



# Lending Club Case Study

By- Vishakha Singh

Tuesday, May 3, 2022

Developed as part of the Exploratory Data Analysis Module required for Executive PG Programme in Machine Learning and AI - IIIT, Bangalore.

# Problem Statement

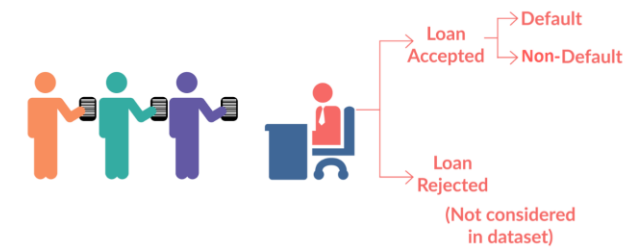
A **consumer finance company** which specializes in lending various types of loans to urban customers receives loan applications, the company has to make a decision for loan approval based on the applicant's profile.

Two **types of risks** are associated with the bank's decision:

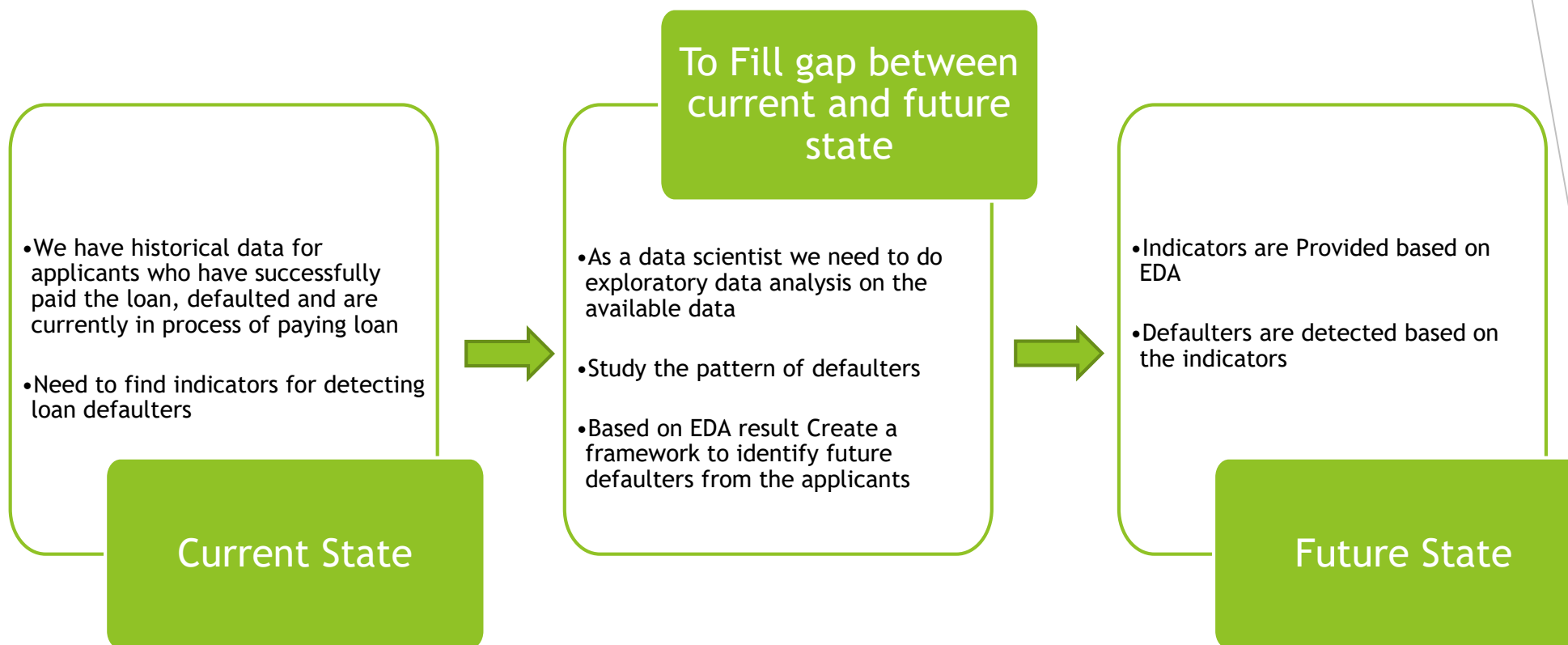
- ▶ If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- ▶ If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

We need to identify indicators which can be used for accepting or rejecting the Loan Application

