

Lending Club Case Study- upgrad

Rakshitha BS

Problem Statement

The aim of this EDA is to help an online lending company analyse risk associated with a loan application. When a lending company receives an application, the borrower is approved or rejected based on his profile. This EDA will help to identify factors which classify a borrower to high-risk or low-risk customer. There are two types of risk associated with lending use-case:

if customer is approved and doesn't repay the loan ,i.e borrower is a defaulter then the company faces a financial loss

If customer is rejected but is likely to repay the loan, then not approving can lead to business loss

With the help of dataset the task is to identify variables or factors that determine if the borrower is likely to be a defaulter or not. The dataset consists of past loan approved loan applications and the corresponding loan status. The loan status "Charged Off" here refers to defaulted loan.

About the Dataset

.The dataset contains the complete loan data for all loans issued through the time period 2007 to 2011. The column `loan_status` determines if the status of the loan is current, fully paid(non-defaulter), or charged off (defaulter).

Data Cleaning

Handling Missing Values

- The dataset has a lot of null values. First task would be to handle null values.
- Columns with all null values are removed
- Columns $> 50\%$ null values are removed
- Columns with lesser % of null values are filled with 0

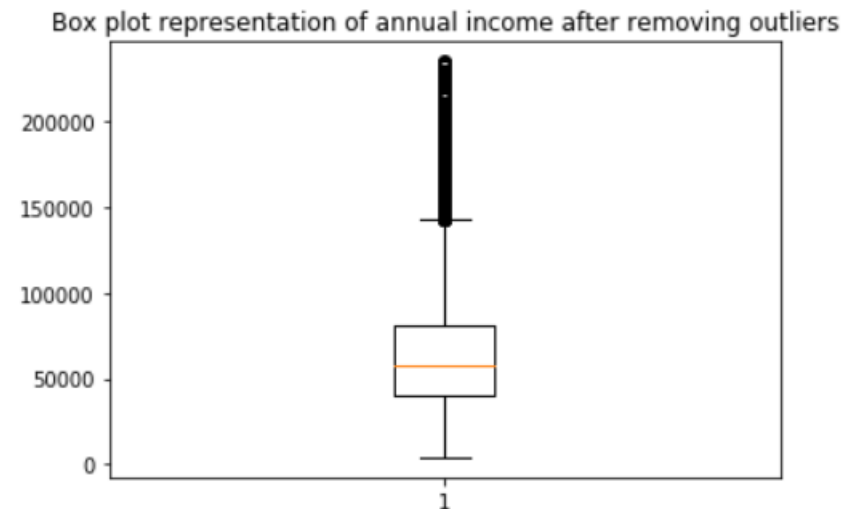
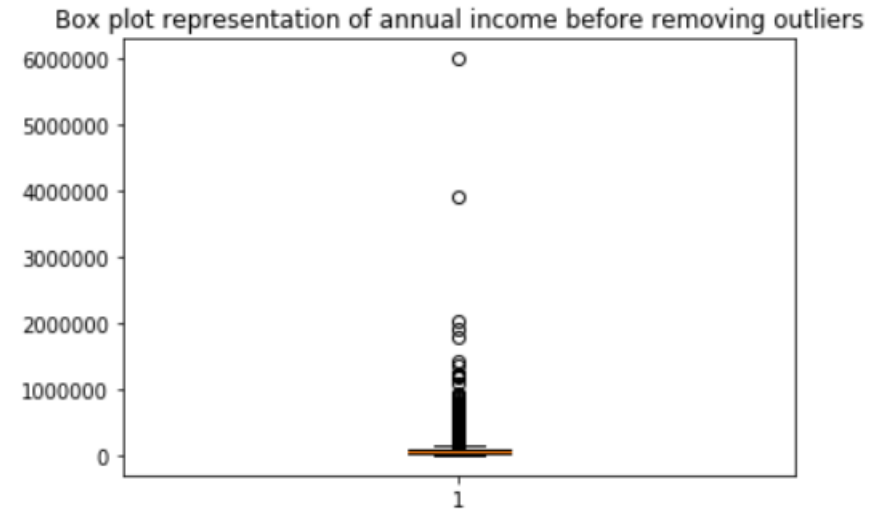
Data Cleaning

Standardizing Data type of columns

Term and int_rate are columns are standardized to int and float.

