Lending Club Case Study

- Ankita Maheshwari
- Nithin Bhardwaj

Objective

Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

The main objective is to be able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.

Identification of such applicants using EDA is the aim of this case study. Perform an analysis to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Dataset details

The data given contains information about past loan applicants and whether they 'defaulted' or not. Data has details regarding approved loan not the rejected ones. It has 3 status of loan which is –
1. Fully Paid
2. Current
3. Charged-Off.
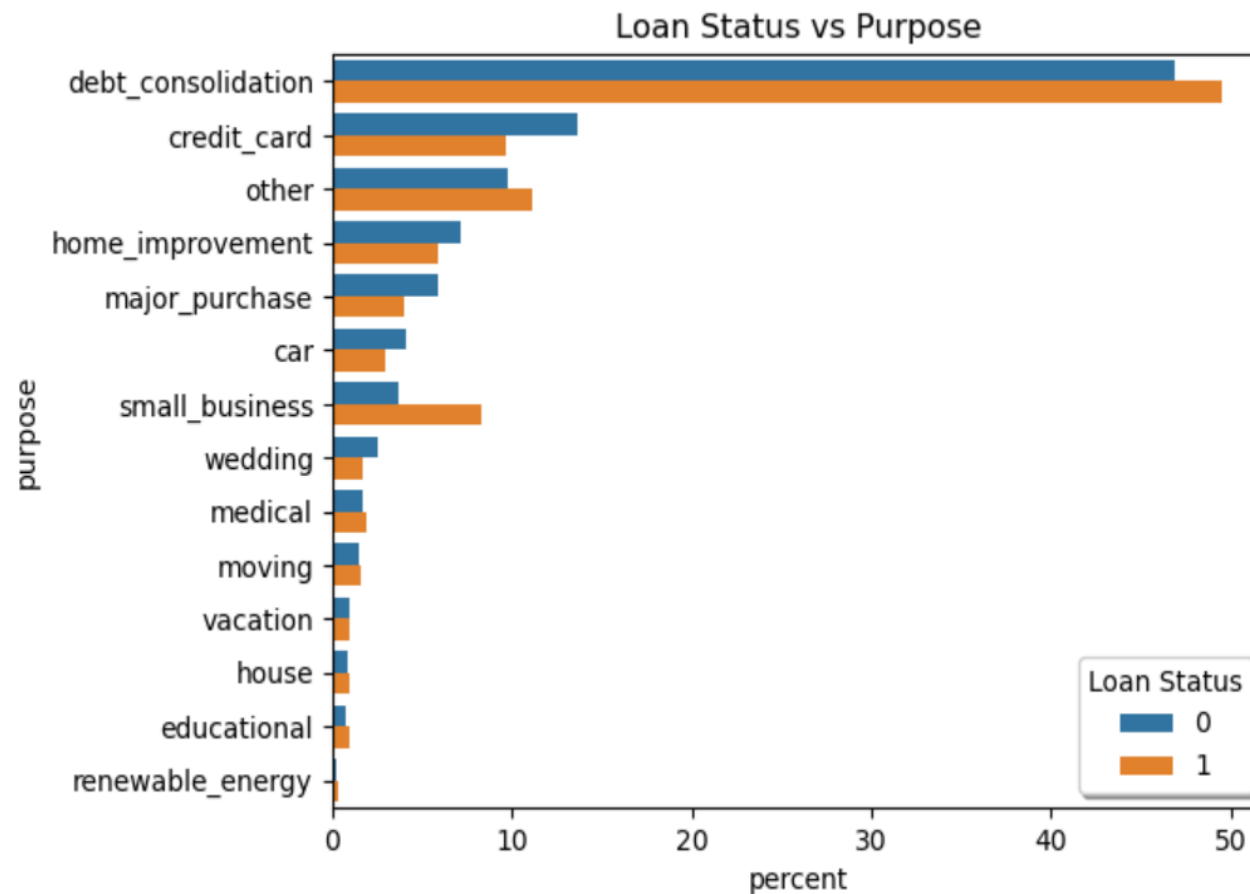
The case study is divided into 4 parts :
1. Data Understanding
2. Data Cleaning
3. Data Analysis
4. Recommendations

