

LENDING CLUB-CASE STUDY

PRIYANKA YADAV



AGENDA

PROBLEM STATEMENT

DATA QUALITY

UNIVARIATE ANALYSIS

BIVARIATE ANALYSIS

FINAL RECOMMENDATIONS

Introduction

Business Understanding

As a worker of a **consumer finance company** which specializes 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 whether an applicant likely to pay the loan or not.

Problem Statement

Base on the loan dataset sample shared give recommendations on likelihood to repay the loan for which use EDA to understand how consumer attributes and loan attributes influence the tendency of default





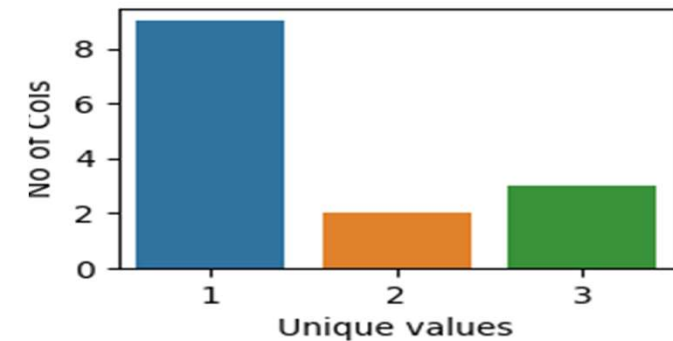
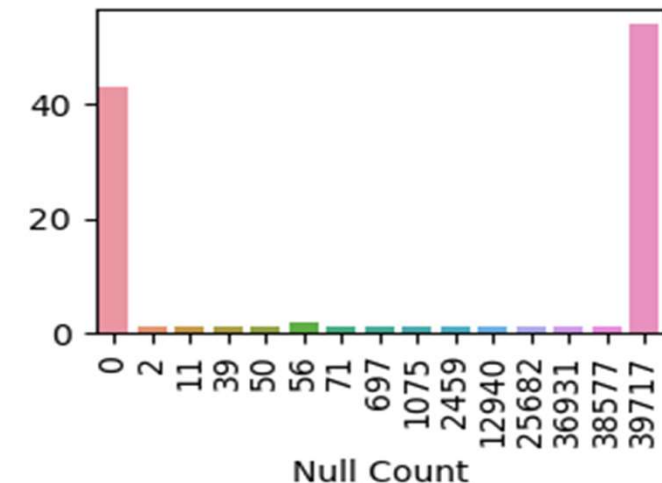
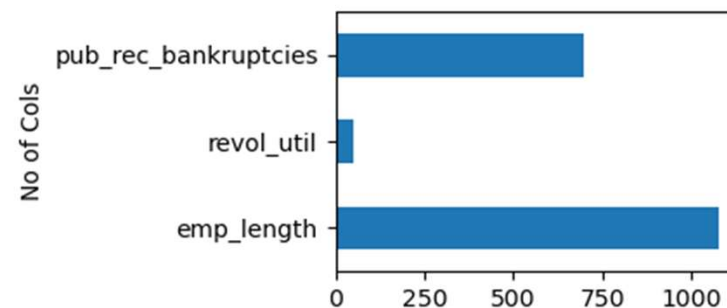
DATA CLEANSING

❑ **Remove Nulls** : As there are more than 50 columns having all rows as null , they can be removed from analysis

❑ **Columns with single unique value**: They cannot aid in any categorical analysis we have around 8 such columns.

❑ **Elimination based on domain knowledge**: Since the case study is for making decision before giving loan based on various demographic details of an applicant columns reflecting payment schedule or customer behavior generates data post loan is approved hence they can not be considered during analysis, which narrows down to approx. 29 columns.

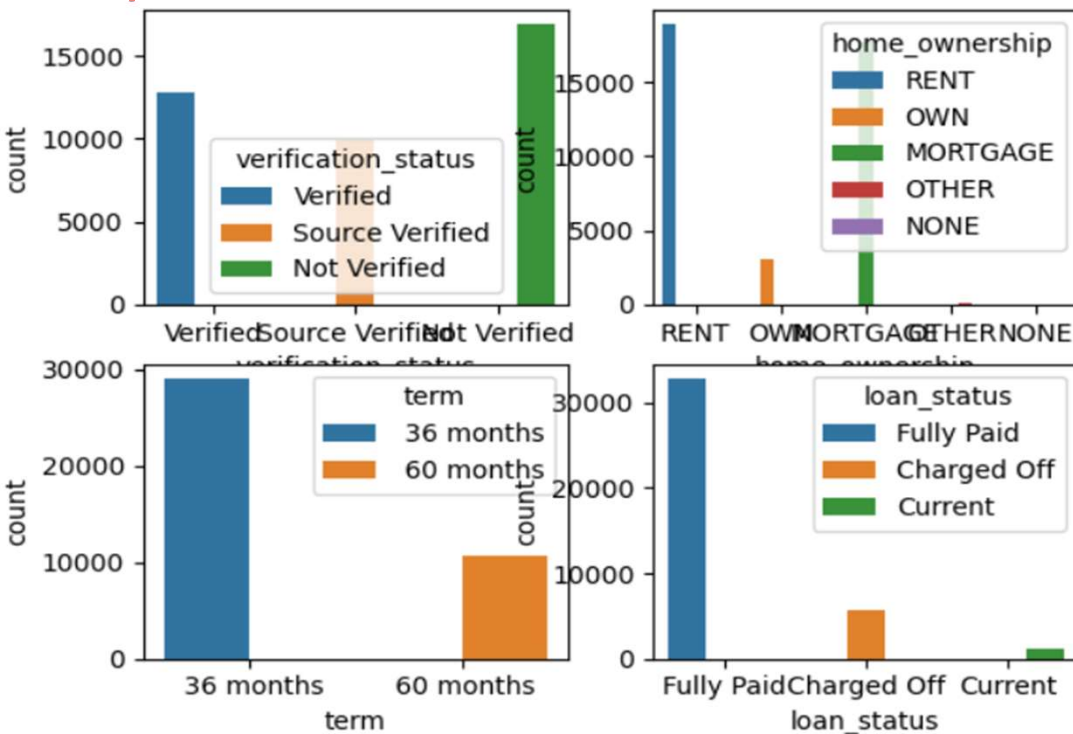
❑ **Filling Nulls**: Updating null values for the inscope columns