Handling Outliers

- Annual income seems to have a lot of outliers. The plot is shown on the top right .
- We can keep values lesser than 99 percentile.
- Box plot after removing outliers is shown on bottom right.



Data Cleaning

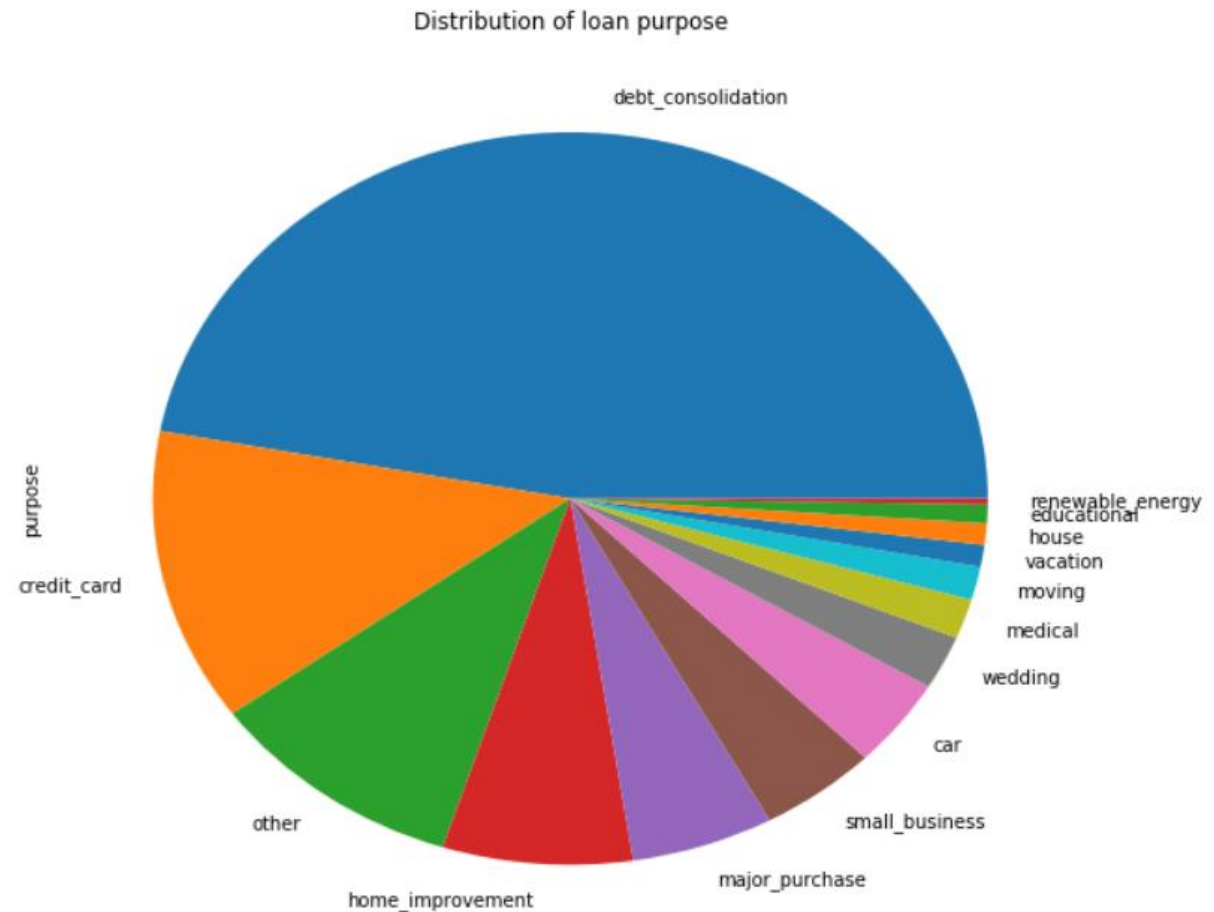
Creating Categorical variables from continuous variables