Bivariate Analysis

# Loan Status and Purpose

INFERENCE:

❑ Charged-Off accounts have high percentage for Small business & Other compared to Fully Paid

Loan Status and Term

INFERENCE:

❑45% Charged-Off accounts have opted for 60 months and remaining 55% of Charged Off accounts have opted for 36 Terms

❑22% Fully Paid accounts have opted for 60 months and remaining 78% Paid Off accounts have opted for 36 Terms

❑Based on this we could infer that, People opting for 60 Months are tend to have high probability of Charge Off as compared to fully paid
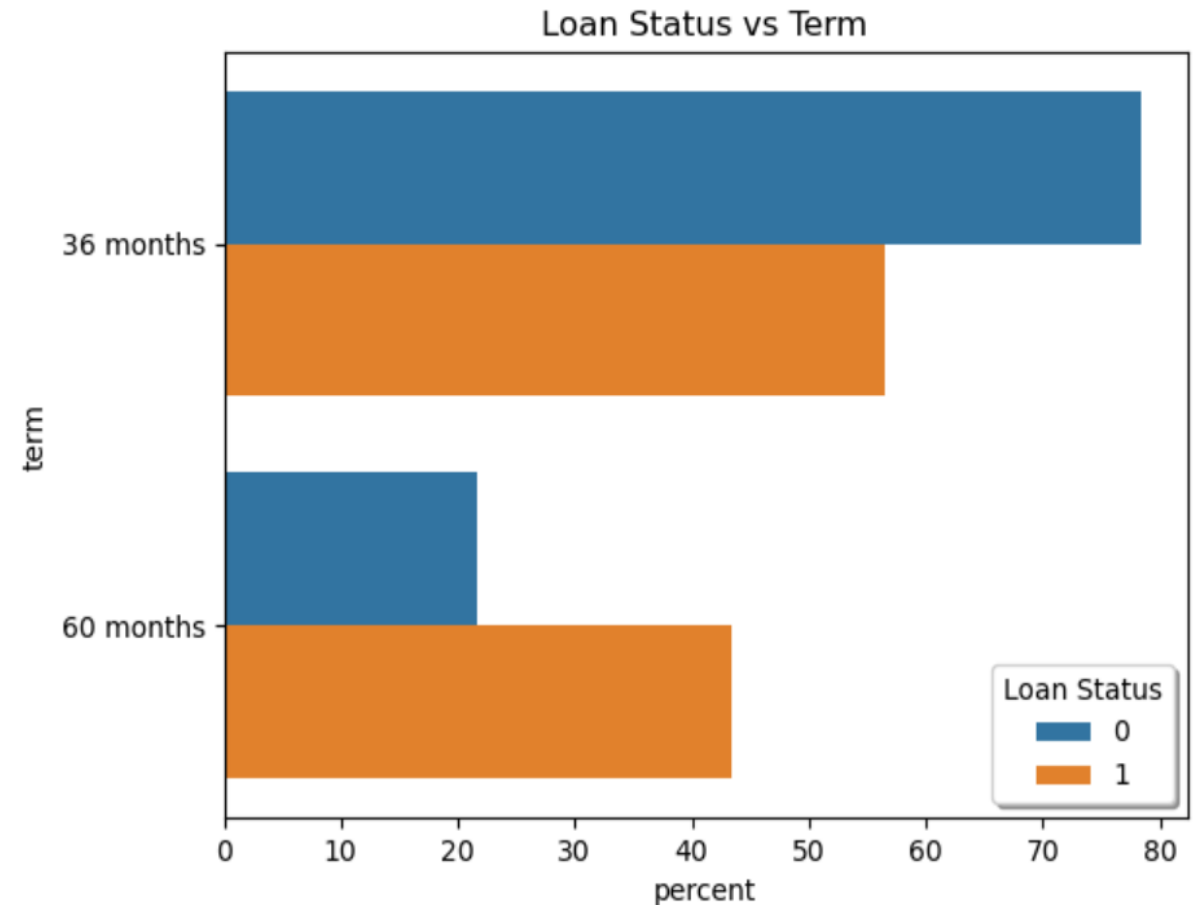
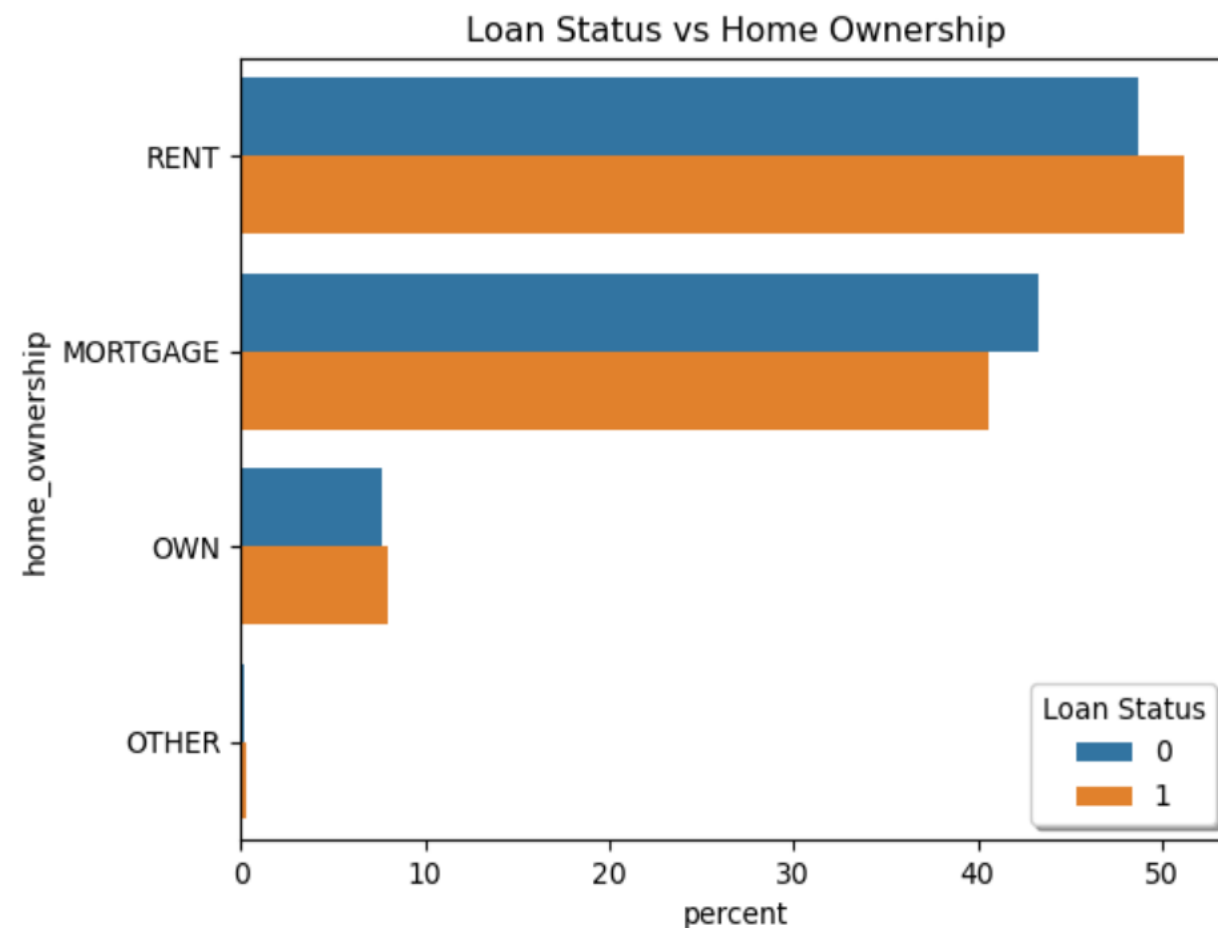# Loan Status and Verification Status