❑ Identifying and Correcting Categorical variables –

Categorical values which has less than 10 unique columns checking for any non-standard value present

- **home_ownership:** NONE can be updated as Other

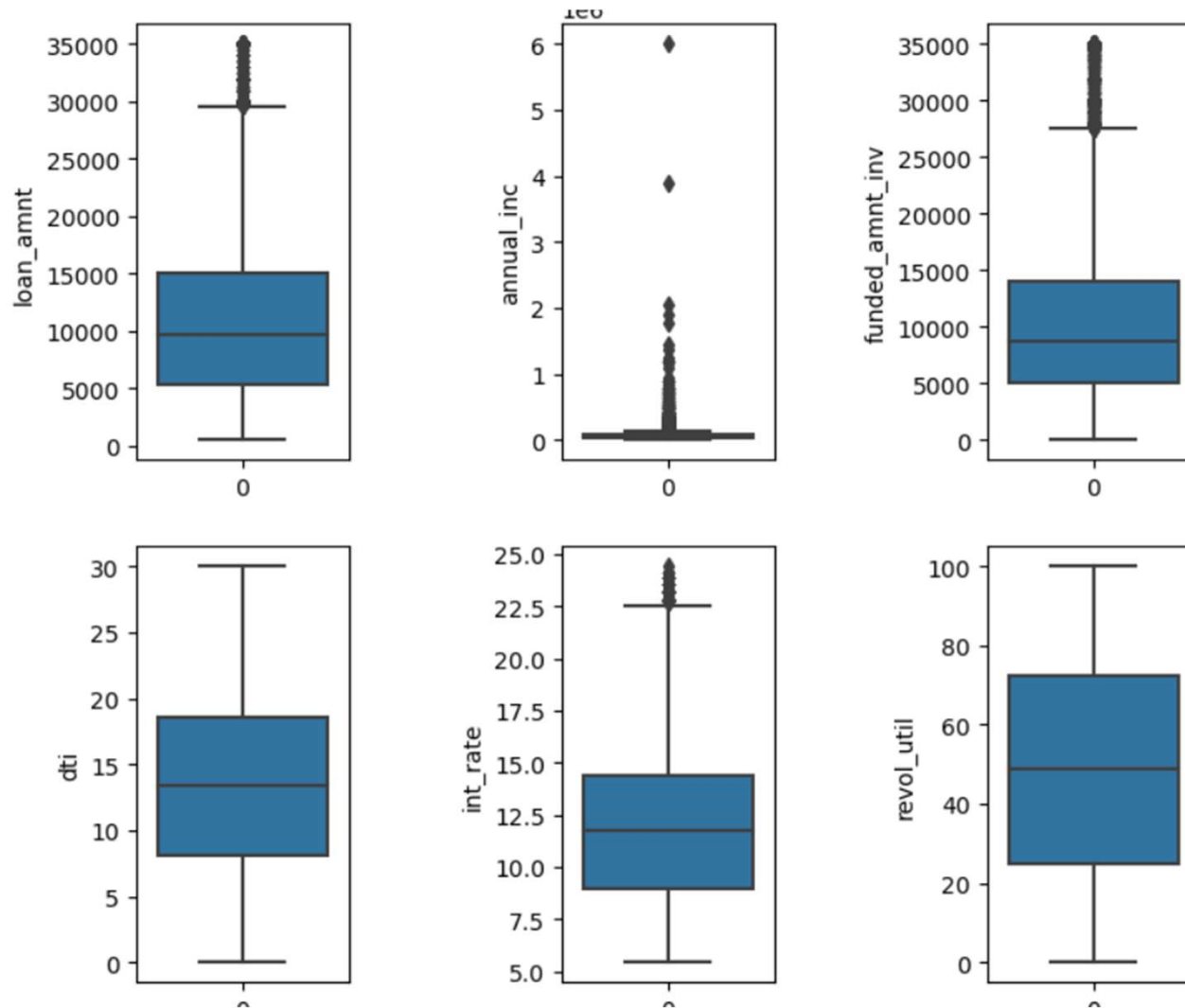


❑ Updating datatypes and deriving columns

- ❑ **Filtering:** As our analysis focus on determining Defaulters or Non defaulters while neither can be concluded for Current applicants

int_rate	installment	grade	sub_grade	emp_length	home_ownership	...	pub_rec	revol_bal	revol_util
10.65	162.87	B	2	10	RENT	...	0	13648	83.7
15.27	50.82	C	1	0	RENT	...	0	1687	0.7