1. **Interest rate bucket :** '0-5%', '5-10%', '10-15%', '15-20%', '20-25%'
2. **Annual income bucket:** '0-25k', '25-50k', '50-75k', '75-100k', '100-125k', '125-150k'
3. **Loan Amount bucket:** '0-7k', '8-14k', '15-21k', '21-28k', '29-35k'

Univariate Analysis

Distribution of Loan Purpose in the dataset

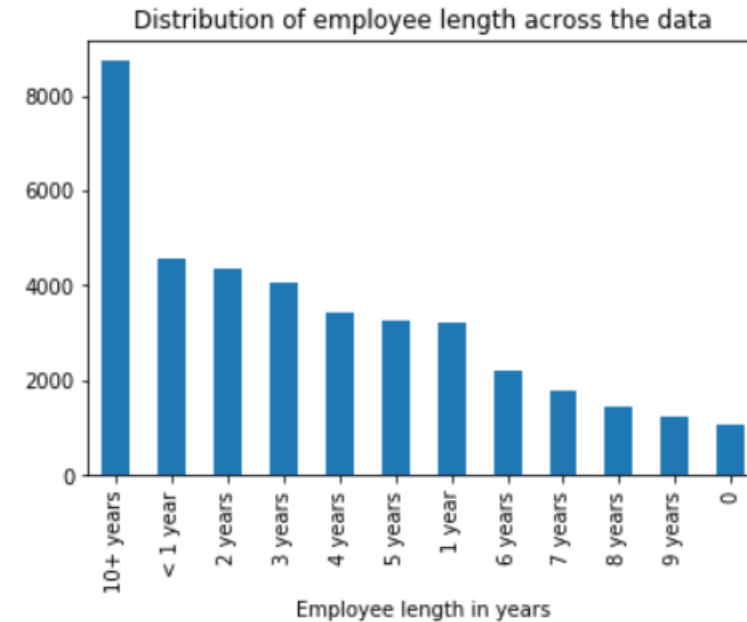
- Most of borrowers take the loan for debt consolidation or for credit cards



Univariate Analysis

Distribution of employee length across the data

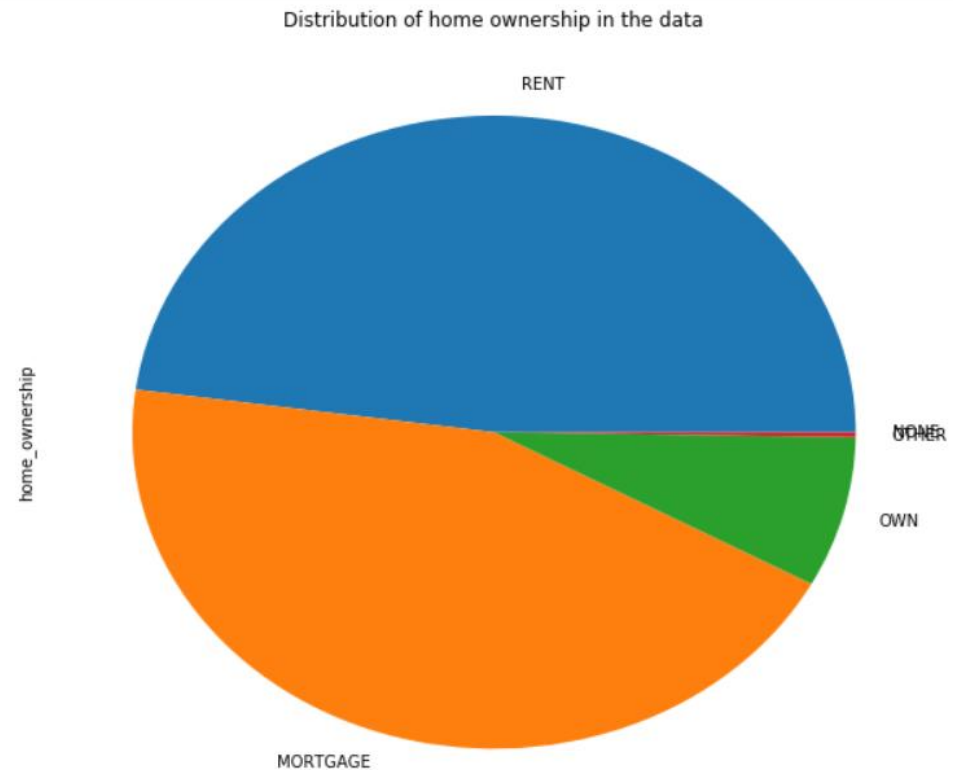
- Borrowers with more than 10 years experience have taken the most loans



