Lending Club Case Study

Evaluation of Loan data to find out possible Defaulters

Prepared by
Mamta Bhat & Jignesh

06/03/2024

Objective and Scope

The objective of this case study is to analyse the given loan data and provide recommendations to decrease the number of defaulters for Lending Club

- Identify patterns and trends to decrease the number of defaulters.
- Analyze various data points to gain insights into the factors that contribute to loan default
- Make informed recommendations to mitigate the risk of default

Data Analysis

We have carried out the following procedures and methods to analyse the data and figure out the possible trends and patterns of defaulters

1. Data cleaning:
    - Fix rows and columns
    - Fix missing values
    - Standardise values
    - Fix invalid values
    - Filter data
    - Outlier treatment
2. Univariate analysis
    - Unordered Categorical Variables
    - Ordered Categorical Variables
    - Quantitative Variables
    - Segmented Univariate Analysis
3. Bivariate analysis
    - continuous variables
    - categorical variables
4. Deriving metrics

Target Variable = “loan_status”= “charged_off”

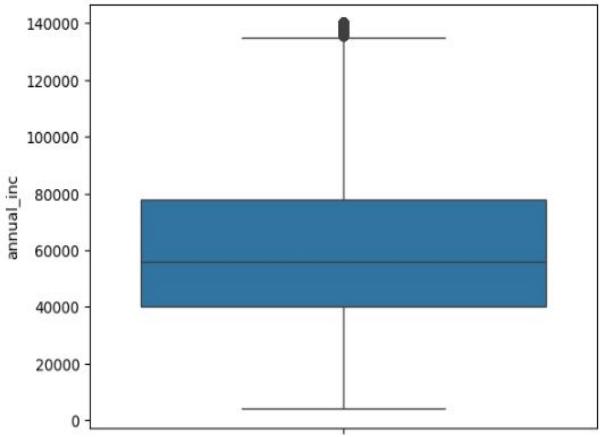
First set of Inference from Treated Data

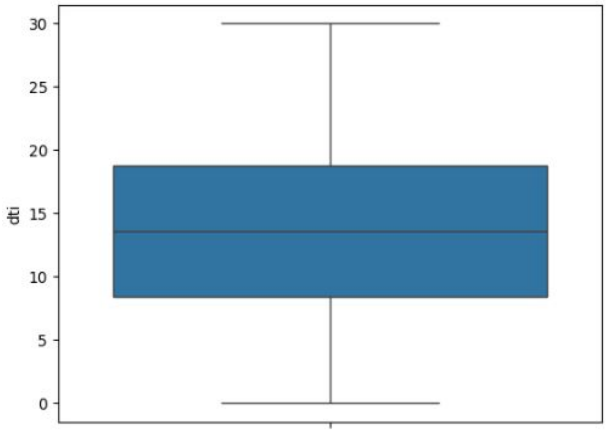
- As per data there are **32950** borrowers who have fully paid the loan
- **5627** borrowers have defaulted
- **1140** applicants are still repaying
- Over all **14.78%** borrowers have defaulted

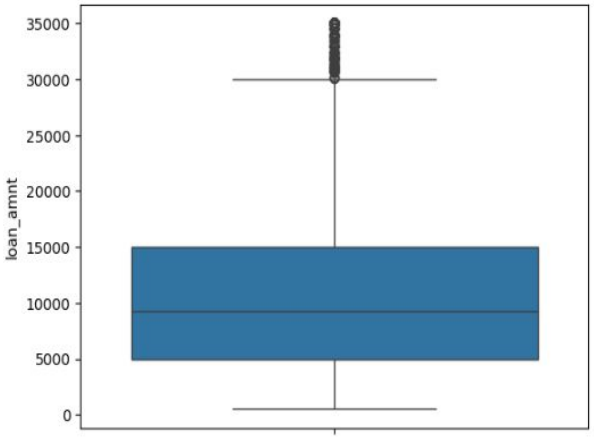
Conclusion from the box plots:

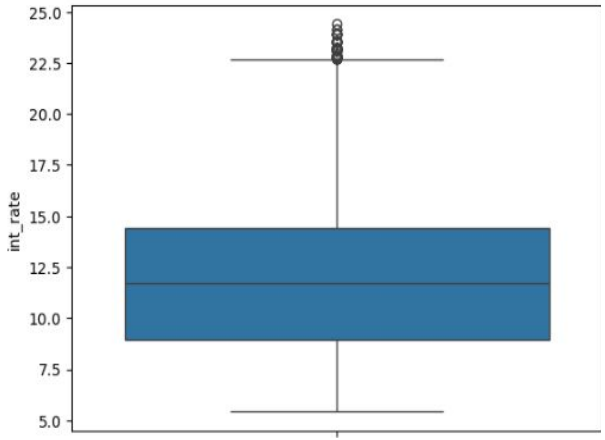
- The **annual income** of majority of the loan applicants is between **40000 - 77000 USD**
- The **loan amount** of majority of the loan applicants is between **5000 - 15000**
- The **funded amount** by investor for majority of the loan applicants is between **4900 - 13600**
- Most offered **interest rate** on the loan is between **9% - 14%**
- The majority of **term** of emi for repayment is between **163 - 408**
- The **debt to income ratio** is between **8.35 - 18.71**