❑ **Outliers** – In the summary datasets checking outliers



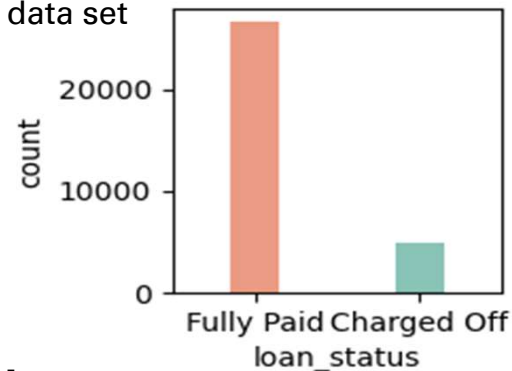
Overall dataset looks good ***except annual income***, we can exclude outliers from the analysis.

UNIVARIATE ANALYSIS



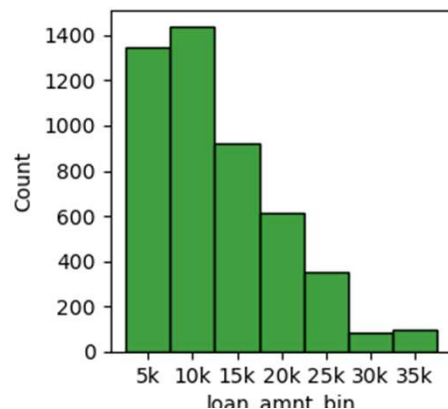
➤ Loan Status

As around 5000 defaulters are there in the dataset we will be performing categorical analysis on the charged off data set



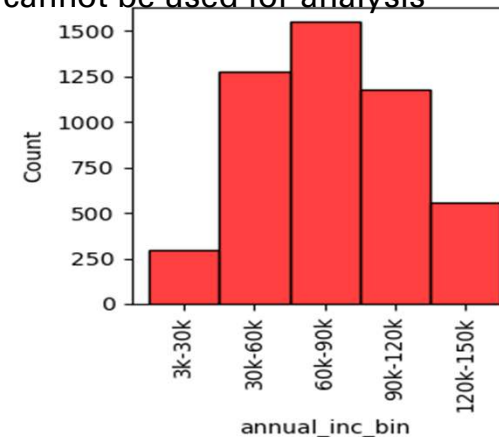
➤ Loan amnt

Frequency distribution of loan amount
Shows higher number of defaulters for Loan_amount 5-15k.



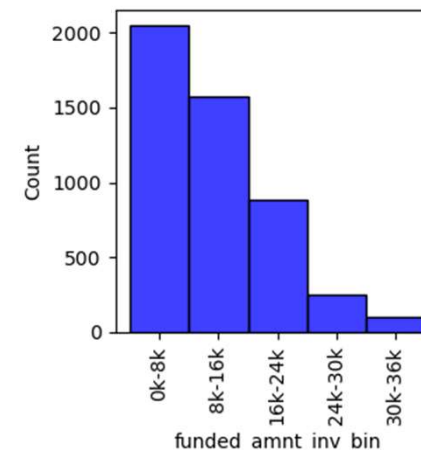
➤ Annual income

As defaulters are spread across various incomes this cannot be used for analysis



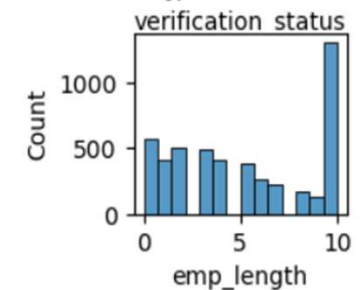
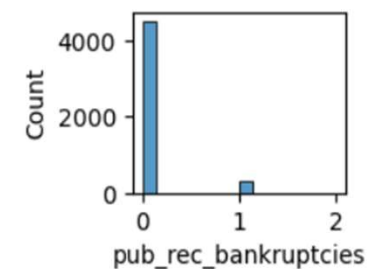
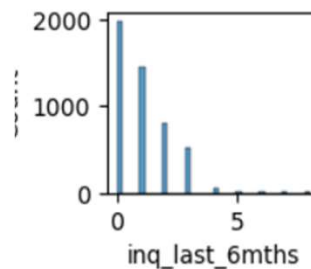
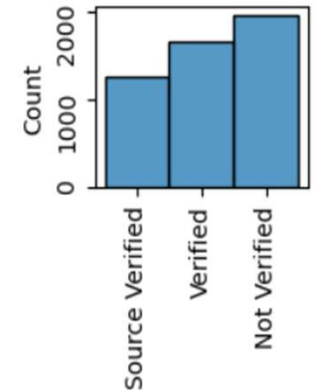
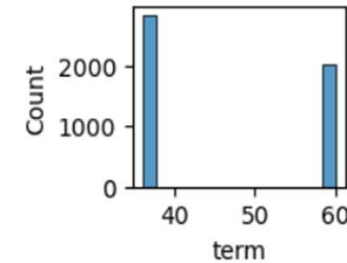
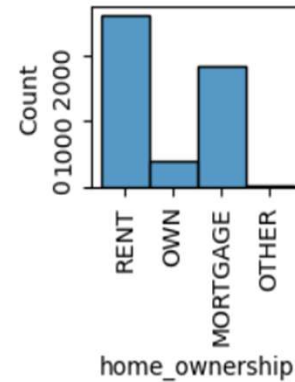
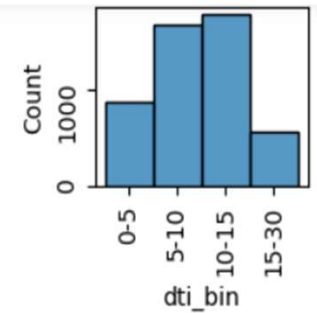
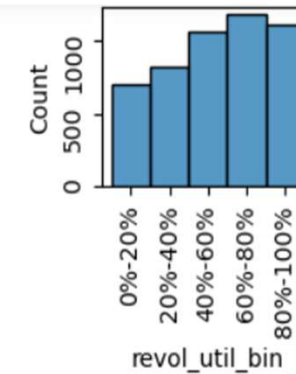
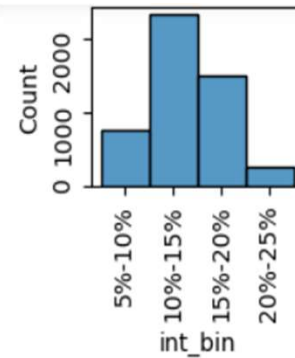
➤ funded amnt inv bin