Univariate Analysis

Distribution of home ownership in the data

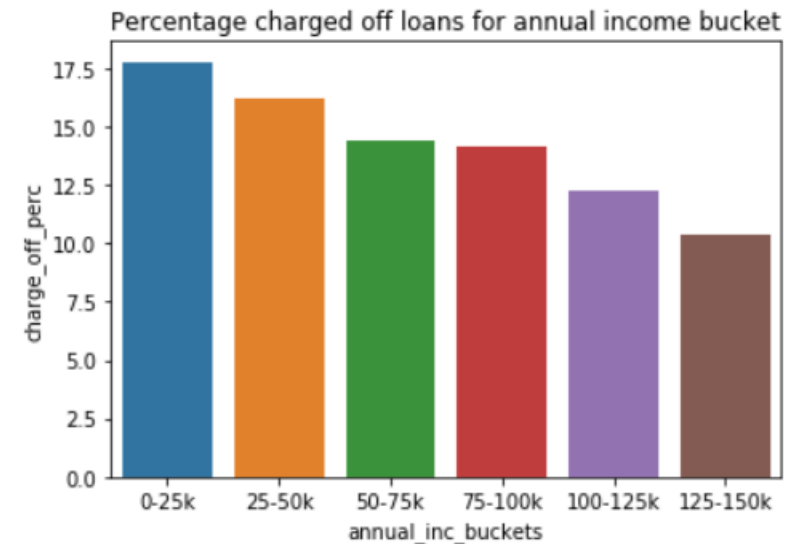
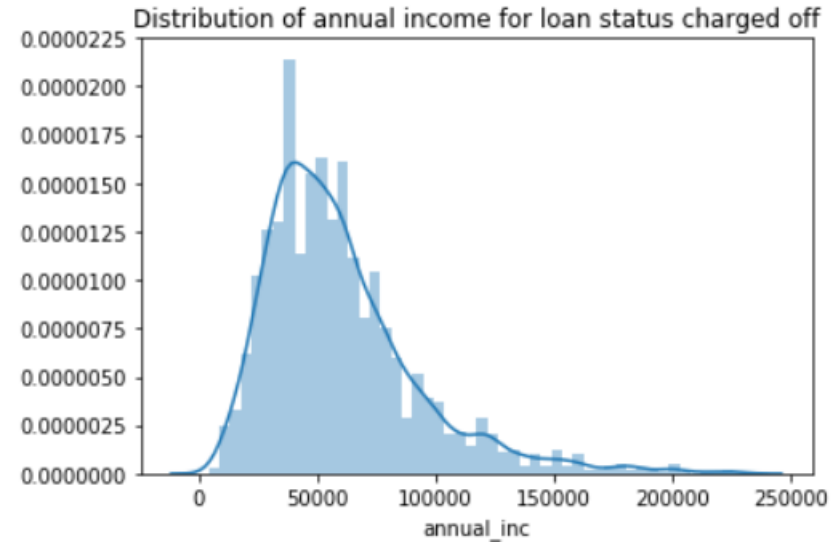
- Majority of the loans are borrowed by customers with rented homes.