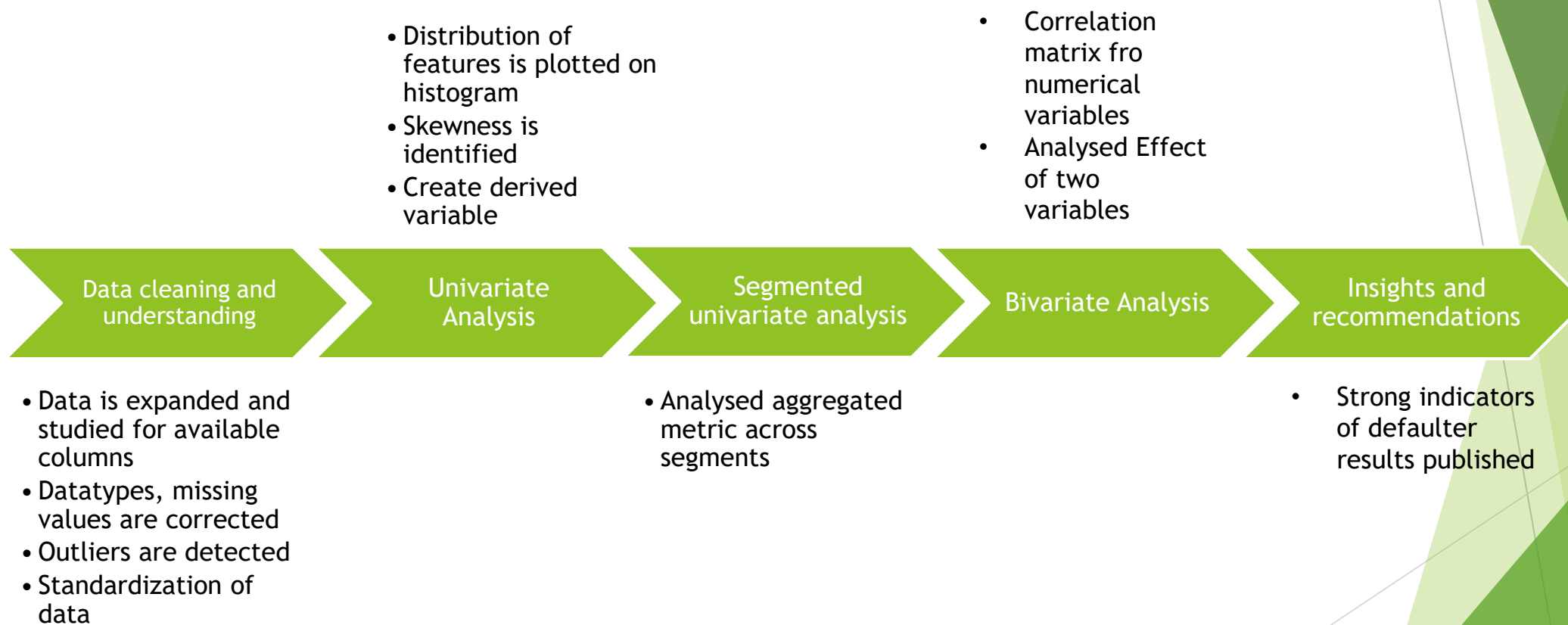
LOAN DATASET



# Problem Analysis



# EDA Approach



# Data Understanding

Categorization of variables happened based on two categories-

- Pre and post loan acceptance variables
- We considered pre loan accepted variables for our analysis as this information would be present at the time of loan application

Pre Loan acceptance variables
loan_amnt
funded_amnt
funded_amnt_inv
term
int_rate
installment
grade
sub_grade
emp_length
home_ownership
annual_inc
verification_status
issue_d
purpose
addr_state
dti
pub_rec
pub_rec_bankruptcies

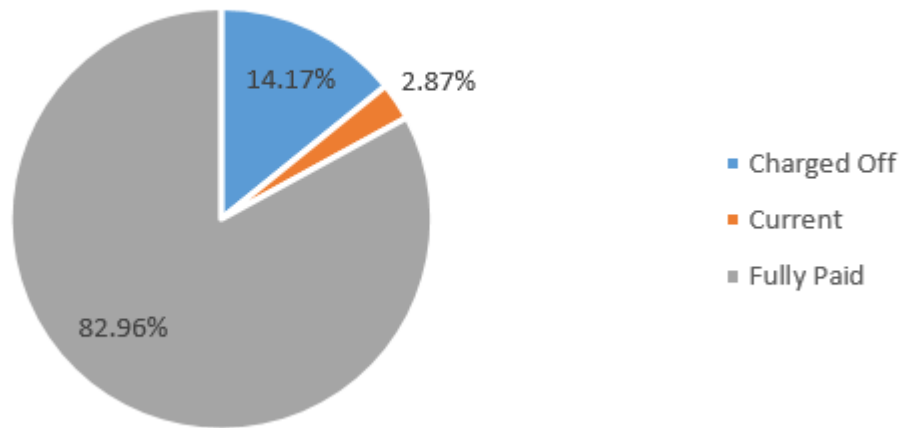
Post loan acceptance variables
delinq_2yrs
earliest_cr_line
inq_last_6mths
open_acc
pub_rec
revol_bal
revol_util
total_acc
out_prncp
out_prncp_inv
total_pymnt
total_pymnt_inv
total_rec_prncp
total_rec_int
total_rec_late_fee
recoveries
collection_recovery_fee
last_pymnt_d
last_pymnt_amnt
last_credit_pull_d
application_type

# Univariate and segmented Univariate analysis

# Loan status distribution

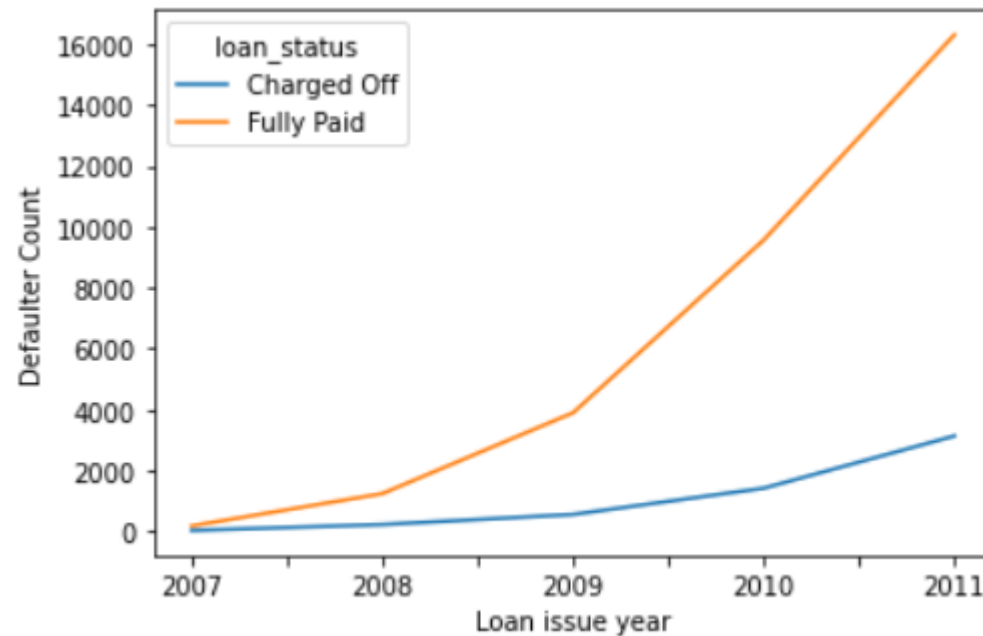
There are ~14% defaulters in the data set