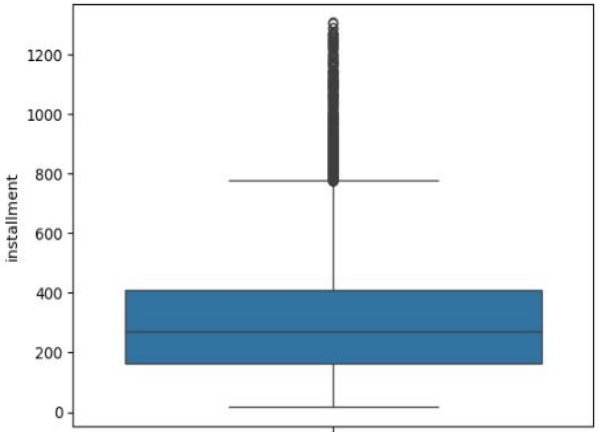
Box Plots

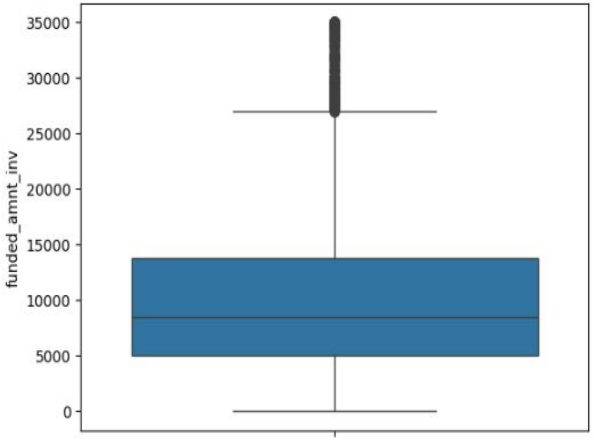

annual_inc
140000
120000
100000
80000
60000
40000
20000
0


dti
30
25
20
15
10
5
0


loan_amnt
35000
30000
25000
20000
15000
10000
5000
0


int_rate
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0


installment
1200
1000
800
600
400
200
0


funded_amnt_inv
35000
30000
25000
20000
15000
10000
5000
0

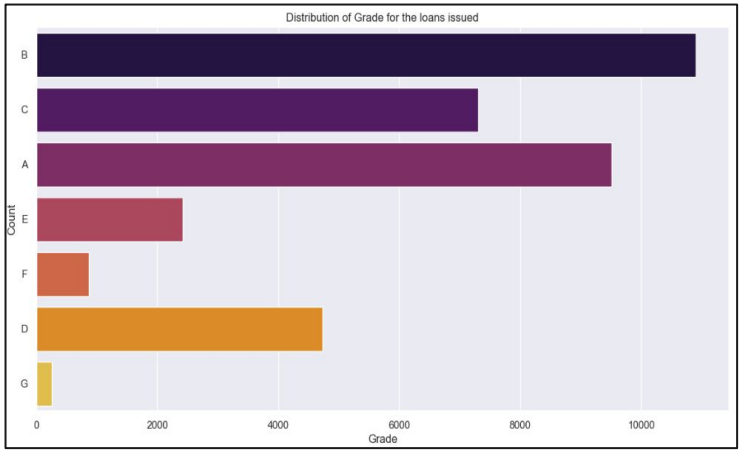
Exploratory Data Analysis

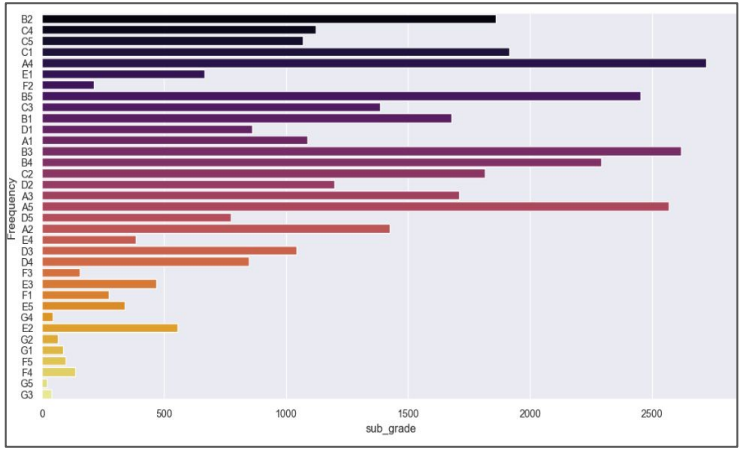
Univariate Analysis -Ordered Categorical Variables

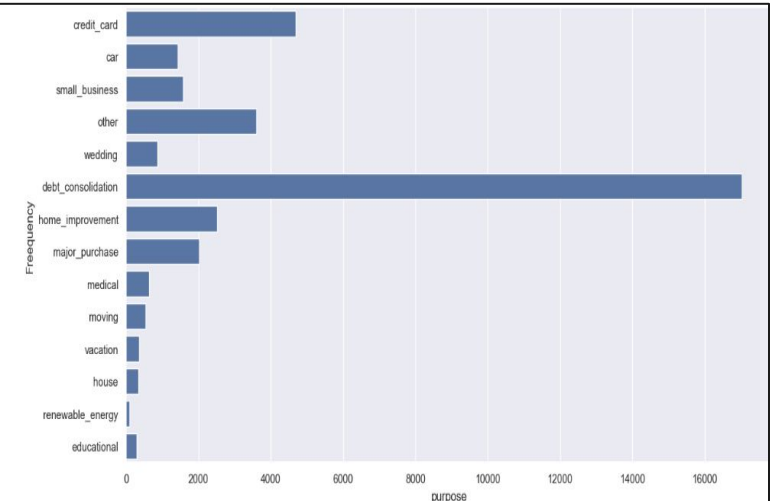
From the univariate analysis we could draw certain observations. The majority of the loan applicants belonged to following categories:

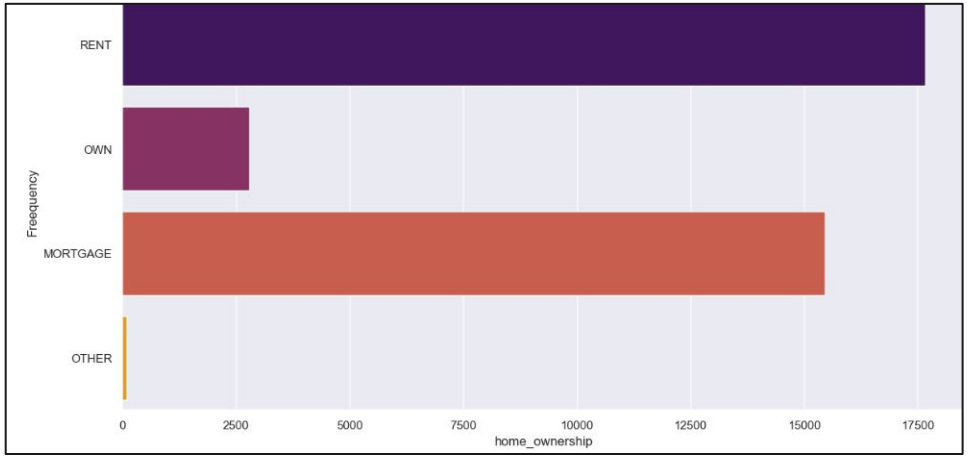
- Loans with a duration of **36 months**; 27063 applicants
- Employment length of **10+ years**
- People who stay in a **rented house**: about 17500 applicants
- Applicants with **Grade B**: frequency more than 10000.
- Subgrades **A4**: about 2723 applicants from this subgrade
- Purpose of loan is **debt consolidation**: more than 16000 applicants