Segmented Univariate Analysis

Distribution of annual income for loan status charged off

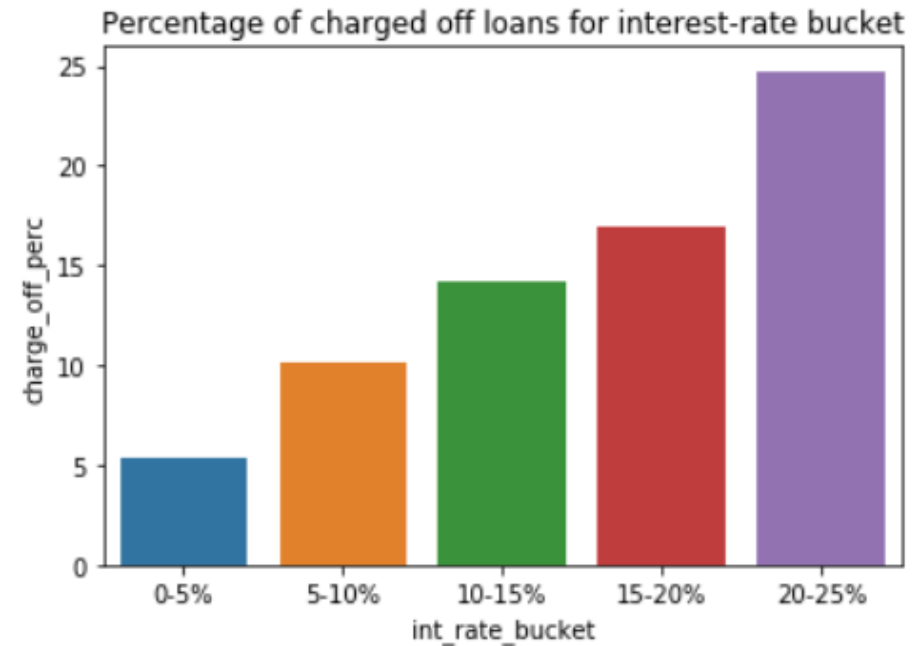
- Borrowers with lower annual income are more likely to charge off.
- By calculating charge off percentage for each annual income bucket, it is observed that lowest annual income bucket(0-25k) has higher chances of charge off.



Segmented Univariate Analysis

Distribution of charged off loans for each interest rate bucket

- Customers with high interest rates are more likely to be charged off



Segmented Univariate Analysis

Employee length for charged off loans

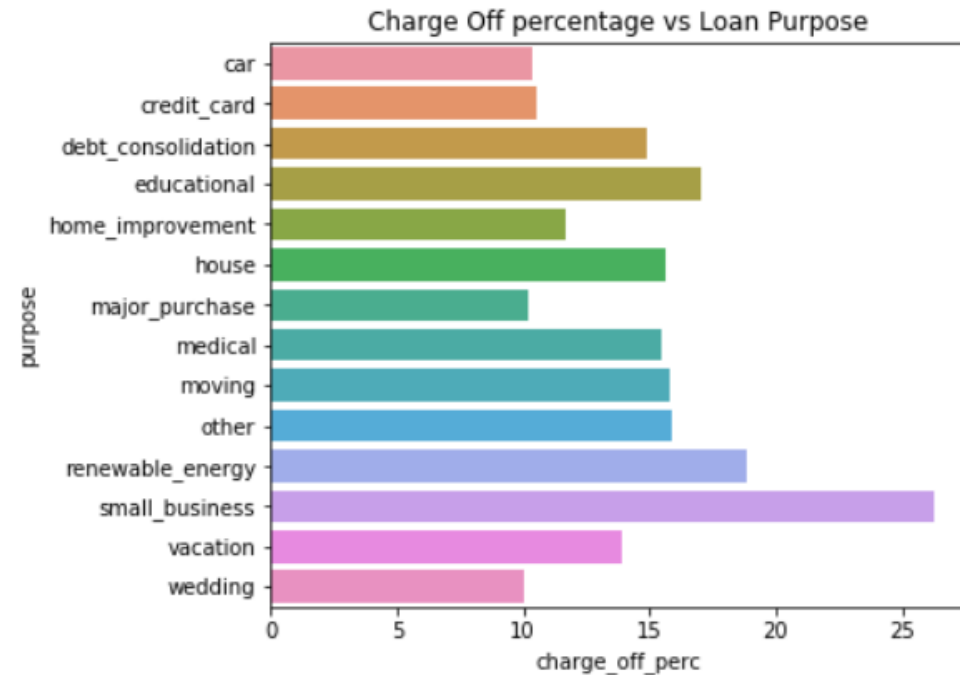
- Customers with 0,<1year,1 year experience are more likely to be charged off