Applicants distribution across loan status



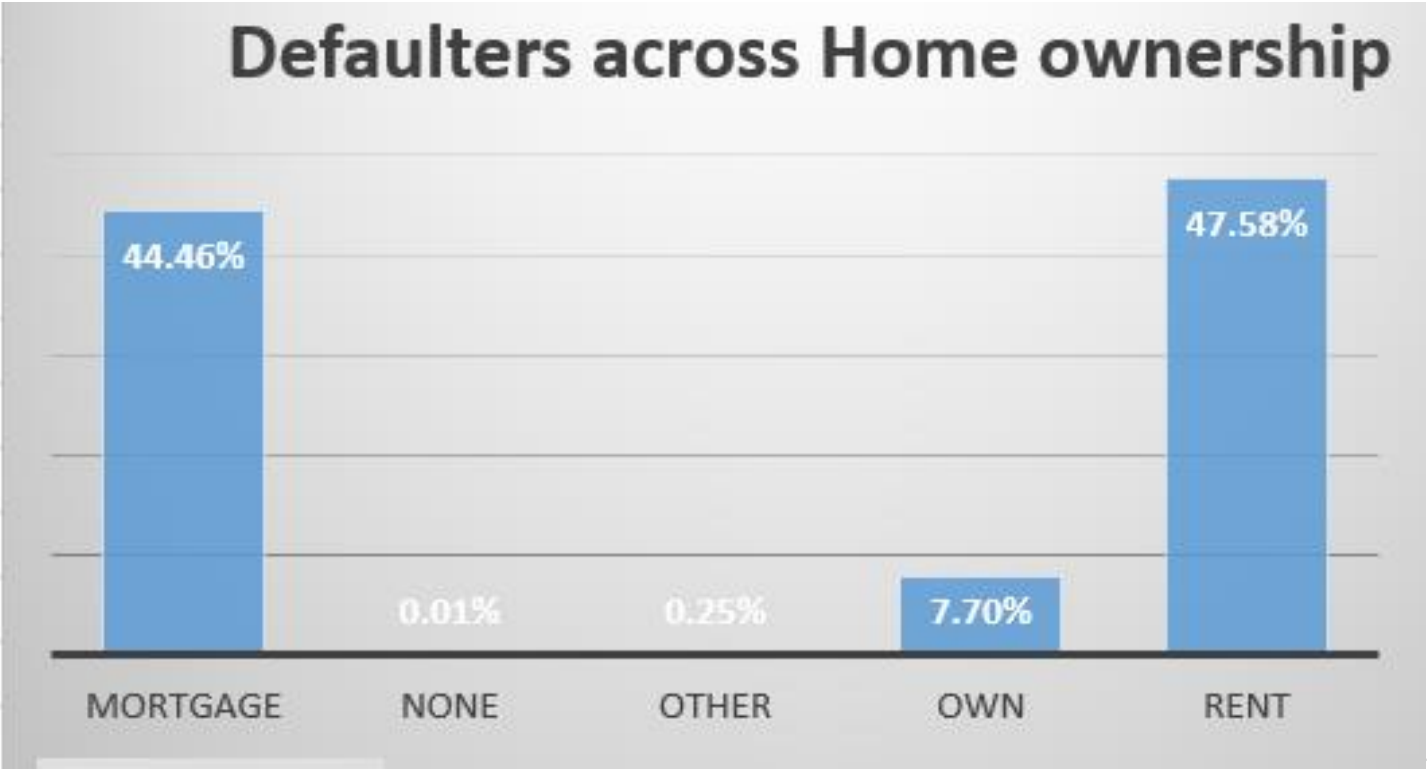
# Analysis across years

- Data is spread across the years 2007 to 2011
- Around 58% loans issued in year 2011 resulted in default

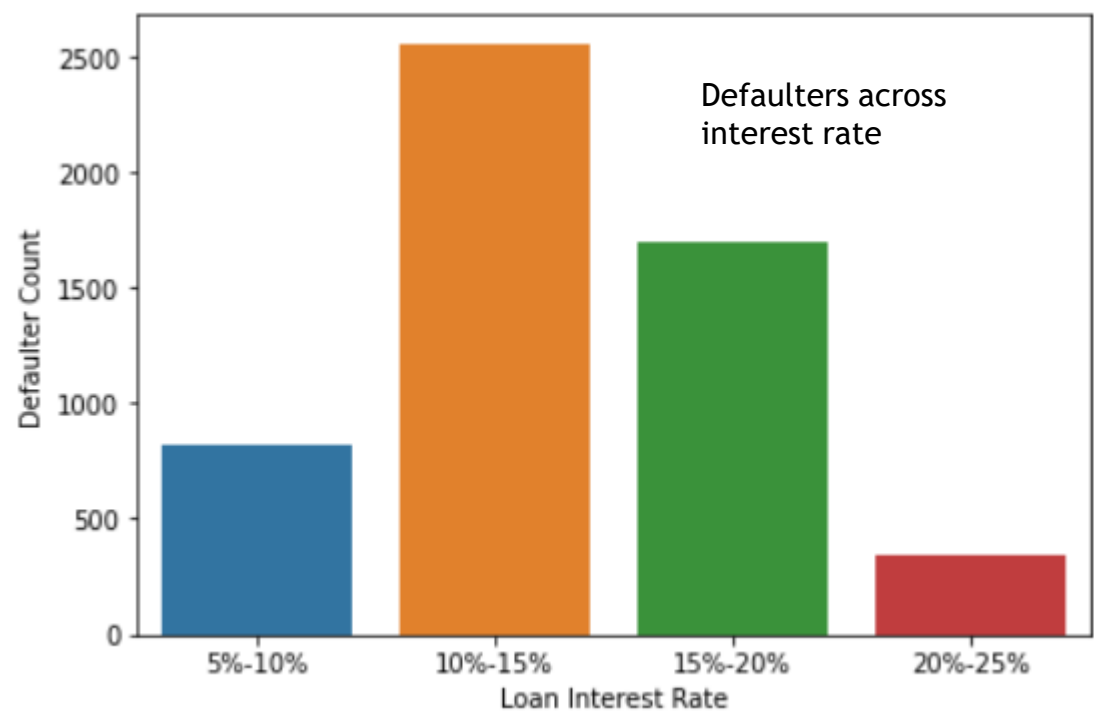




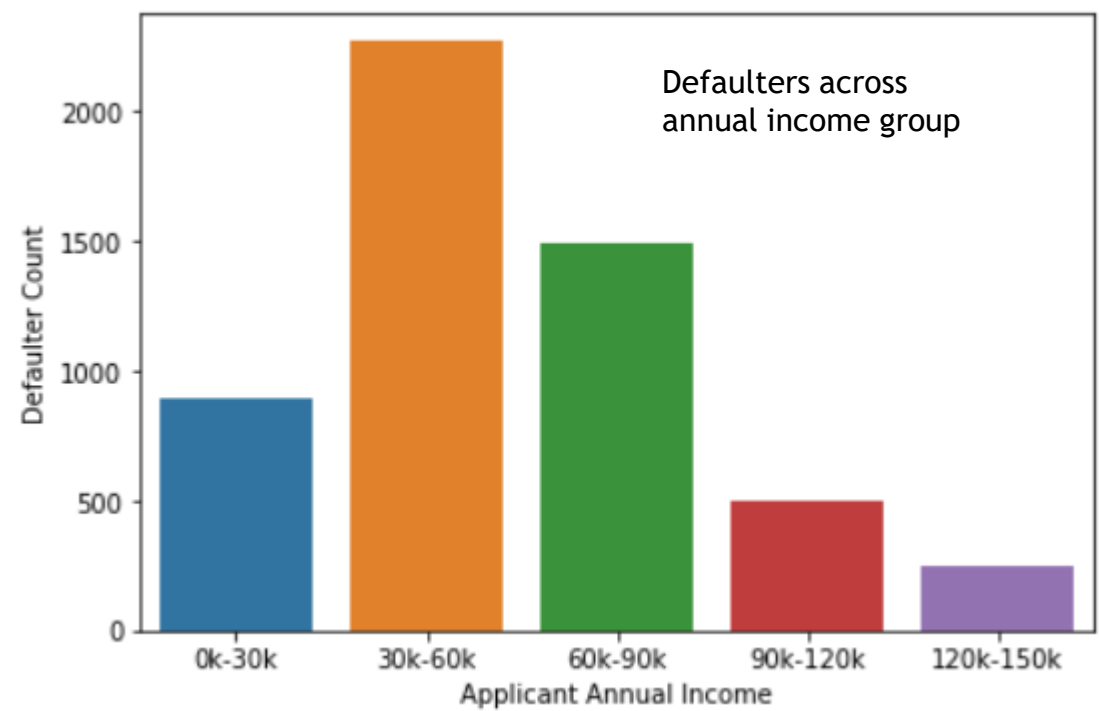
Around 92% applicants having home ownership status as rented or mortgage were defaulters