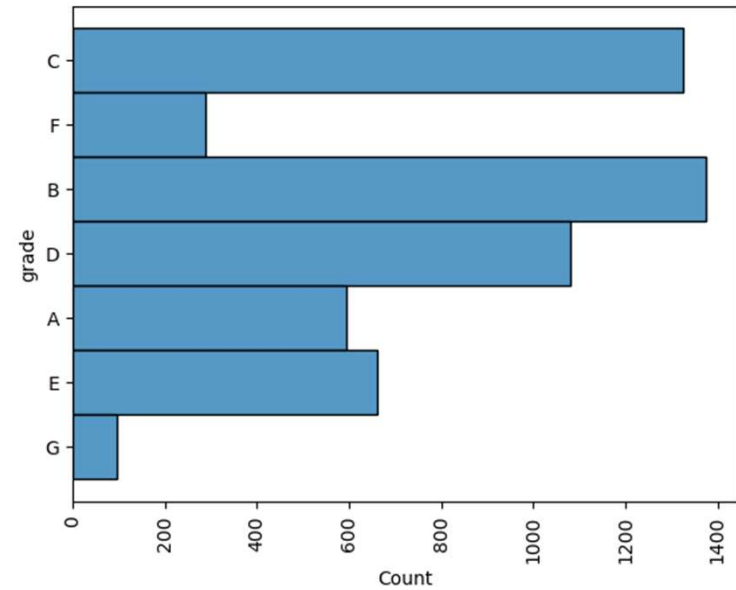
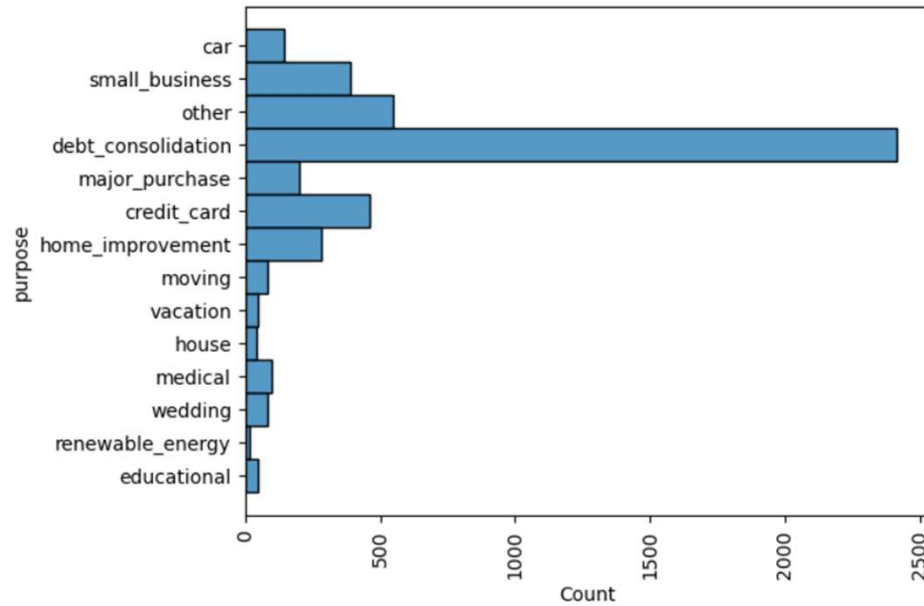
Low funded amount invested higher frequency of default



Highest number of defaulters have

- opted for **interest** rate **10-15%**,
- Has debt to income ratio (**dti**) 5-15%,
- Mostly having **rented** property or other **mortgages**.
- Taken loan for **duration** of **36** months
- Long **employment** length more than **10 years**
- The number of **inquiries** in past 6 months were nearly **null**
- income was **not verified** by LC
- Purpose was debt consolidation
- LC assigned loan **grade** was **B,C**





From the univariate analysis of the given columns ,below were the major factors affecting loan approval