Univariate Analysis -Ordered Categorical Variables

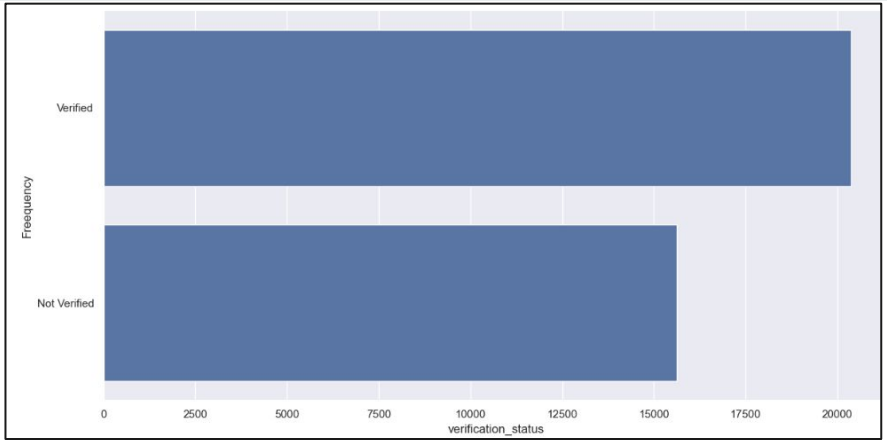


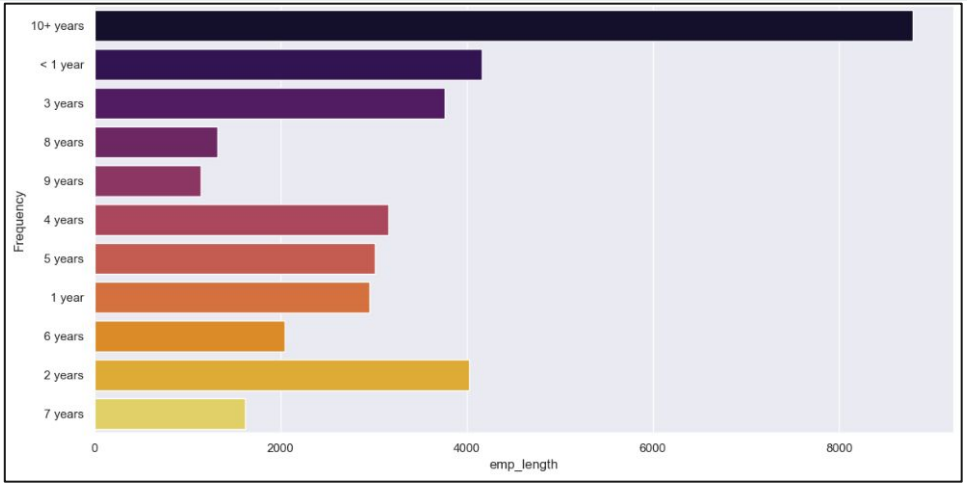


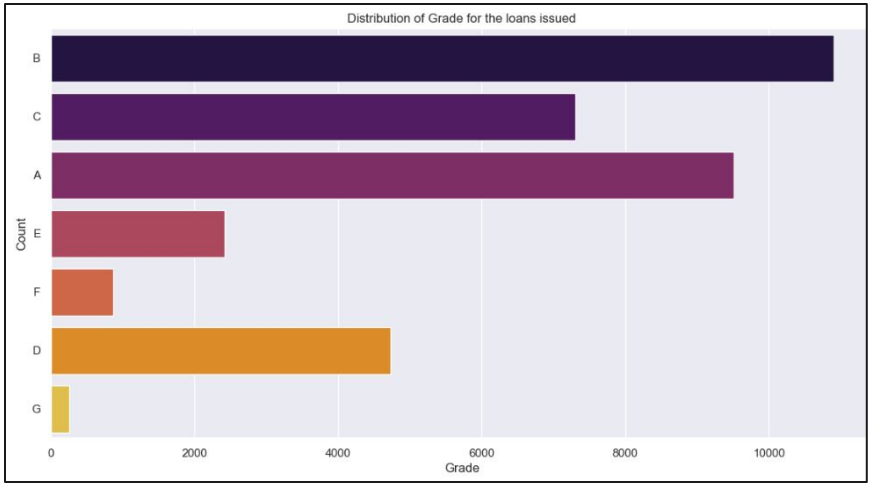


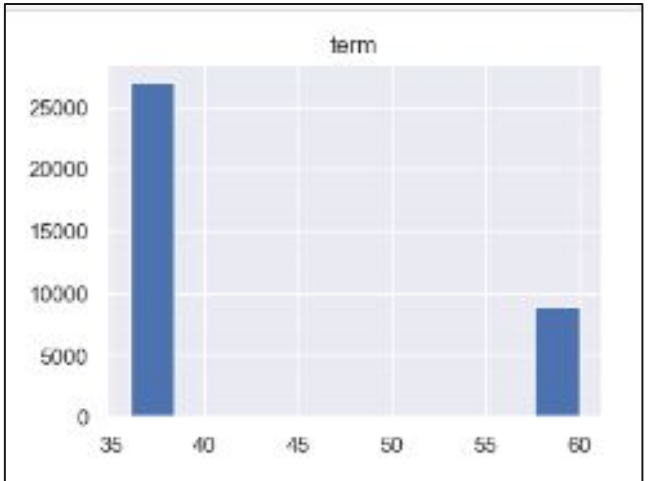


Univariate Analysis -Ordered Categorical Variables









Bivariate Analysis of Quantitative variables

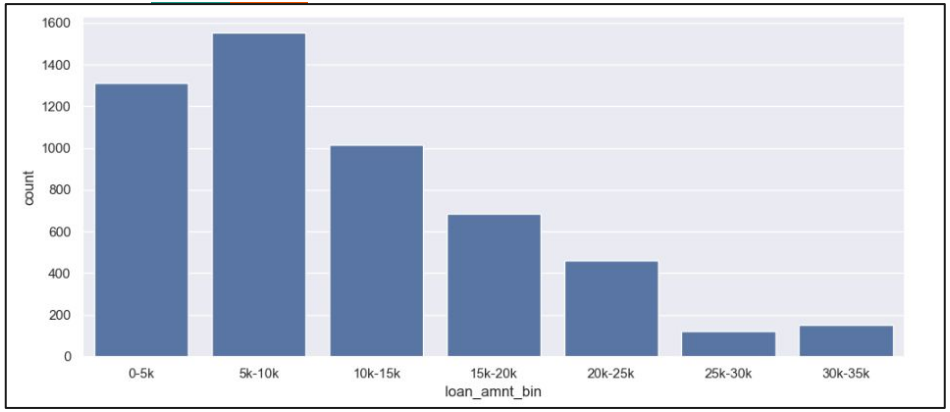
When analysed quantitative variables with respect to the **charged off** loan, there is a more probability of defaulting when :

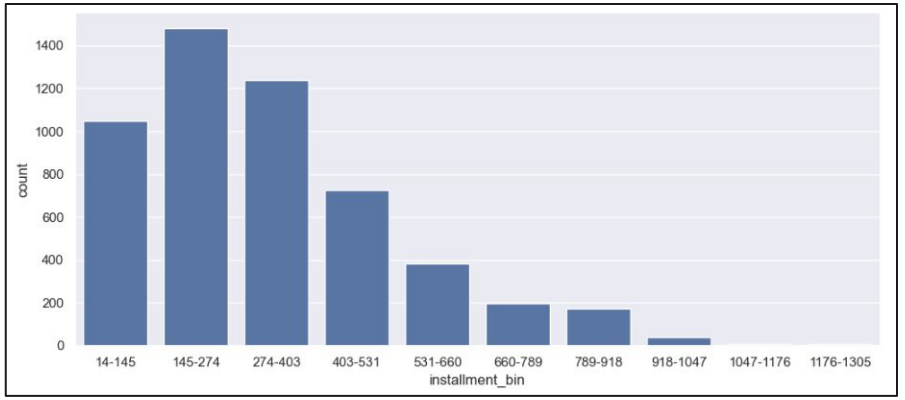
- Applicants who receive **interest** at the rate of **10-15%**
- Applicants who have an **income** of range **31201 - 58402**
- **Loan amount** is between **5429 - 10357: majority is about 10000 loan amount**
- When **funded amount** by investor is between **5000-10000**
- **DTI** is between **12-18** ; 18%-24% also has a higher number of defaulters
- When the number of **installments** are between **145-274**
- Also there is a very interesting observation from the date issued. The late months of an year indicated the high possibility of defaulting. Especially in **December**
- The high number of loan defaults in **2011** could be due to the financial crisis in USA (Assuming the data is of US origin) -more than 3000 applicants defaulted in 2011 and more than 600 of them have defaulted in the month of december

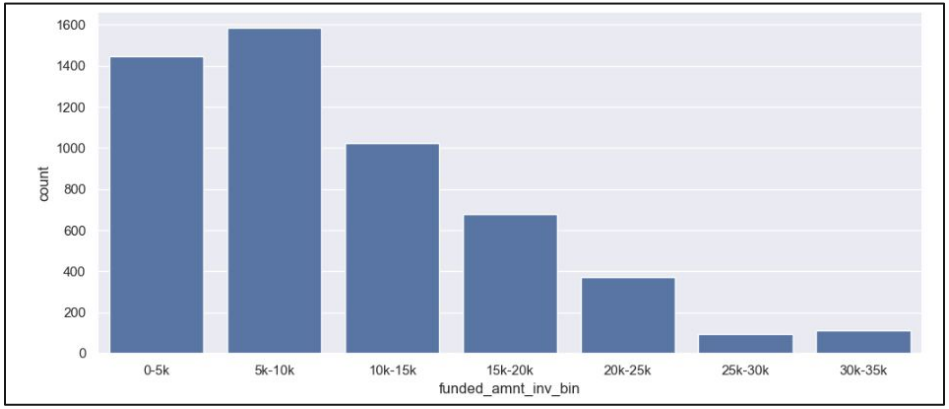
Loan Amount Analysis

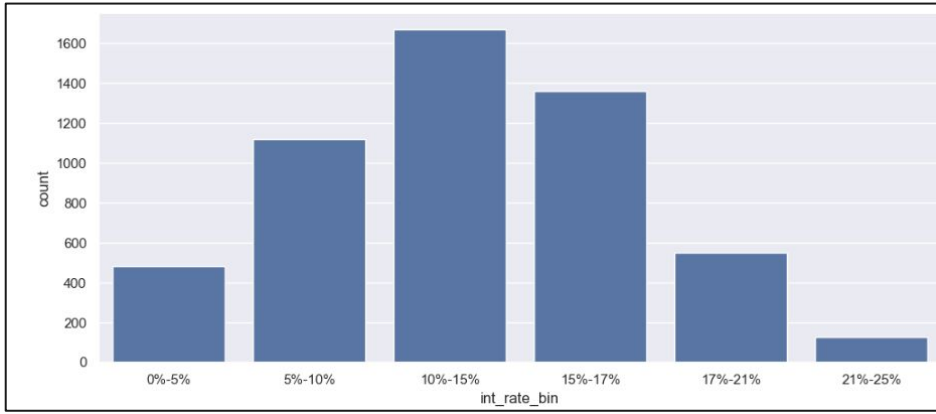
- The majority of loan defaulters fall within the **$5,000 - $15,000** range.
- Default rates decrease as the loan amount increases beyond **$15,000**.
- Loans less than $5,000 and more than $25,000 have relatively lower default rates.