INFERENCE:

❑From the above graph, we can infer that Verified accounts have high percentage of Charge Off accounts compared to Fully Paid

# Loan Status and Home Ownership

INFERENCE:

❑People living in a rented house have slightly high percentage of Charge Off compared to Fully Paid.

Loan Status vs Home Ownership

Loan Status and Grade

INFERENCE:

❑65% Charge Off accounts belong to Grade C,D,E,F where as only 42% of Fully Paid Accounts belong to these Grades

❑14% Charge Off accounts belong to Grade A where as 28% of Fully Paid Accounts belong to Grade A

❑Hence, people belong to Grade A has low probability of becoming Charged-Off, whereas people belonging to Grade C,D,E,F has high probability of becoming Charged-Off

Loan Status vs Grade
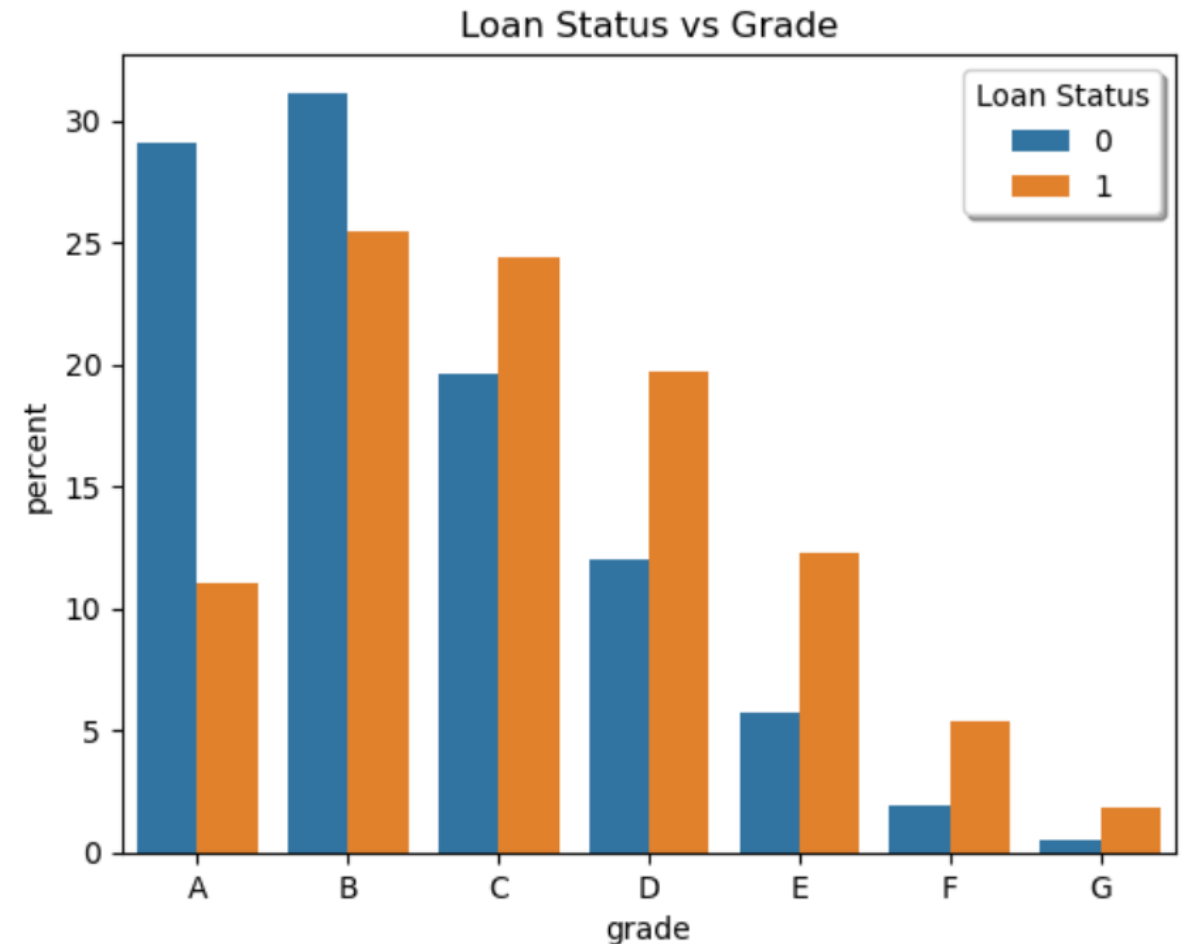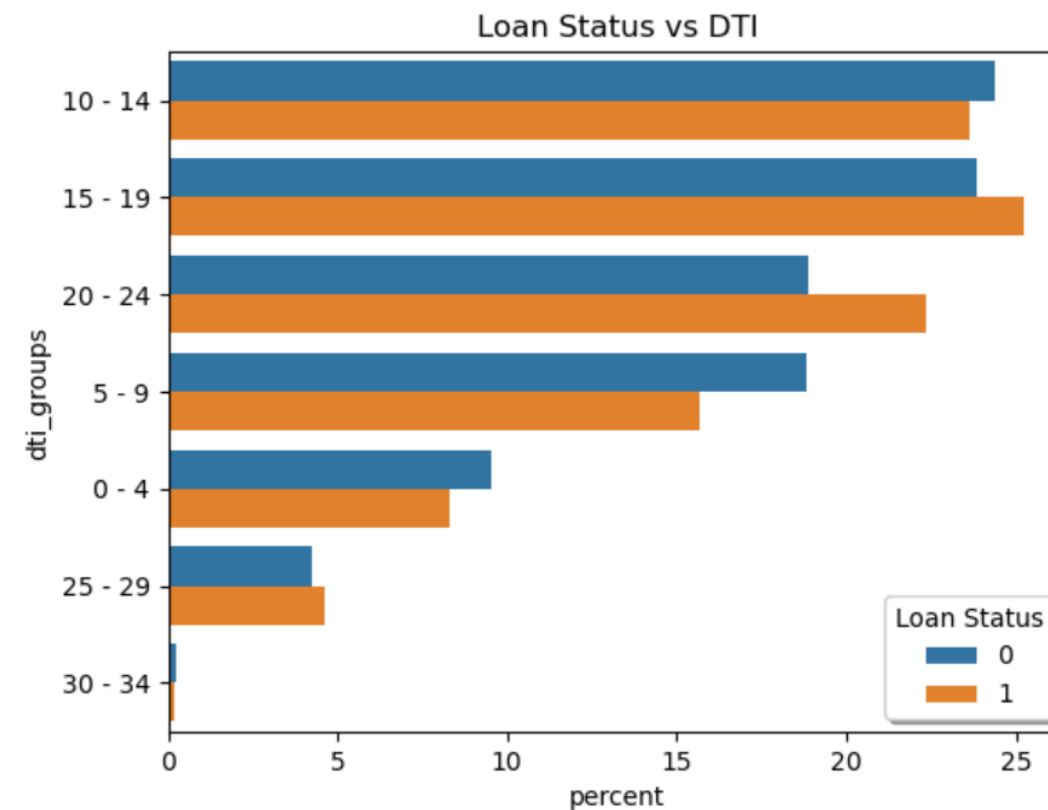
Loan Status and Annual Income