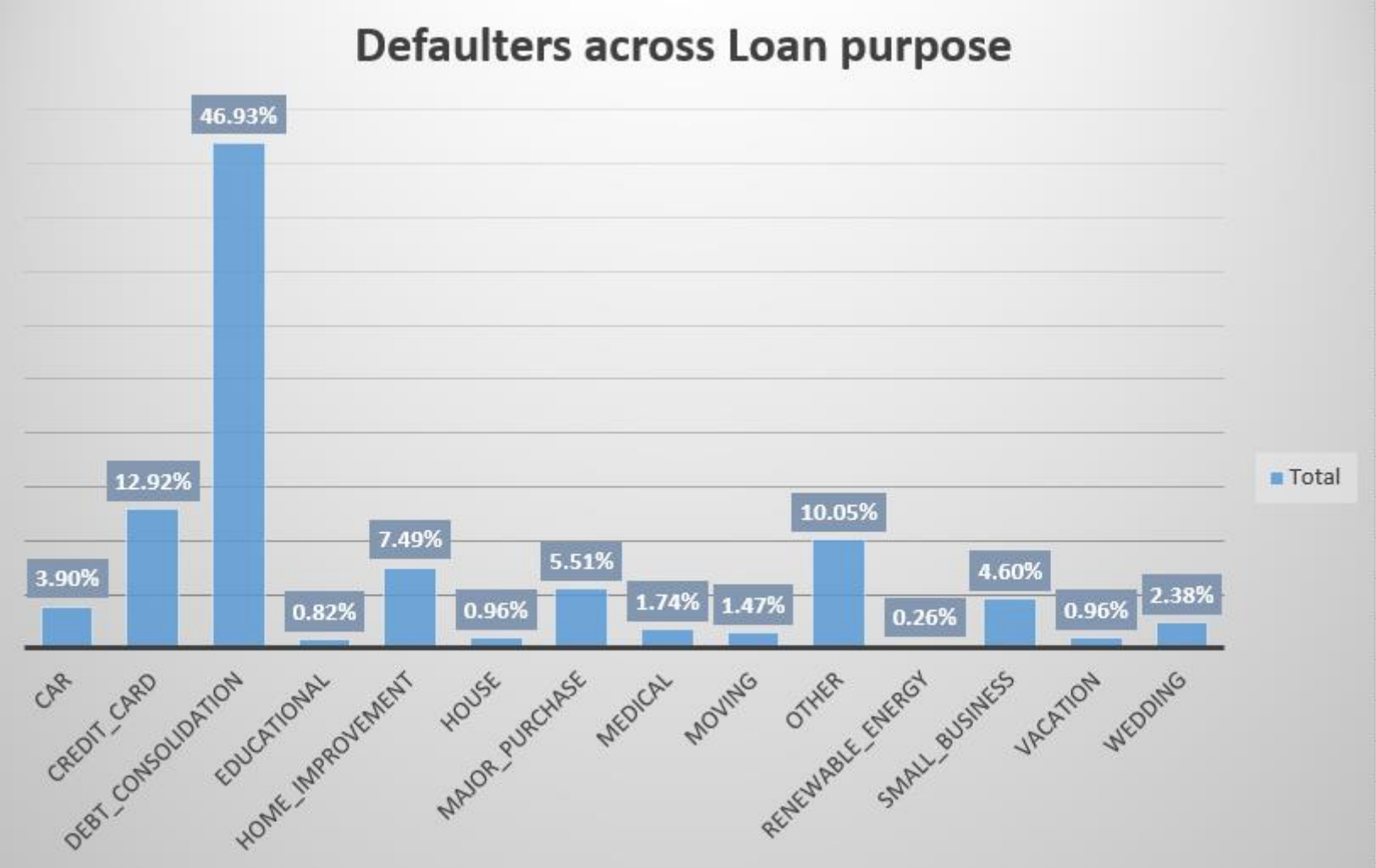
Around 46% Applicants who had an interest rate in range 10%-15% defaulted



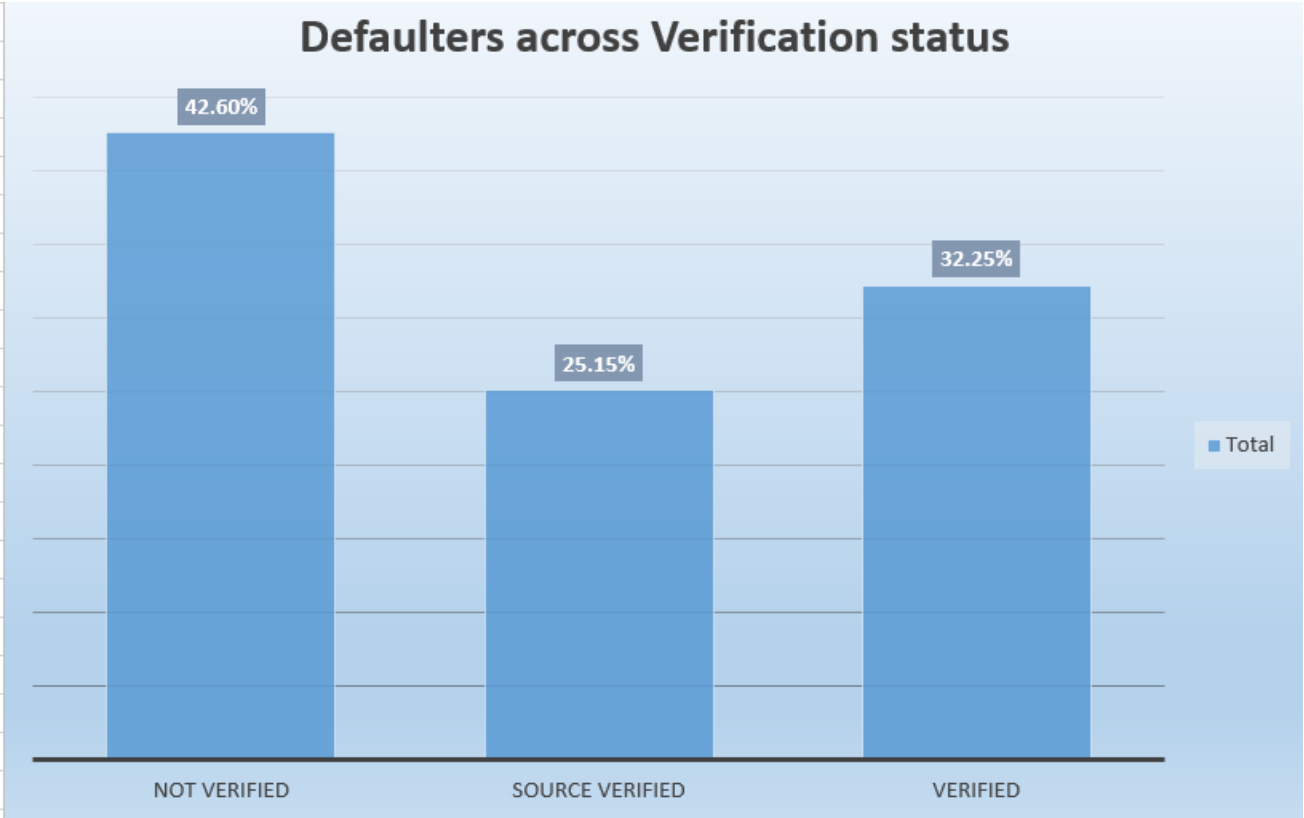
# 42% Defaulters had an annual income in range 30k-60k