Segmented Univariate Analysis

Distribution of loan purpose for charged off loans

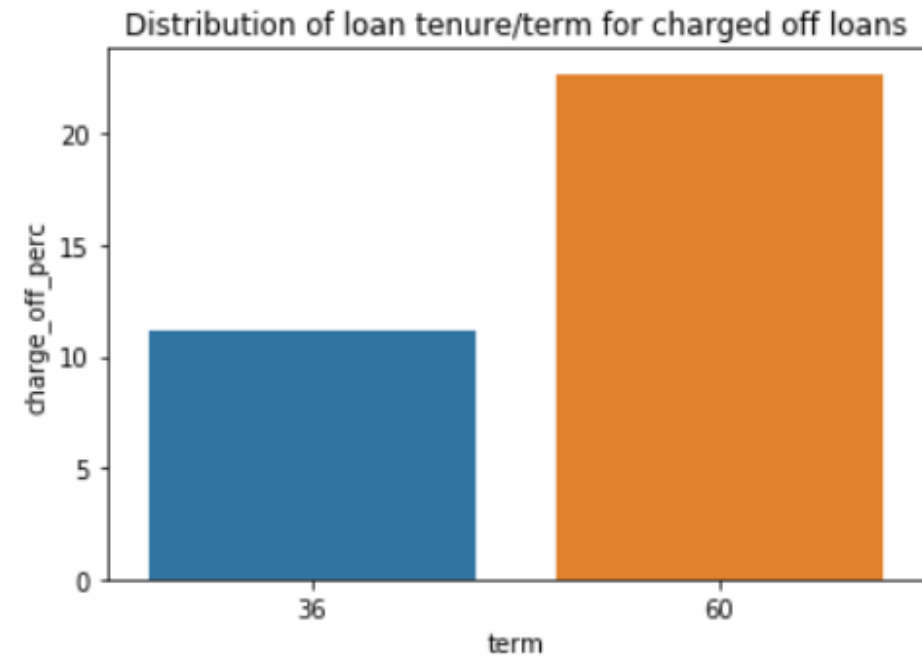
- Loans with purpose as small business are more likely to be charged off



Segmented Univariate Analysis

Distribution of loan tenure/term for charged off loans

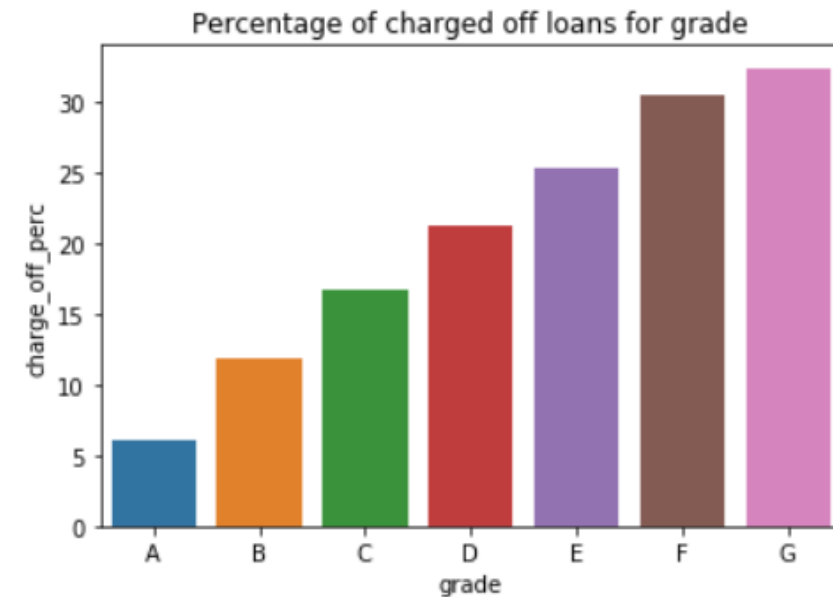
- loans with more no of installments are more likely to default