Bivariate Analysis of Quantitative variables - Plots

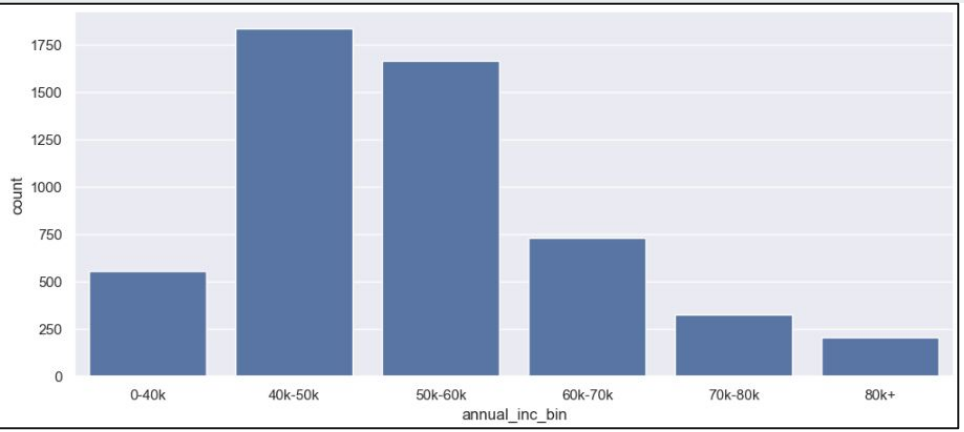


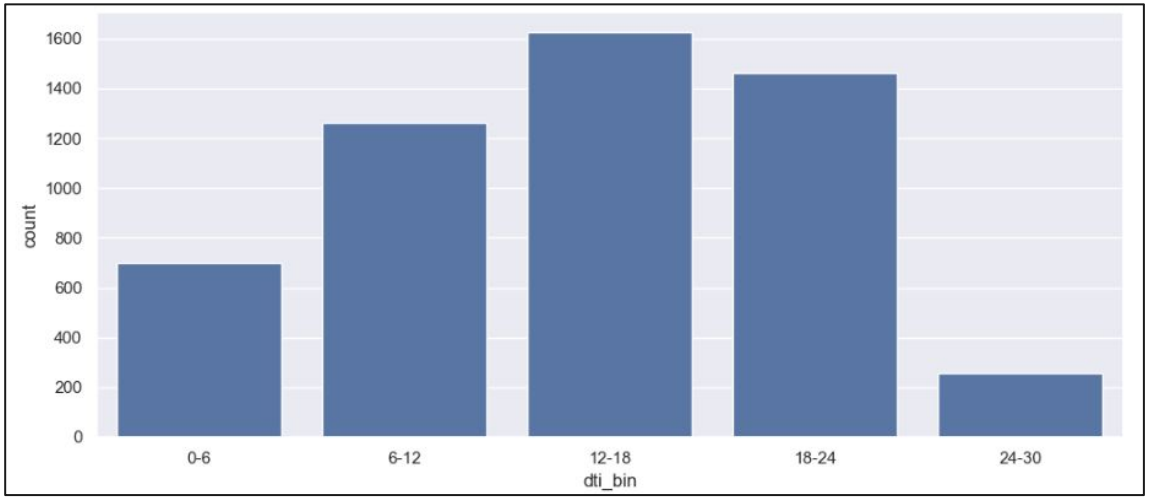






Bivariate Analysis of Quantitative variables - Plots





Bivariate Analysis on Categorical Variables

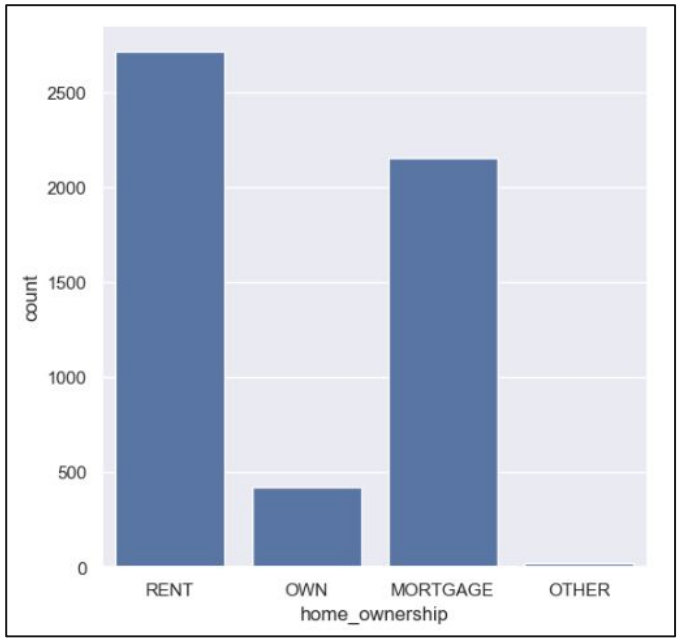
When analysed categorical variables with respect to the **charged off** loan, there is a more probability of defaulting when :