INFERENCE:

❑Annual Income in the range 0-20K, 20k-40k, 40k-60k has high percentage of Charge Off accounts compared to Fully Paid Accounts have high probability of Charge-Off
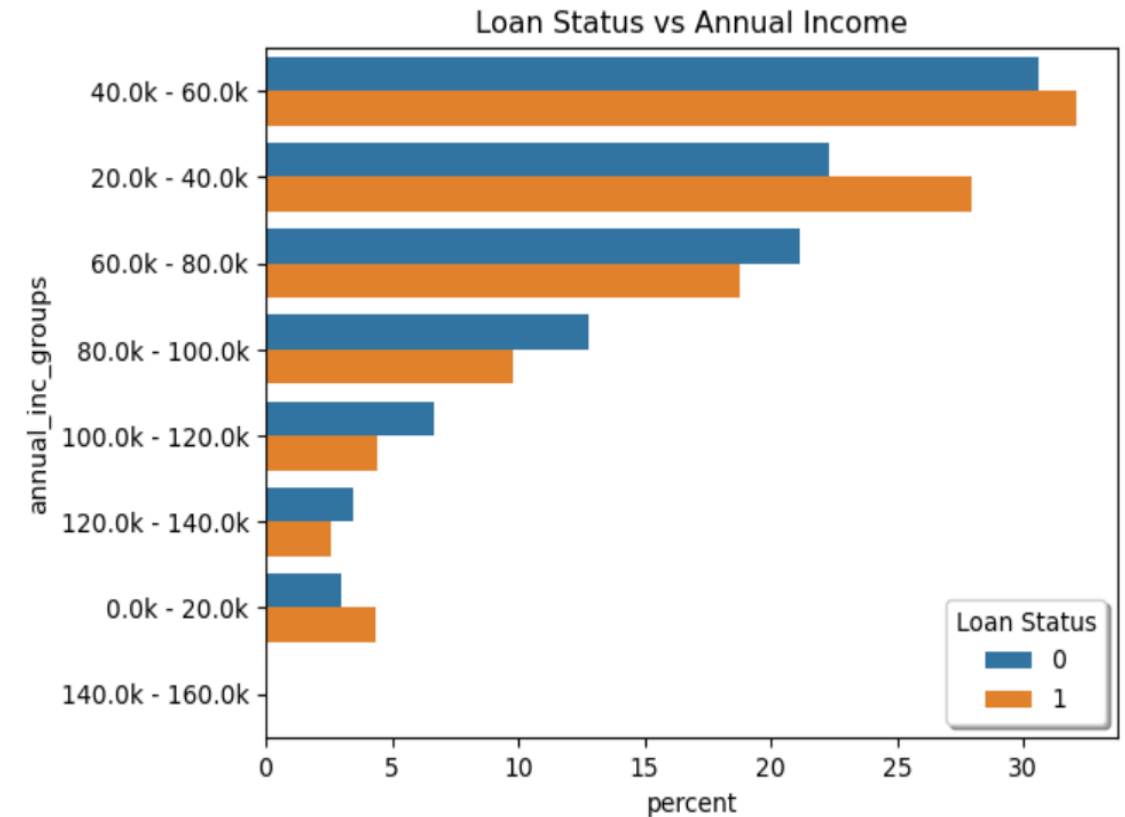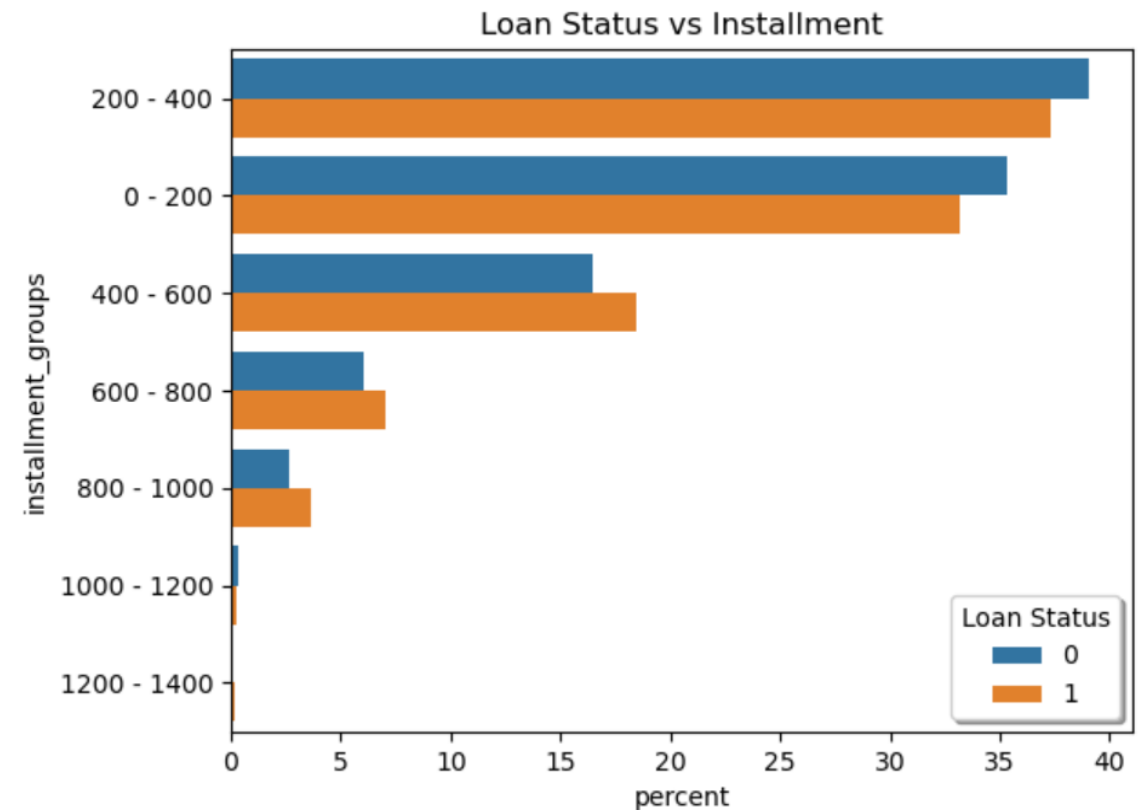
Loan Status and DTI