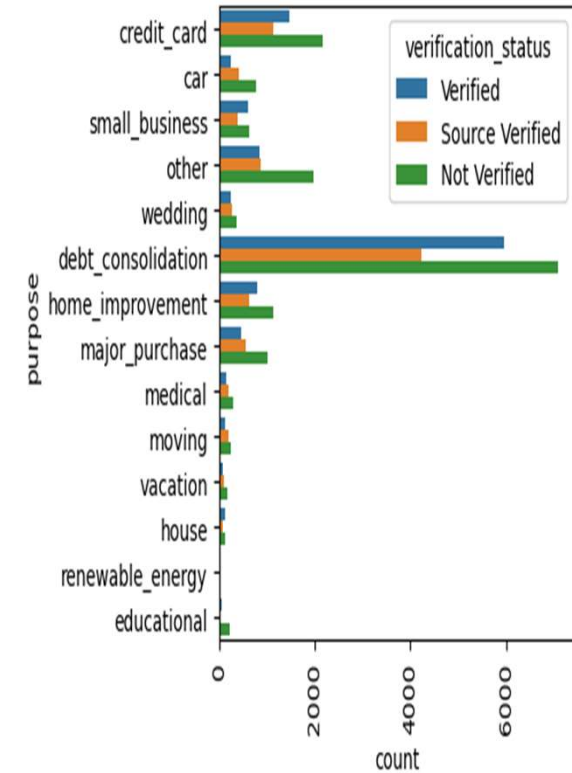
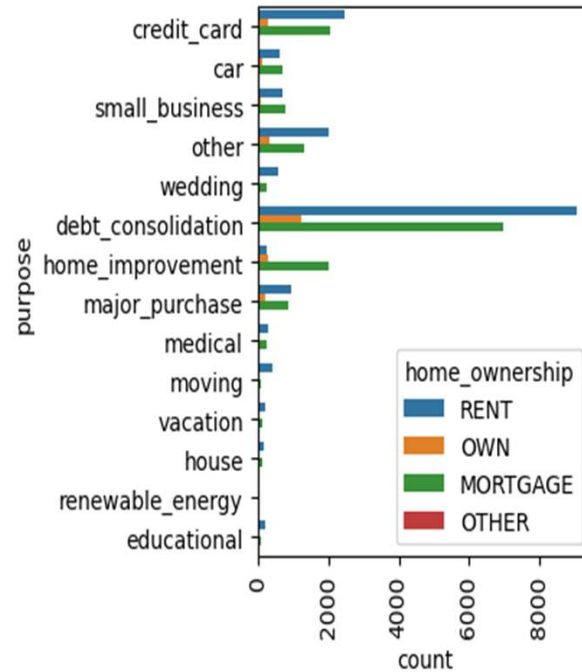
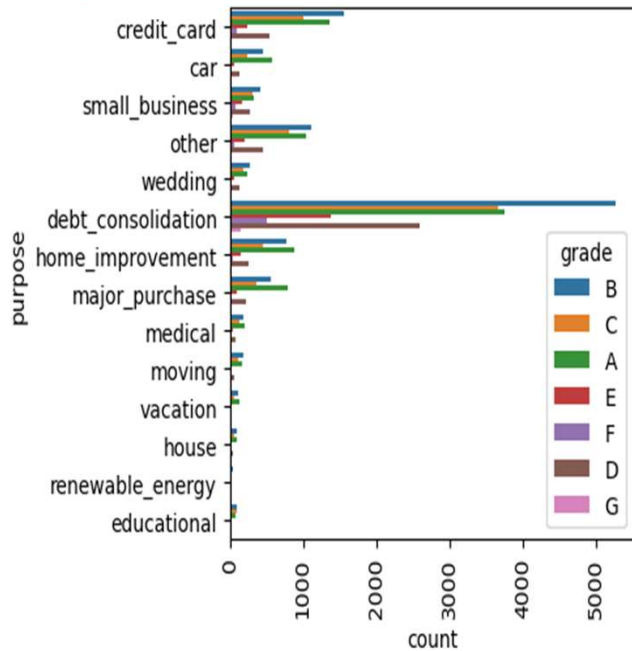
- Term
- Inq_last_6mnths
- Emp_length
- Home_ownership
- Verification status
- Purpose
- Grade
- DTI