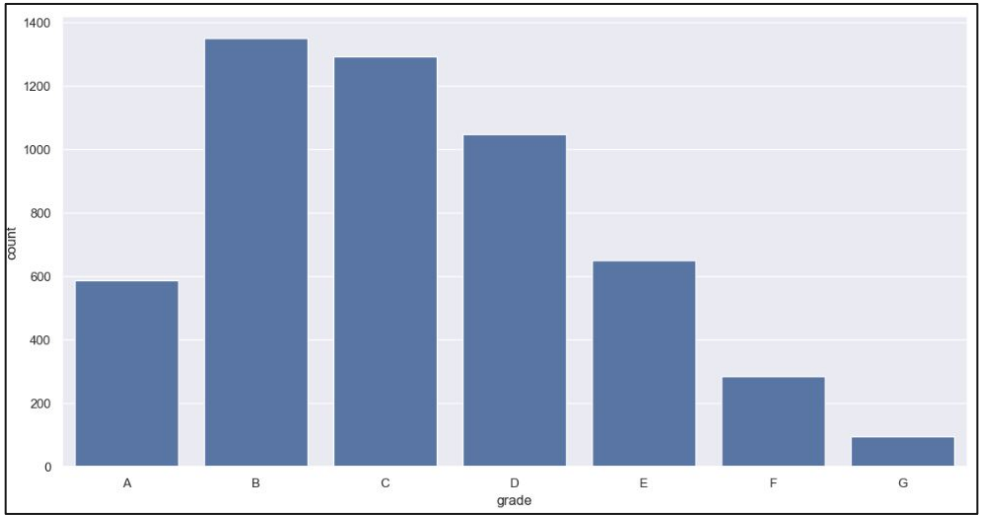
- Applicants having house_ownership as '**RENT**': 2500 applicants
- Applicants who use the **loan to clear other debts**: more than 2500 defaulters took loan for debt_consolidation
- Applicants who have **20-37 open_acc**
- Applicants with **employment length of 10 plus years : out of 5306 defaulters around 1331 defaulters are from 10+ years exp. Which is significant.Which is about 25.1%**
- **Term** of **36** months
- When the loan application is **Verified,** especially for higher loan amounts (more than 10k)
- When the no of enquiries in last 6 months is 0
- When the number of derogatory public records is 0
- Grade **B** borrowers have defaulted more and **C** is almost close to **B -** 1376 charged of cases are of grade B
- And majority defaulters are of subgrade 'B5' - close to 350

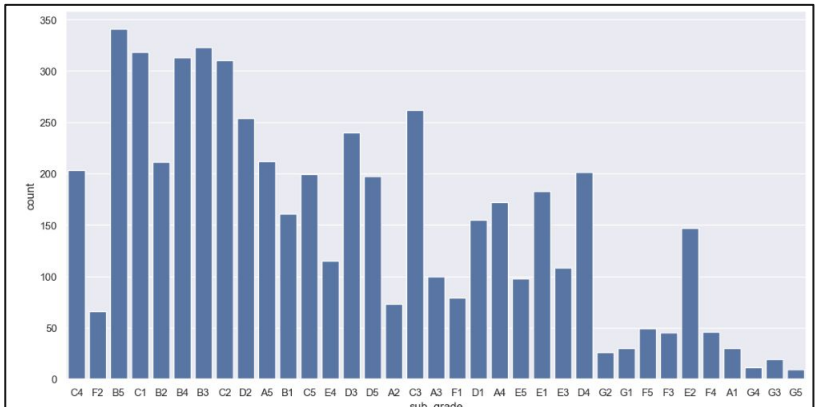
Though it may not be a deal breaker:

- Loan applicants from the state of California (CA), Florida (FL), and New York (NY) are most likely to default

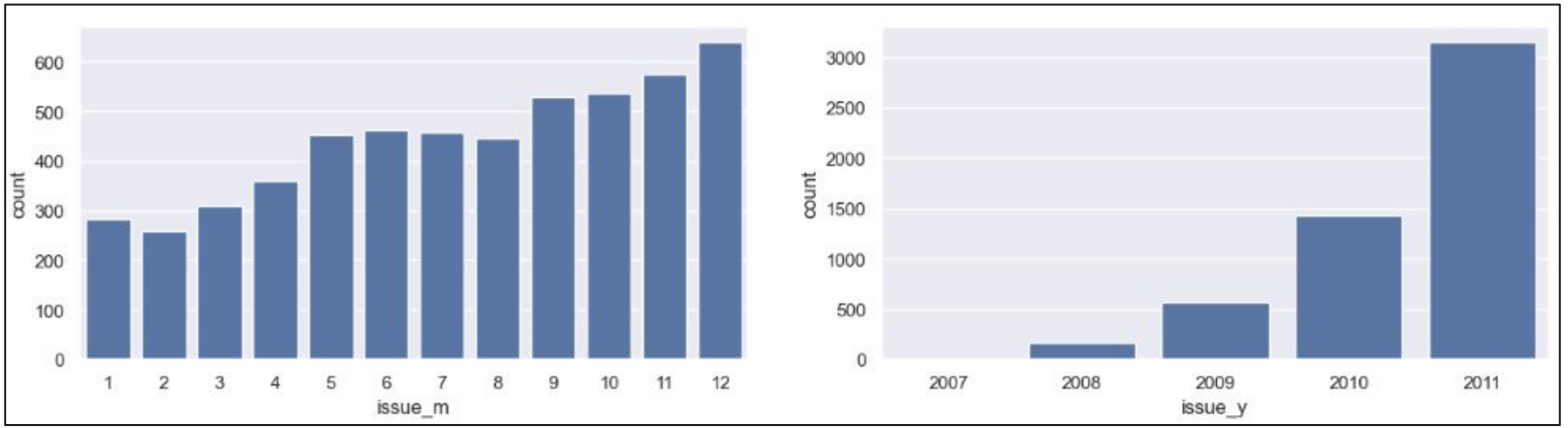
Bivariate Analysis on Categorical Variables