Segmented Univariate Analysis

Distribution of grade of loan for charged off loans

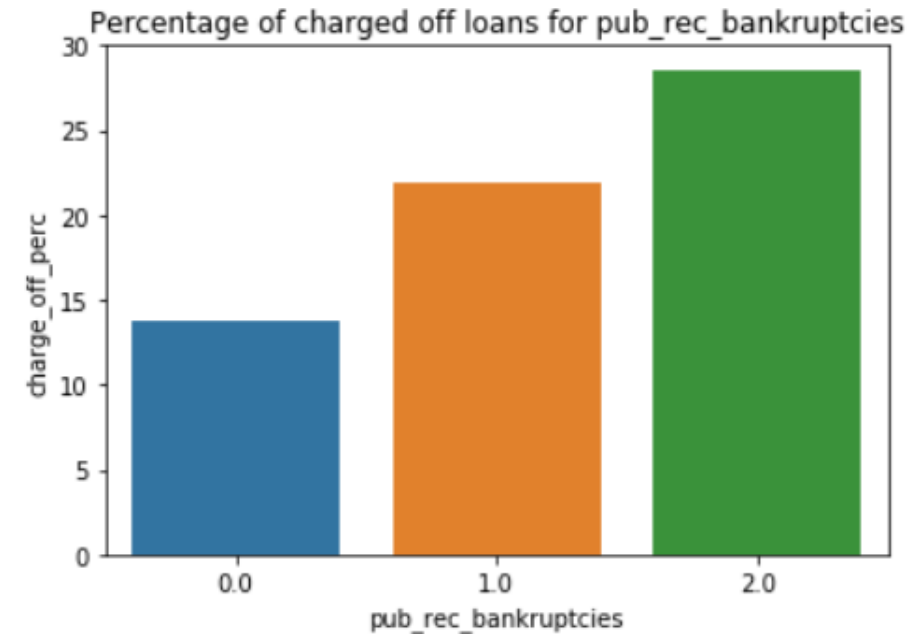
- Loans with grade G are more likely to be charged off



Segmented Univariate Analysis

Public record bankruptcies for charged off loan

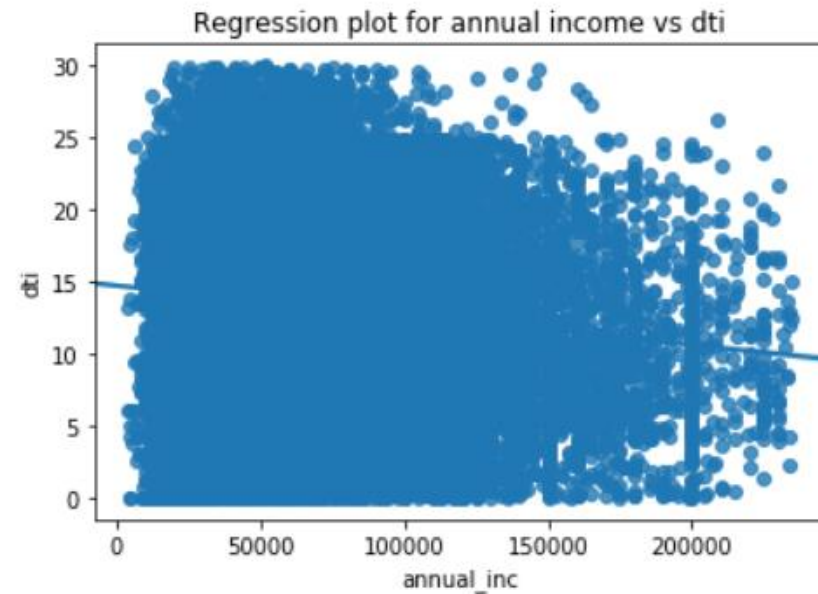
- Customers having more number of public record bankruptcies are more likely to default