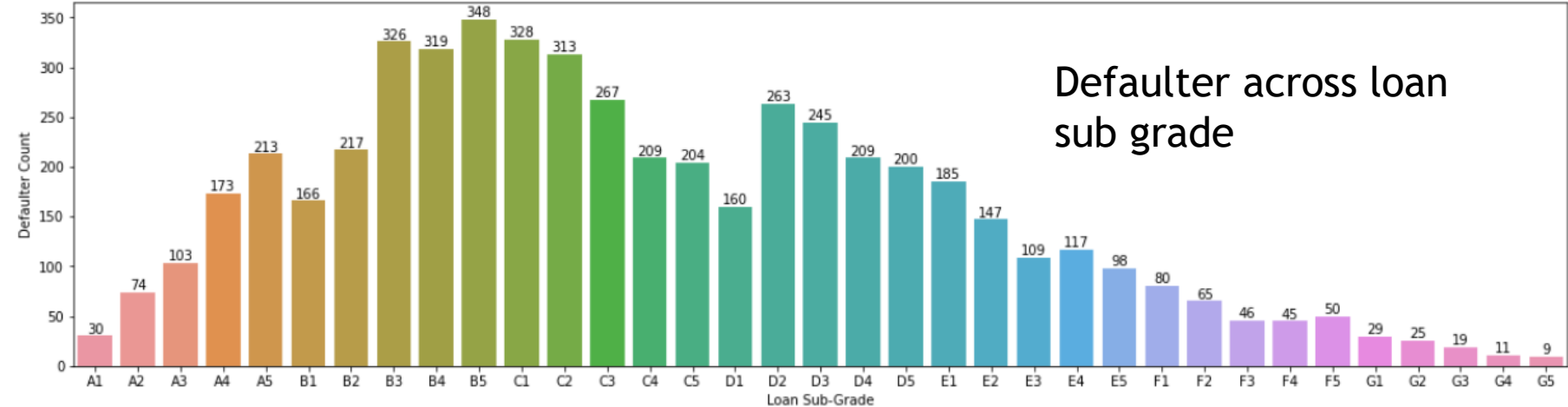
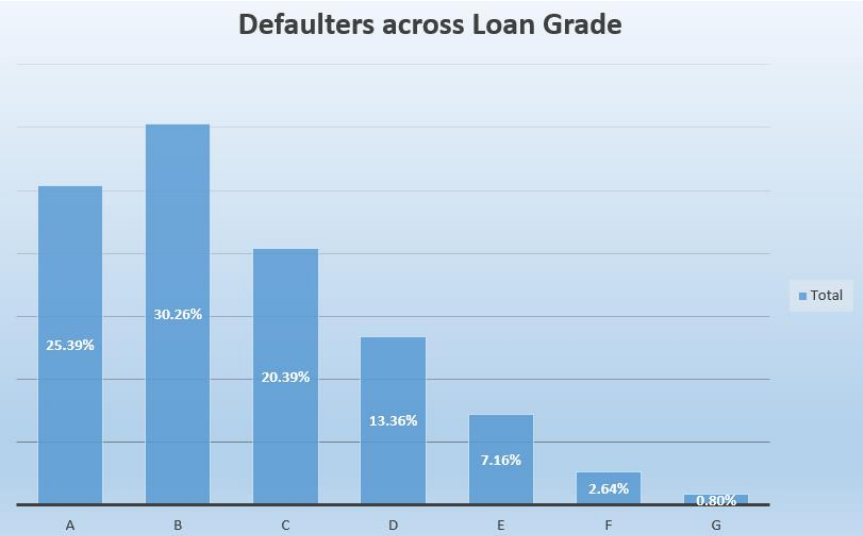
Debt\_consolidation was the purpose of loan for 47% defaulters