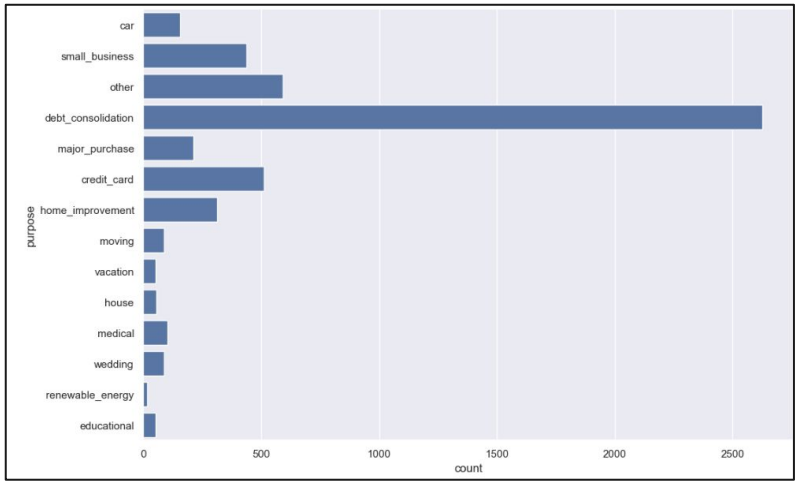






Bivariate Analysis on Categorical Variables

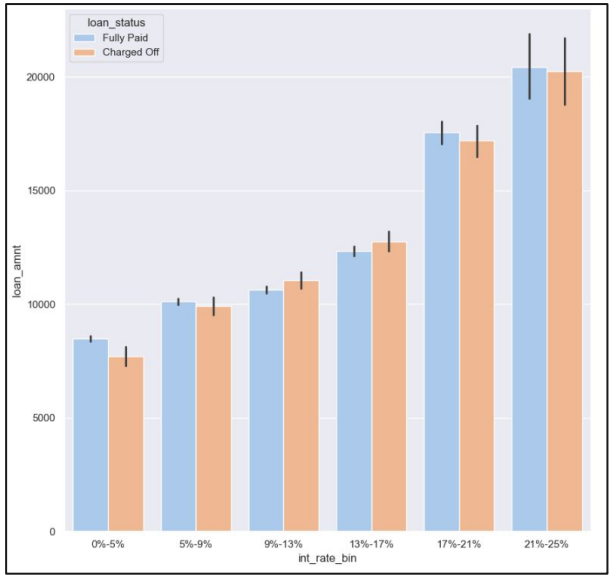




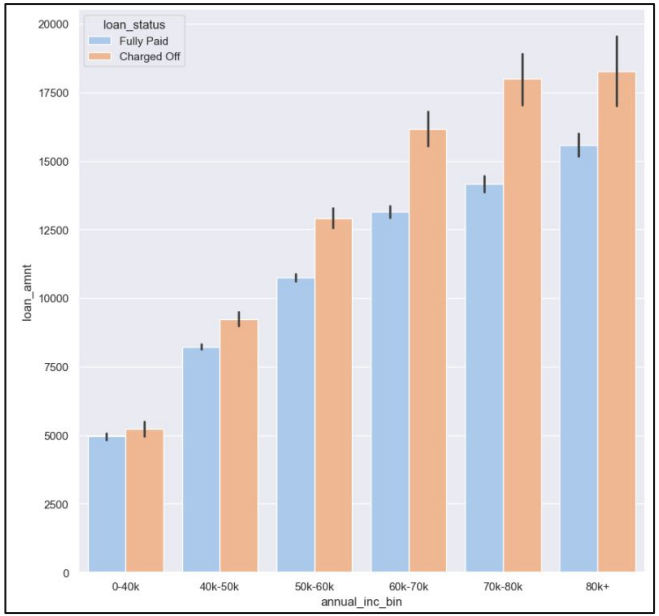
Multivariate Analysis of Continuous Variables

- The interest is high for higher loan amount and the number of defaulters are also the highest in this range.
- In every interest rate interval the number of fully paid applicants is almost equal to the number of defaulters
- The interest rate is more (21-25%) when the loan amount is more(15-20k)
- Highest defaulters are observed with the income range of 70 to 80k and above, and those who have availed loan for more than 17500 USD.
- Across all ranges of annual income, defaulters rate increases with the increase in loan amount and therefore increase in interest rate
- People with annual income range 60k to 80k+ have defaulted when the interest rate is above 14%
- people with more bankruptcy records have defaulted more and they are in the income range of 60 to 70k, also 50k-60k and 70-80k with very small gap
- People who are in the salary range of 55k to 65k and who have borrowed money for Home improvement, house construction and small business have defaulted more than any other category of purpose
- small business, debt consolidation, credit card are the main purposes where people have defaulted when the loan amount is more than 12k
- Across all types of house ownership people have defaulted when the loan amount is more than 10k.
- Among all ownership categories, mortgage and other have more defaulters and the loan amount is more than 12k