Bivariate analysis

Correlation between dti and annual income

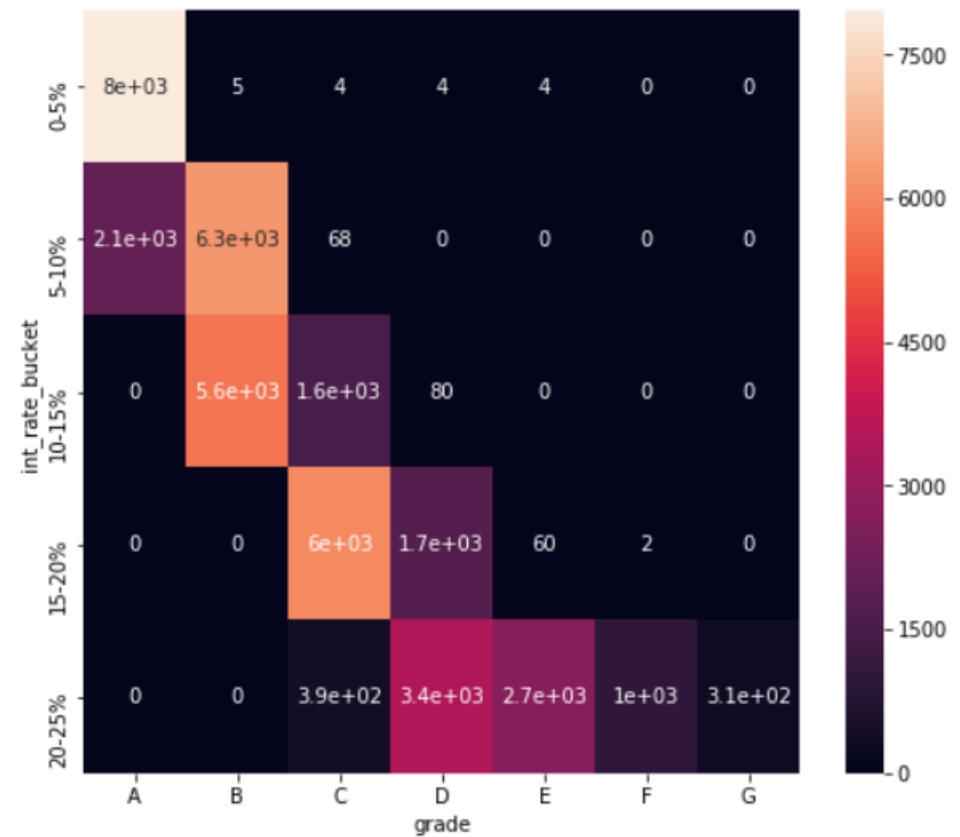
- Annual income has negative correlation with dti, i.e if annual income is low, dti is high
- Dti here is the debt to income ratio.



Bivariate analysis

Relationship between grade and interest rate

- Grade A has low interest rate and grade G has high interest rate



Conclusion

- Borrowers with low income are more likely to be defaulted
- Loan Amounts with higher interest rates have more chances to be defaulted
- Borrowers with few years of work experience have more chances to be defaulted
- Loans with higher tenure/ number of installments have more chances of defaulting
- Low annual income have higher dti(debt to incom ratio)



THANK
YOU!

The image features the words "THANK YOU!" arranged in two rows. Each letter and the exclamation mark are printed in a bold, black, sans-serif font on individual yellow rectangular sticky notes. The notes are slightly crumpled at the corners, giving them a realistic appearance. Each note is held in place by a small blue pushpin at its top edge. The entire arrangement is set against a plain white background.