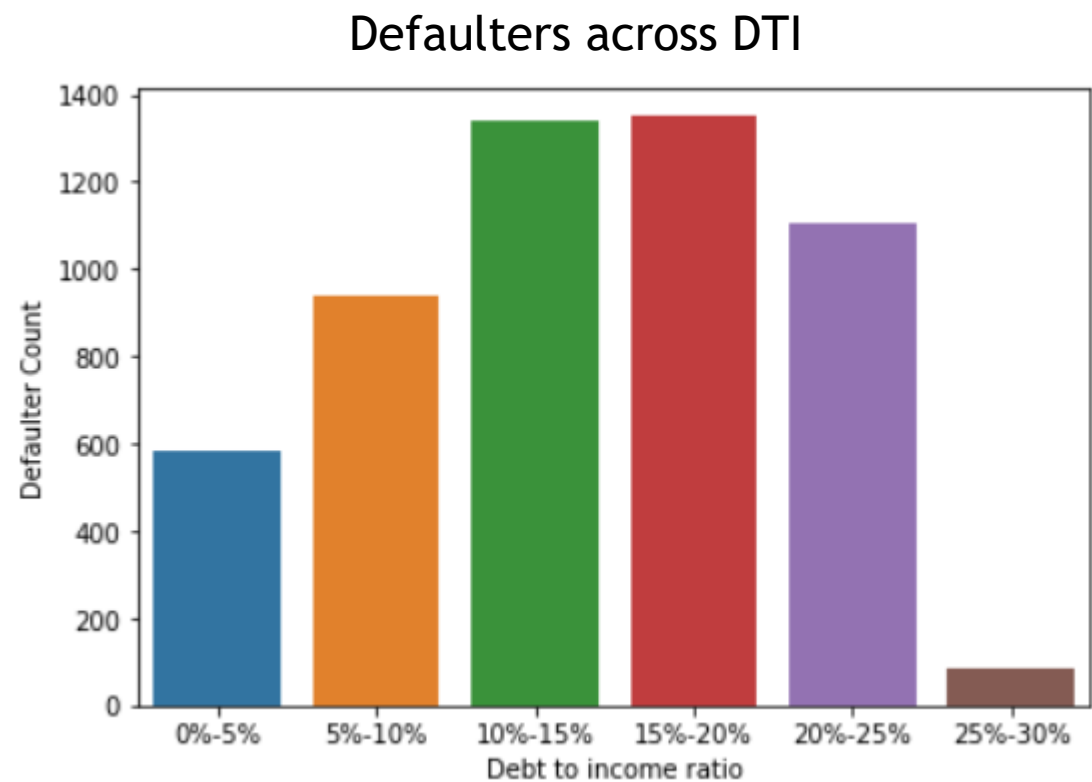
# 42% Defaulters had not have their stated income verified



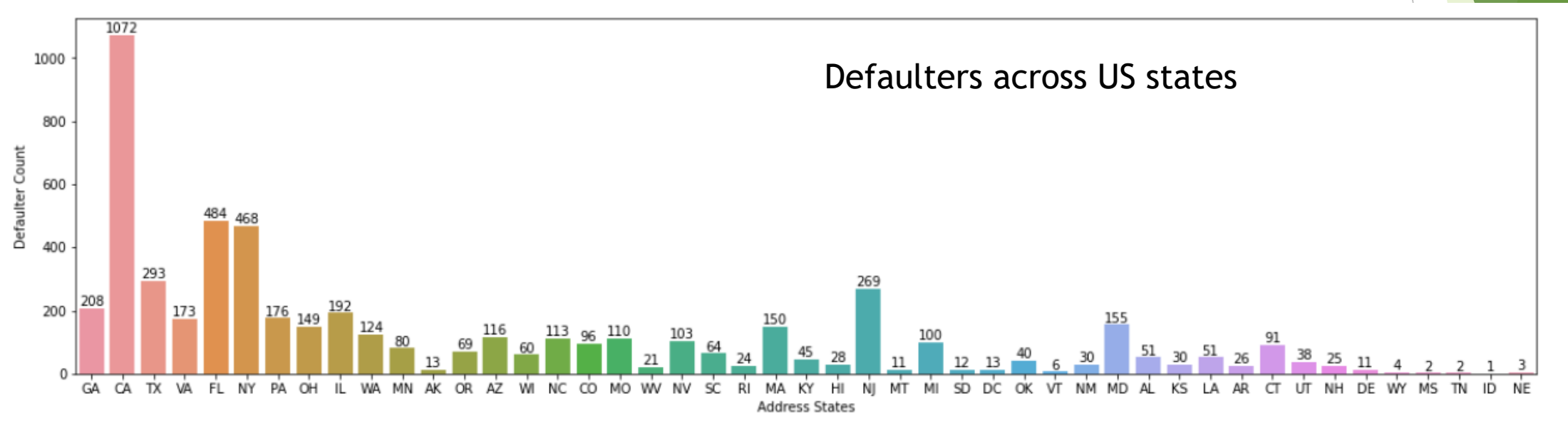
Loan Grade B had ~30 % of the defaulters, under which B5 was the worst performer



Around half of the total defaulters had Debt to income ratio in range 10%-20%



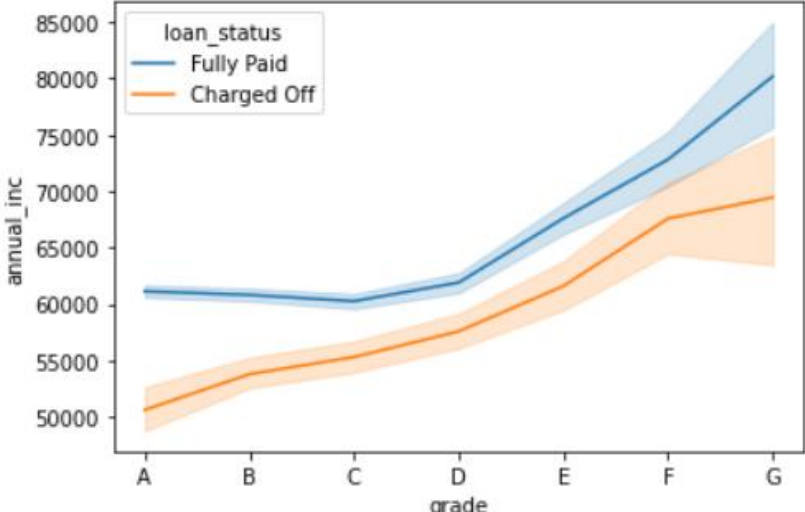
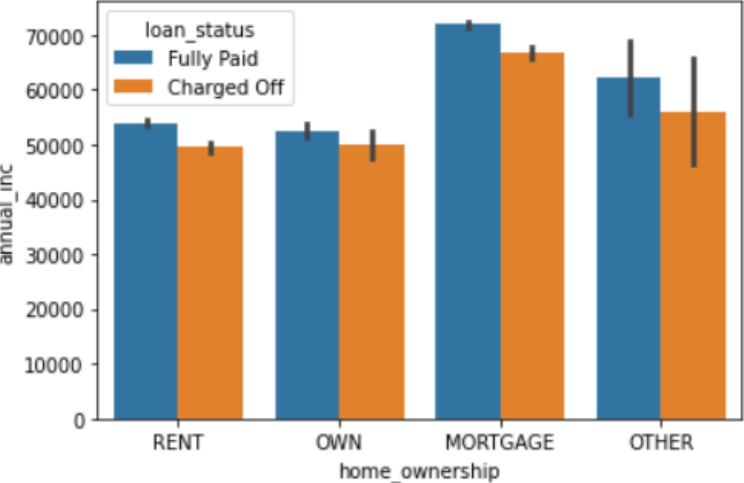
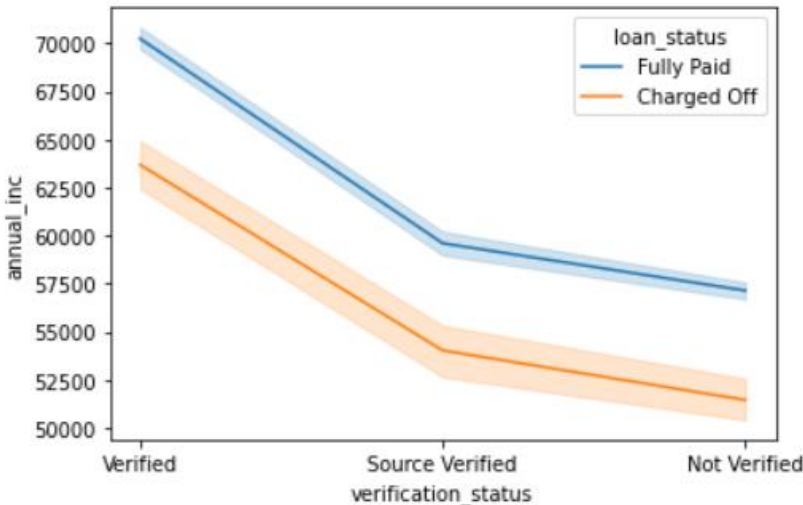
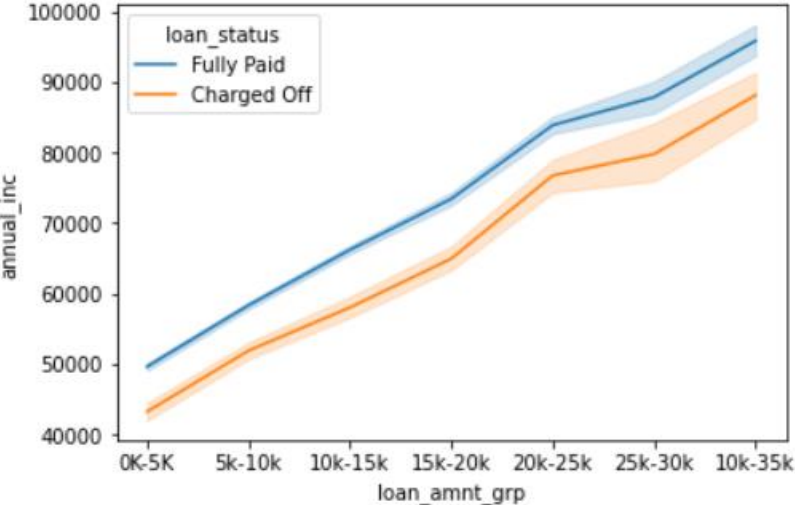
~20% defaulters are from California state