Plots for Continuous Variables Multivariate Analysis

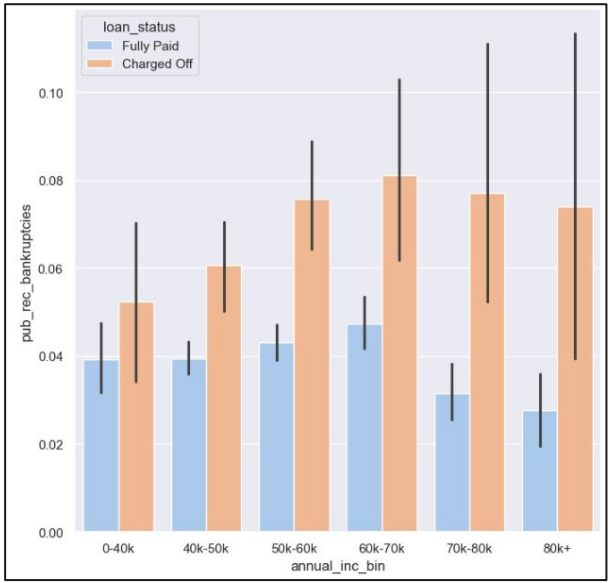


Interest Rate Vs. Loan amount

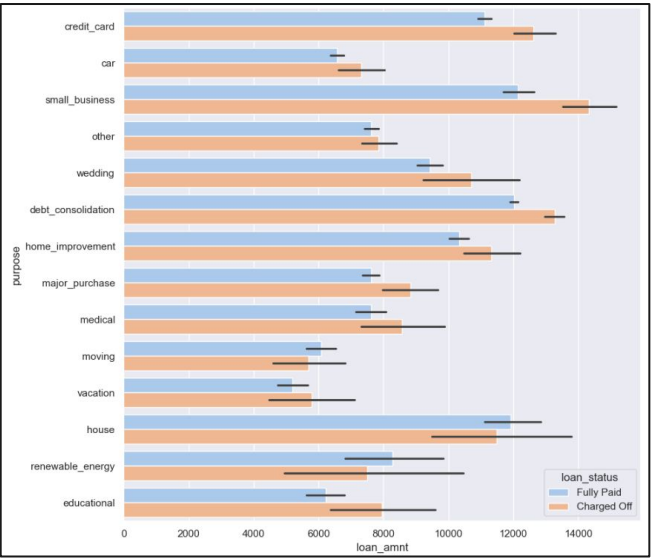


Annual Income Vs. Loan Amount

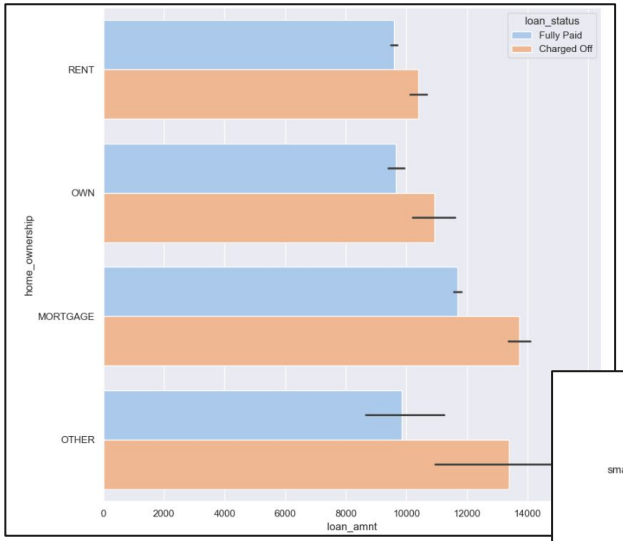
Annual Income Vs. Loan Public Records Bankruptcy



Plots for Continuous Variables Multivariate Analysis

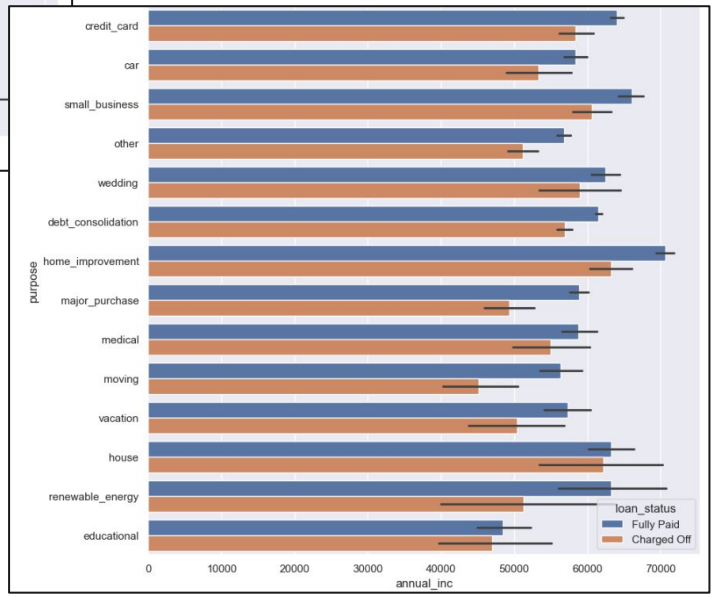


Loan Amount Vs. Purpose



Loan Amount Vs. House Ownership

Annual Income Vs. Purpose



Summary

- The **loan amount** is more for grades F, G and E respectively: more than 15k
  - number of defaulters are also more in the same category.
  - So every comparison and analysis boils down to the same reason that:
    - Number of defaulters are more when the loan amount is more.
    - Higher the loan amount, higher the interest rate
    - Also when the term is longer - 60 months
- Out of 5306 defaulters, **1468** defaulters are from 10+ years exp. I.e **27.67%** of total defaulters.
- 1349 defaulters are with Grade B: **25.4%**
- **2627** defaulters took loan for **debt consolidation**. I.e. **49.5%** of defaulters.
- 2714 applicants who have defaulted have **rented house**. Which contributes to again **51.15% of total defaulters**
- People with more bankruptcy records have defaulted more and they are in the income range of 60 to 70k
- **2364** defaulters have the repayment term as **60 months**. Which is **43.4%** of total defaulters
- **1149** defaulters have their income range from **5k to 15k**: **21.65%**

Recommendations

Recommendations to decrease the loss by Defaulters:

- Ensure only borrowers with a high likelihood of repayment are approved for loans.
- Consider factors like credit line, income level, and debt-to-income ratio when evaluating loan applications.
- Analyse the future situations of economy such as recession or financial meltdown before hand
- Check the house ownership, employment length, grade and annual income of the borrowers
- Strengthen the verification process as more number of defaulters are all verified and verified by sources
- Revisit the interest rate for evaluated borrowers as the majority of defaulters are the ones who are paying high interest rates
- Rework on the repayment term with people who have a bad credit history, as with longer term people are likely to have more financial commitments and are more likely to default
- Research more on the impact of geography on borrowers
- Higher the number of experience doesn't mean that the possibility of repayment is more. Have better loan lans for applicants with 10+ years of experience

Thank You