INFERENCE:

❑DTI, in the range of 15-19 and 20-24, has high percentage of Charge Off accounts compared to Fully Paid Accounts

Loan Status vs DTI
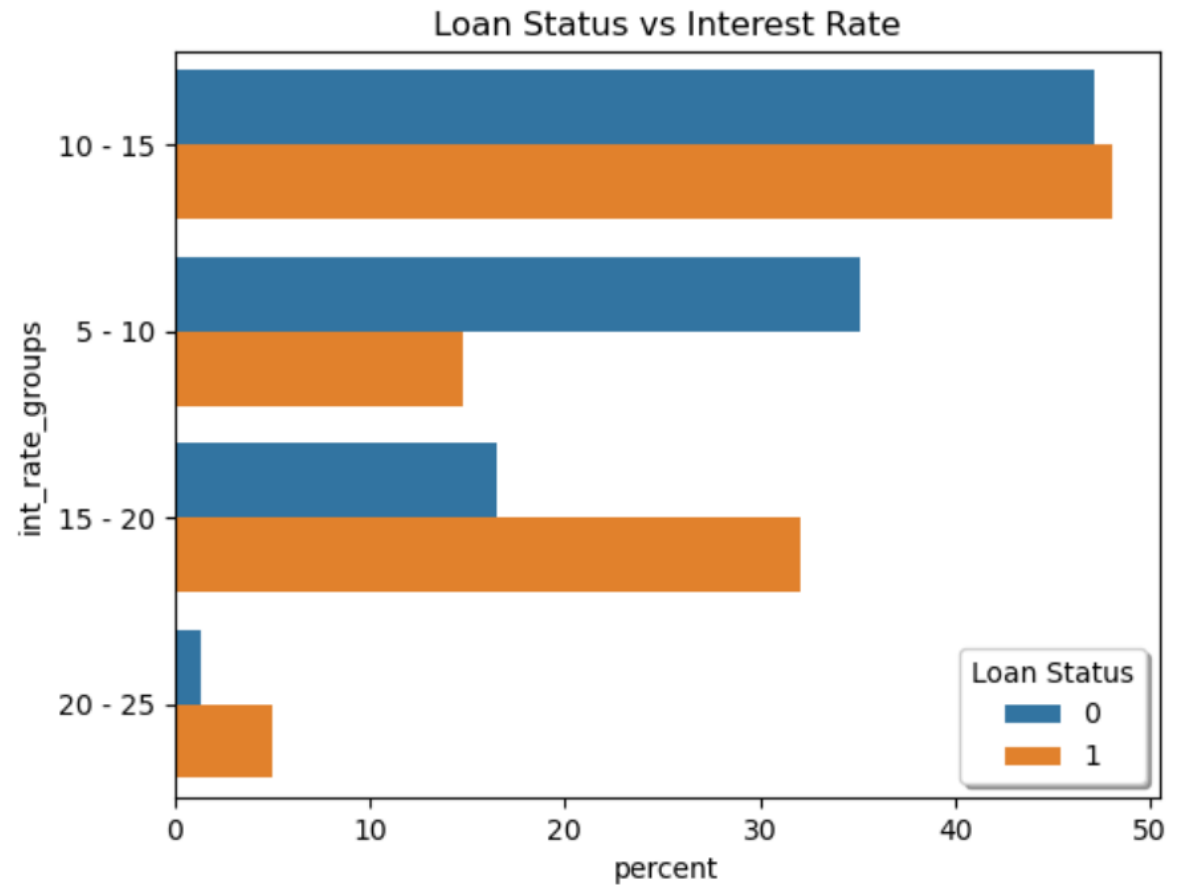
Loan Status and Installment

INFERENCE:

❑Percentage of Charge Off accounts is higher with installment greater than 400 compared to Fully Paid

Loan Status and Interest Rate

INFERENCE:

❑40% of the Charge Off accounts have interest rate in the range of 15-25% where as only 20% of Fully Paid accounts have interest rate in this range.