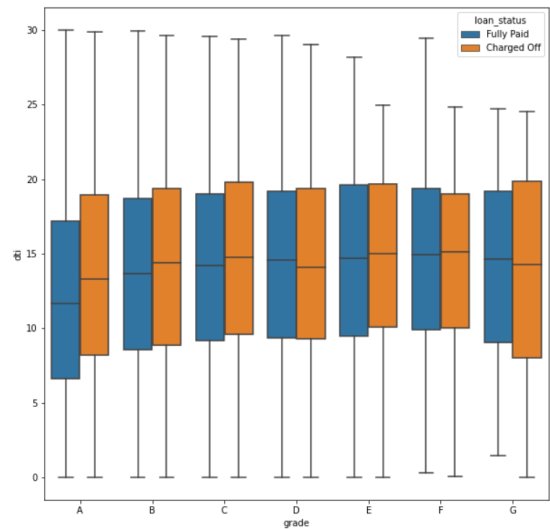
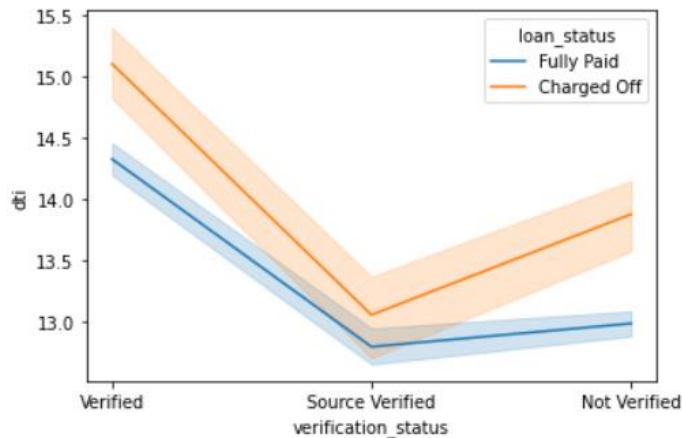
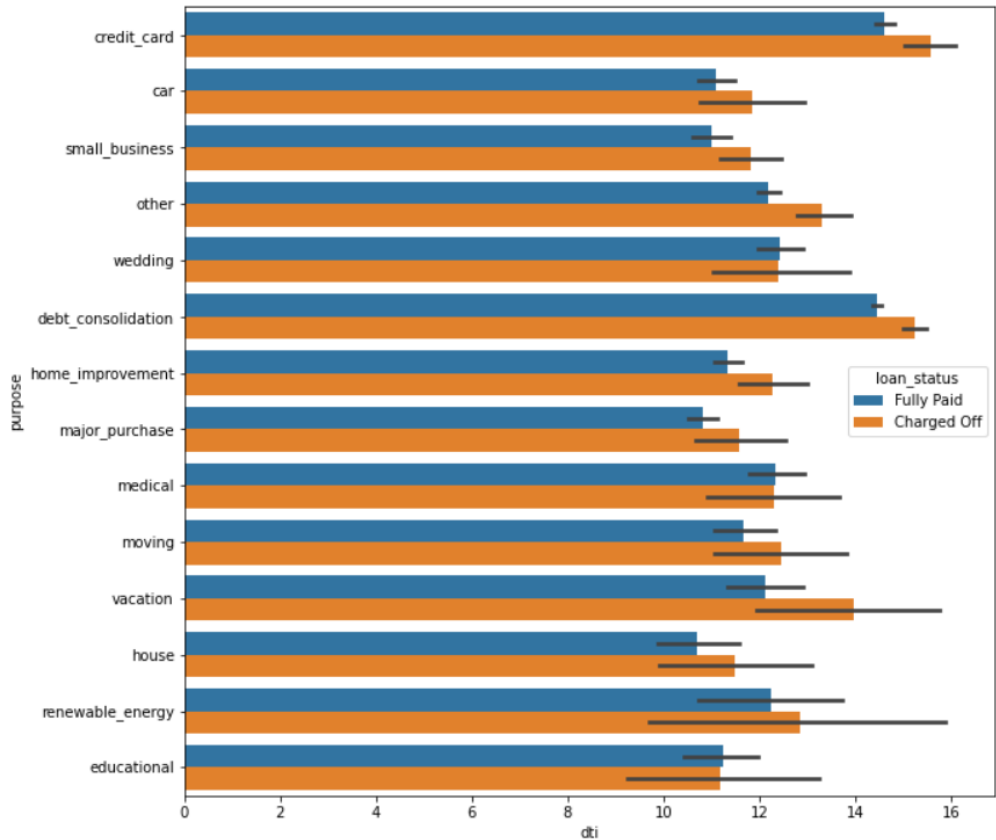