BIVARIATE ANALYSIS



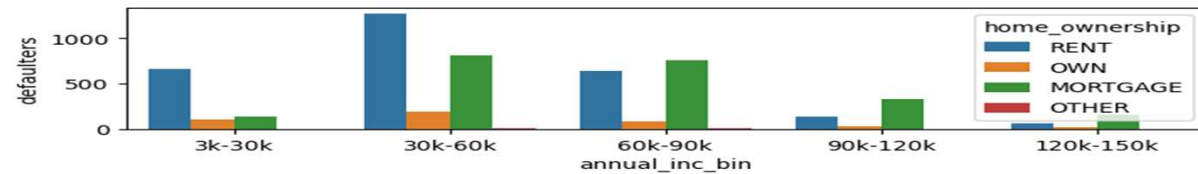
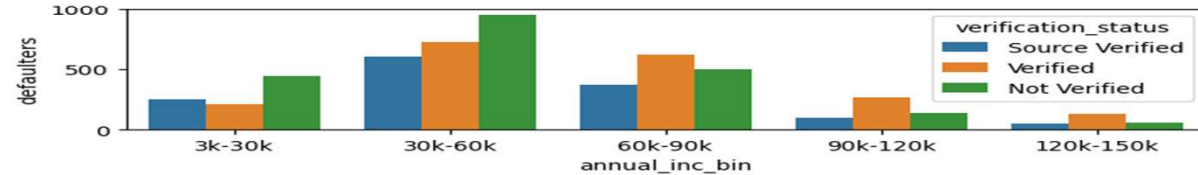
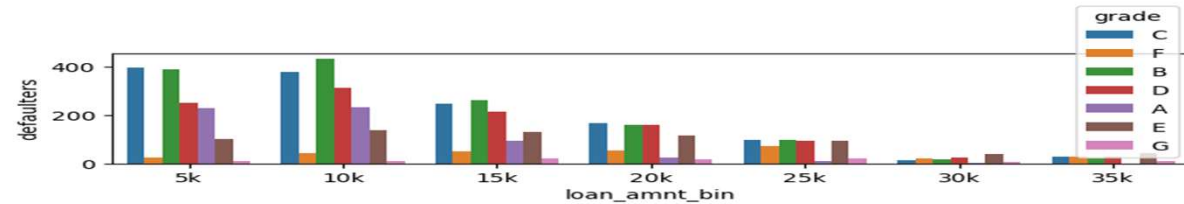
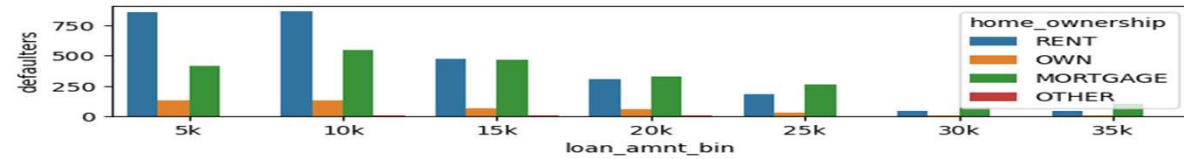
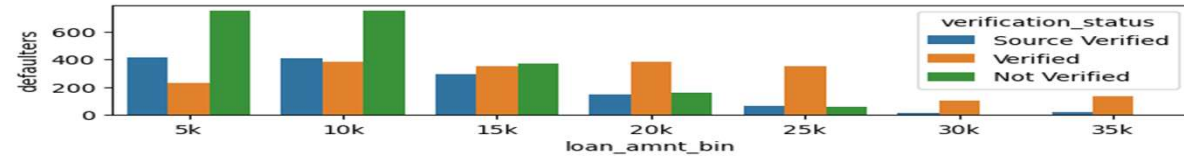
❖ Categorical vs Categorical

- It is interesting to not that purpose **debt consolidation** positive relation with applicants being **on rent, grade B and not verified**

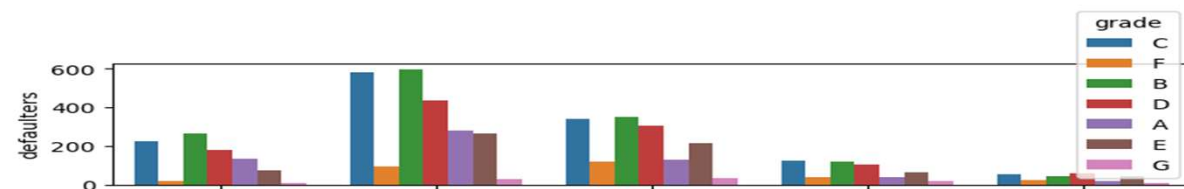


❖ Categorical vs Numerical

- Most defaulters with income not verified, has rented property, belong to grade B,C has loan_amount in the range 0-10k
- Annual income 30-60k has large number of defaulters directly linked with income not verified, has rented property, belong to grade B,C



