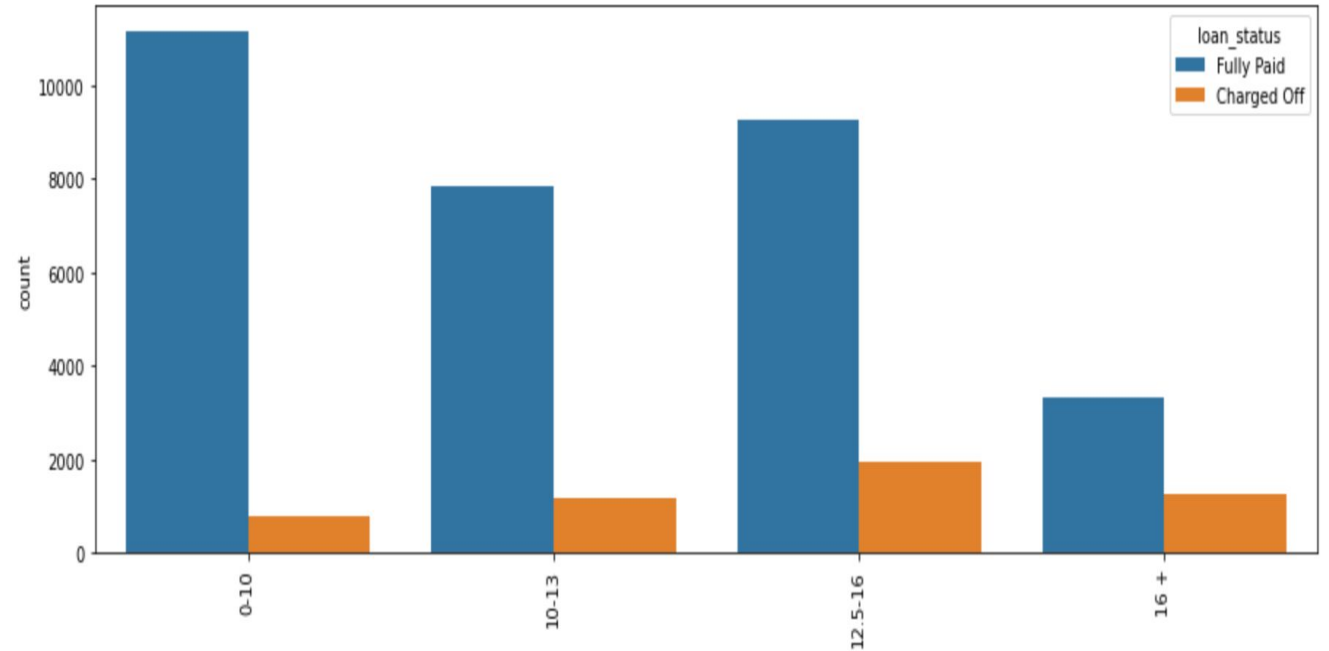
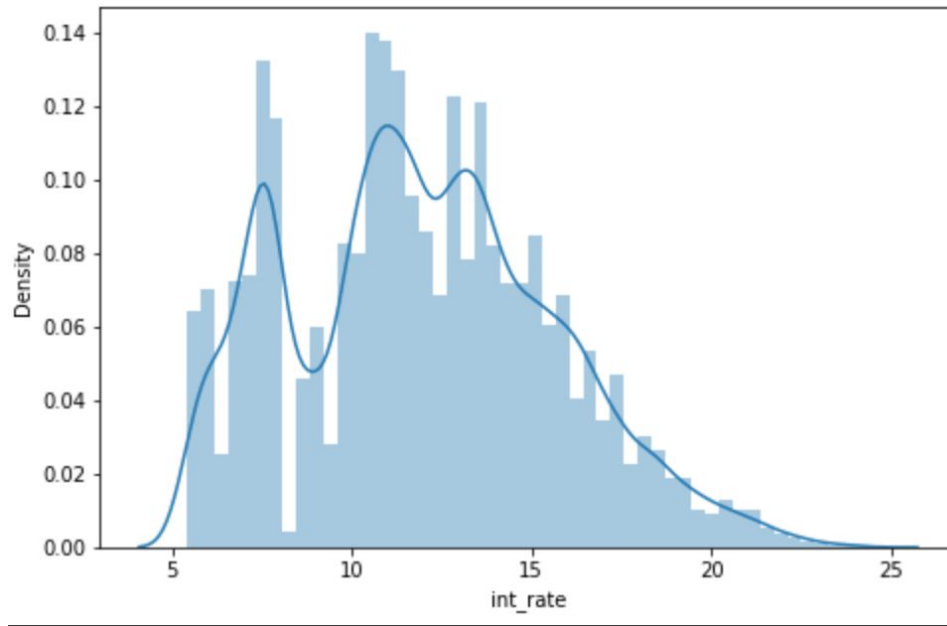
## Purpose

Debt Consolidation is the highest type of loan

Debt consolidation, Small Business are the two purposes where we see a large percentage of defaults



## Interest Rate



- As interest rate increases the percentage of defaulted loans increases

## Conclusions

In this case study we are trying to figure out the important features that contribute toward default.

### **Recommendations:**

1. Grading system of the LC is working as expected.
2. Clients belonging to state CA has high rate of default so extra scrutiny is necessary. More background checks of the loan applicants will be a good start.
3. when the purpose is debt consolidation or small business there is a high risk of default so we can either increase the interest rate or reduce the loan amount.
4. In segmented analysis of loan amount we have discovered that higher loan amount has higher risk of default. It would be a cautious step to see if the applicant has all the factors to be a good applicant.
5. Higher interest rates have higher chance of default so needs extra scrutiny.