# BIVARIATE ANALYSIS

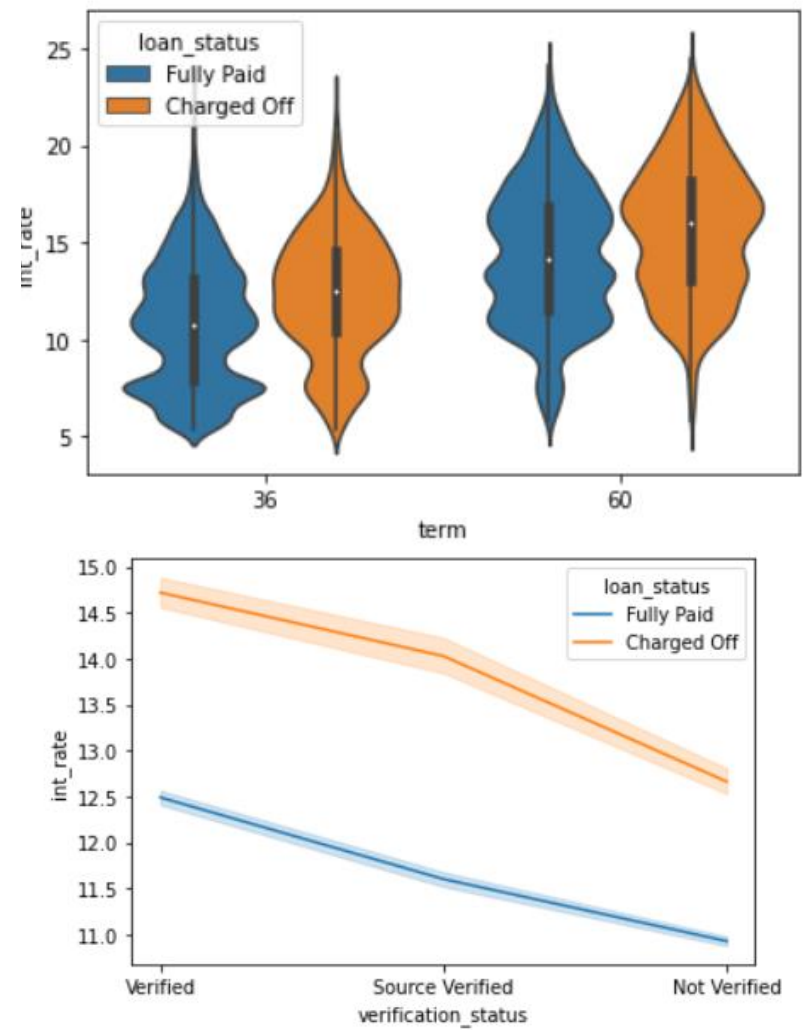
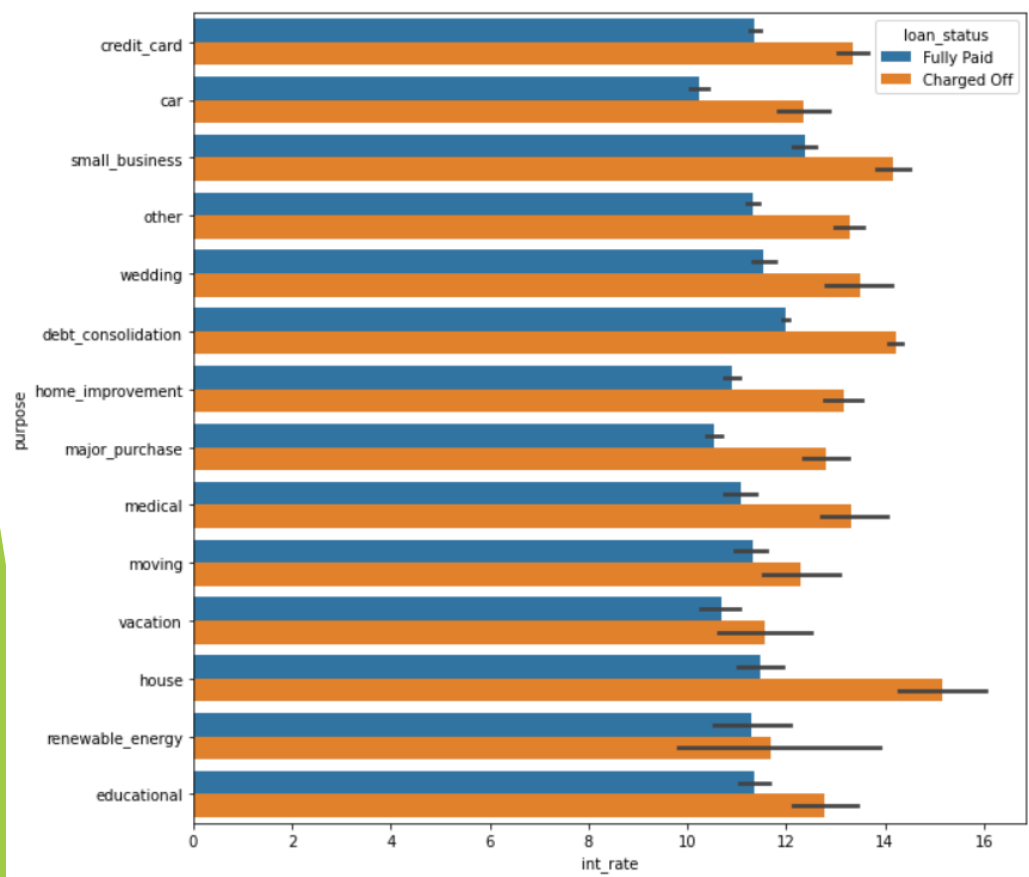
# Mean Annual Income across various variables is less for defaulters compared to “fully paid” customers



# Mean Debt to income ratio across various variables is high for defaulters compared to “fully paid” customers



Mean interest rate across various variables is high for defaulters compared to “fully paid” customers



# Recommendations

# Insights and Recommendation

Look for the following indicators for new applicants who are likely to default

- ▶ income is not verified
- ▶ People with annual income in range 30k-60k
- ▶ Purpose of loan is debt\_consolidation
- ▶ Debt to income ratio 10%-20%
- ▶ Home ownership as RENT or MORTGAGE
- ▶ Loan grade assigned as B

Thank You!!!!!!!!!!