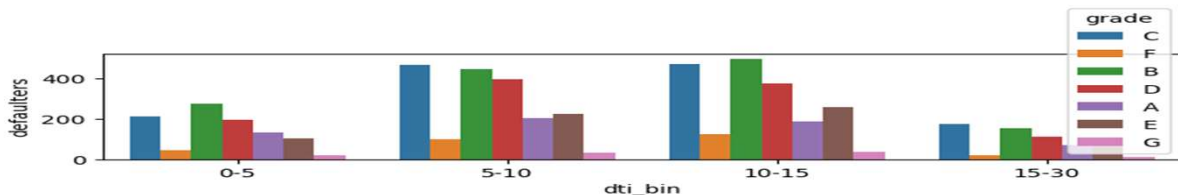
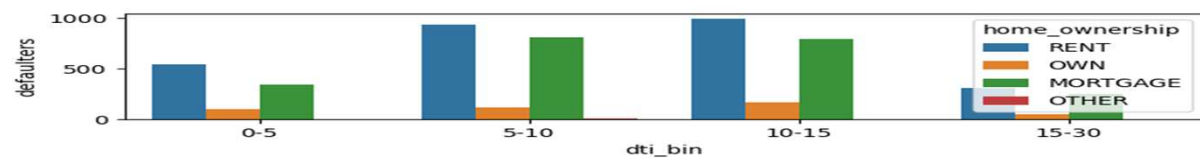
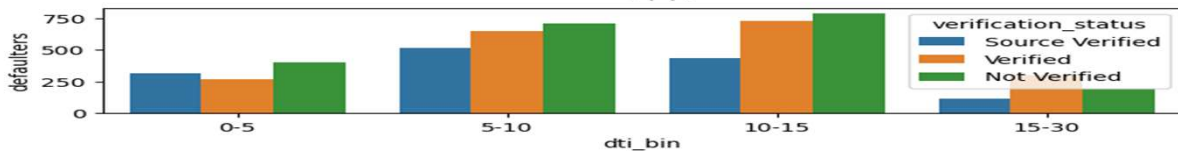
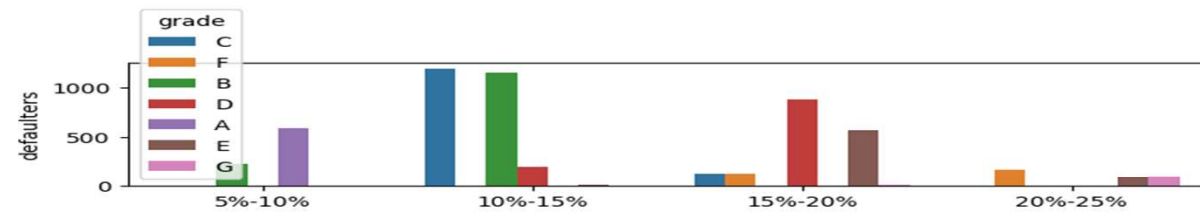
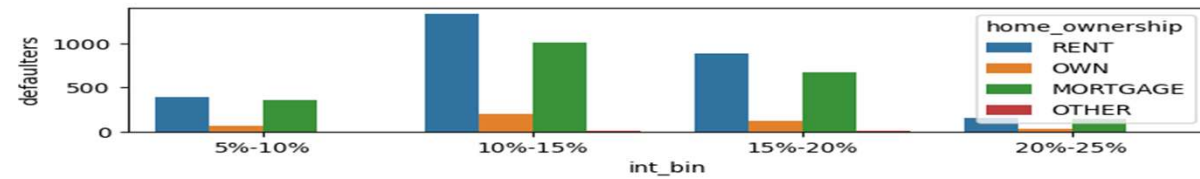
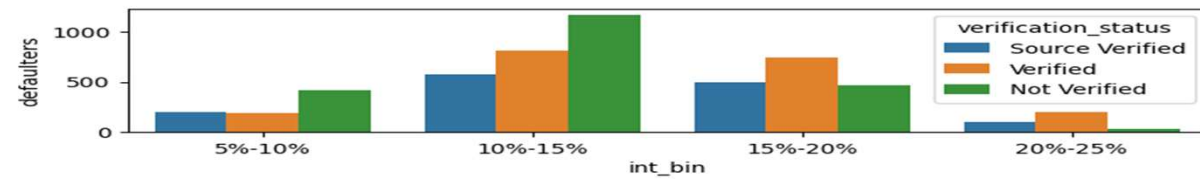
PRES



- Debt to income ratio: has most defaulters in 5-10 and interest_rate 10-15%

Applicants in higher risk category are from

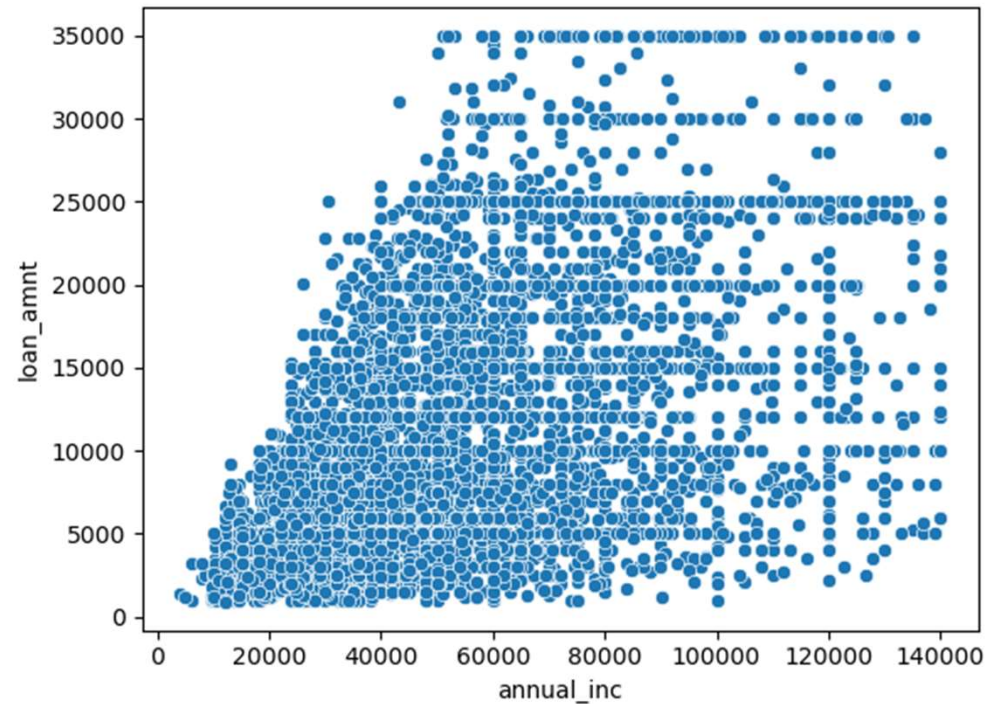
- > Grade B and C
- > Have rented property
- > Not verified
- > has debt to income ratio 5-15
- > annual income of 30-60k
- > purpose debt consolidation
- > not been inquired since last 6 months
- > opted for shorter loan duration



PRE

❖ Numerical vs Numerical

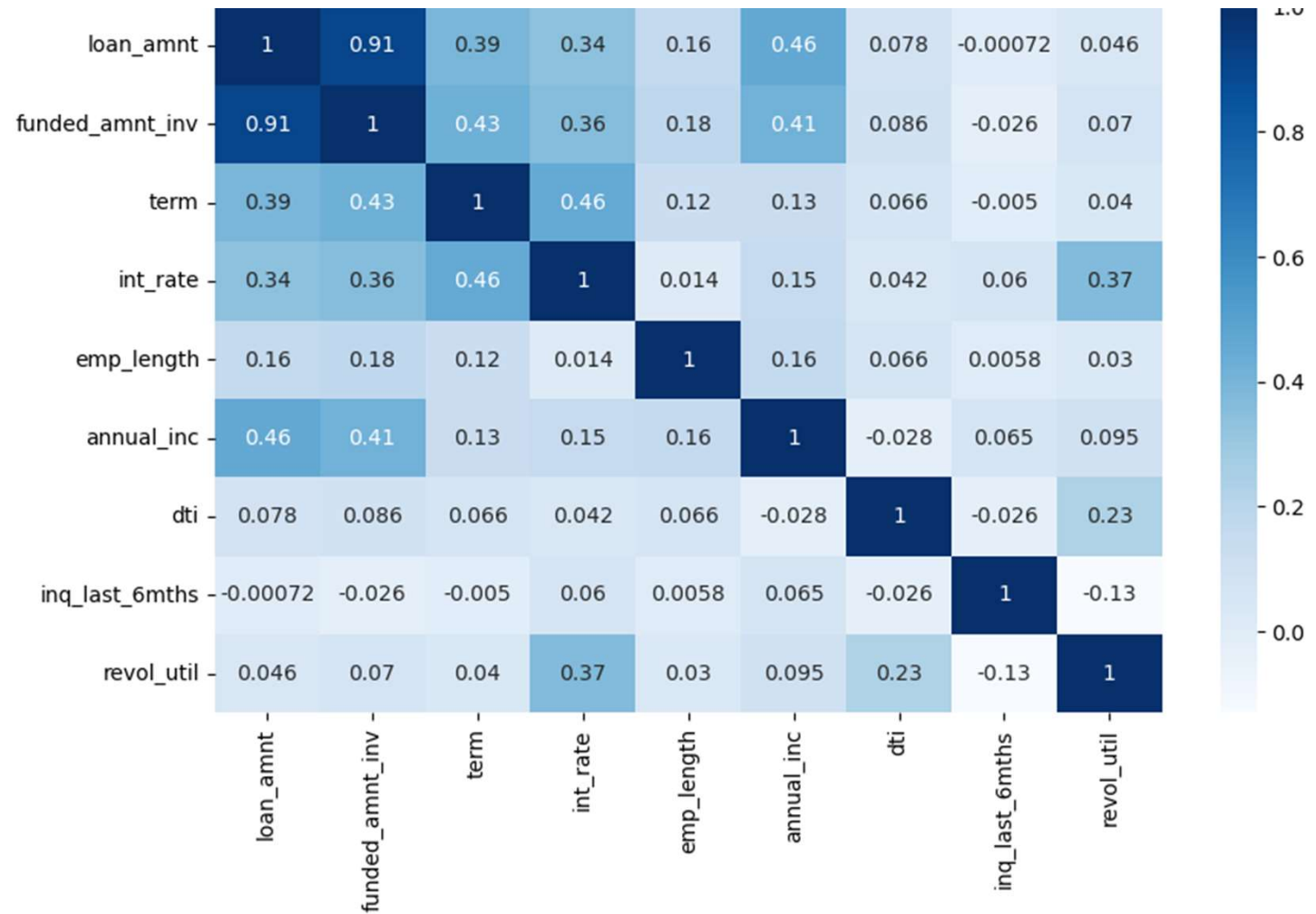
From the scatterplot its evident that loan_amnt increases abruptly between annual_inc 20k-60k which shows maximum densite of defaulters. Applicants in this income group have increased their loan capacity drastically despite slower rise in income



MULTIVARIATE ANALYSIS



- Loan amount strongly influences funded amnt invested.
- Debt to income ratio is reversely proportional to annual income
- Overall loan_amnt ,annual_income,int_rate,funded_amnt_inv are the major influencers



Final Insights from the Analysis

- ✓ Overall there were **13.6%** defaulters in the dataset.
- ✓ Defaulters took higher loan_amount compared to their annual income.
- ✓ Also dataset reflected those who paid fully had lower dti(debt to income ratio) , took loan_amount comparable to their annual income.
- ✓ Purpose “debt_consolidation” was a clear indicator where defaulters mainly stated that they wanted to payoff some previous debt
- ✓ Most applicants whose annual income were not verified by the Lending Company had defaulted.
- ✓ Also those who had previous mortgages, on a Rented property were in high risk category.
- ✓ Constantly grade B,C showed positive relation ship with other risk category –“Not verified”, “Rent”, “higher dti”