❑Accounts having interest rate in the range of 15-25% has high probability of becoming Charged-Off

Loan Status vs Interest Rate

Loan Status and Amount Invested

INFERENCE:

❑Percentage of Charge Off accounts is high when the amount funded by investor is in the range of 15000-25000

Loan Status vs Funded Amount Invested
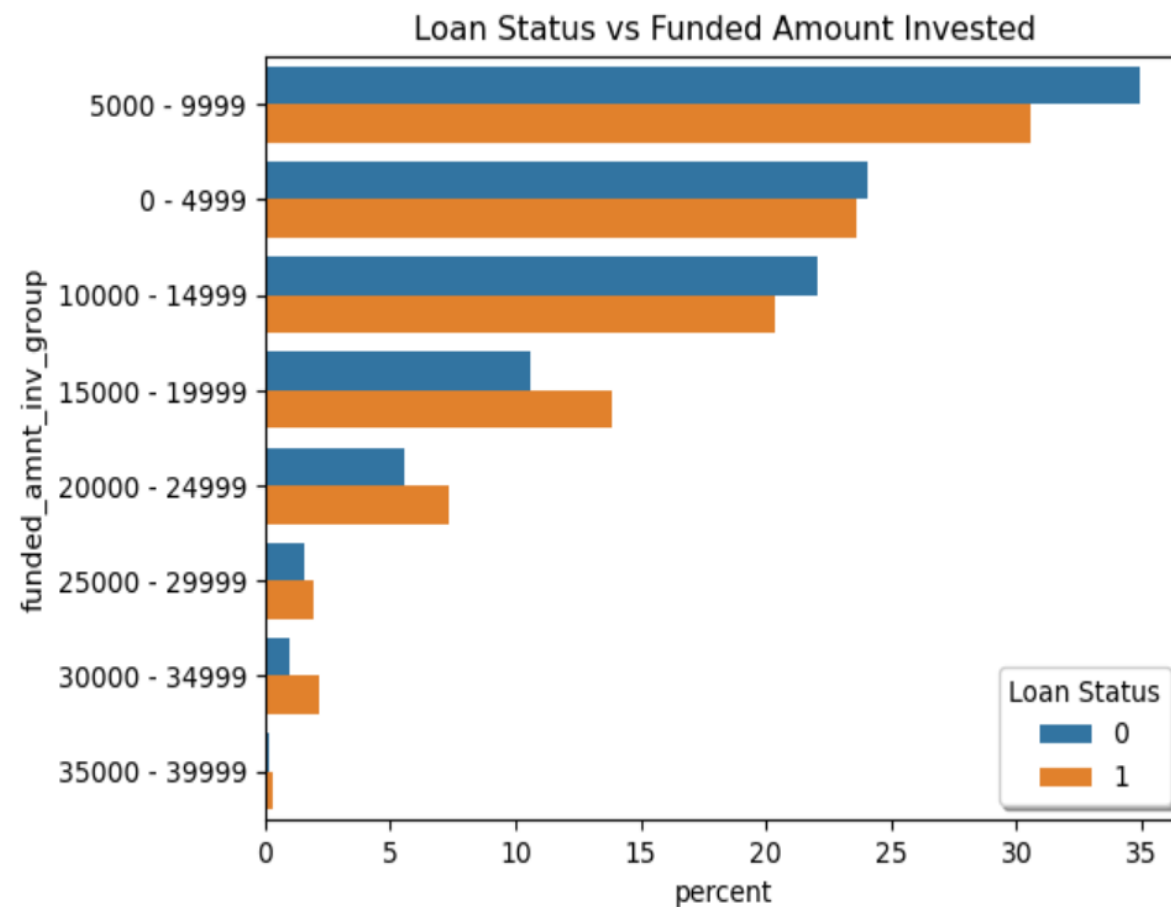
Loan Status and Loan Amount

INFERENCE:

❑Percentage of Charge Off accounts is high when the amount applied for loan is greater than 14999 compared to Fully Paid
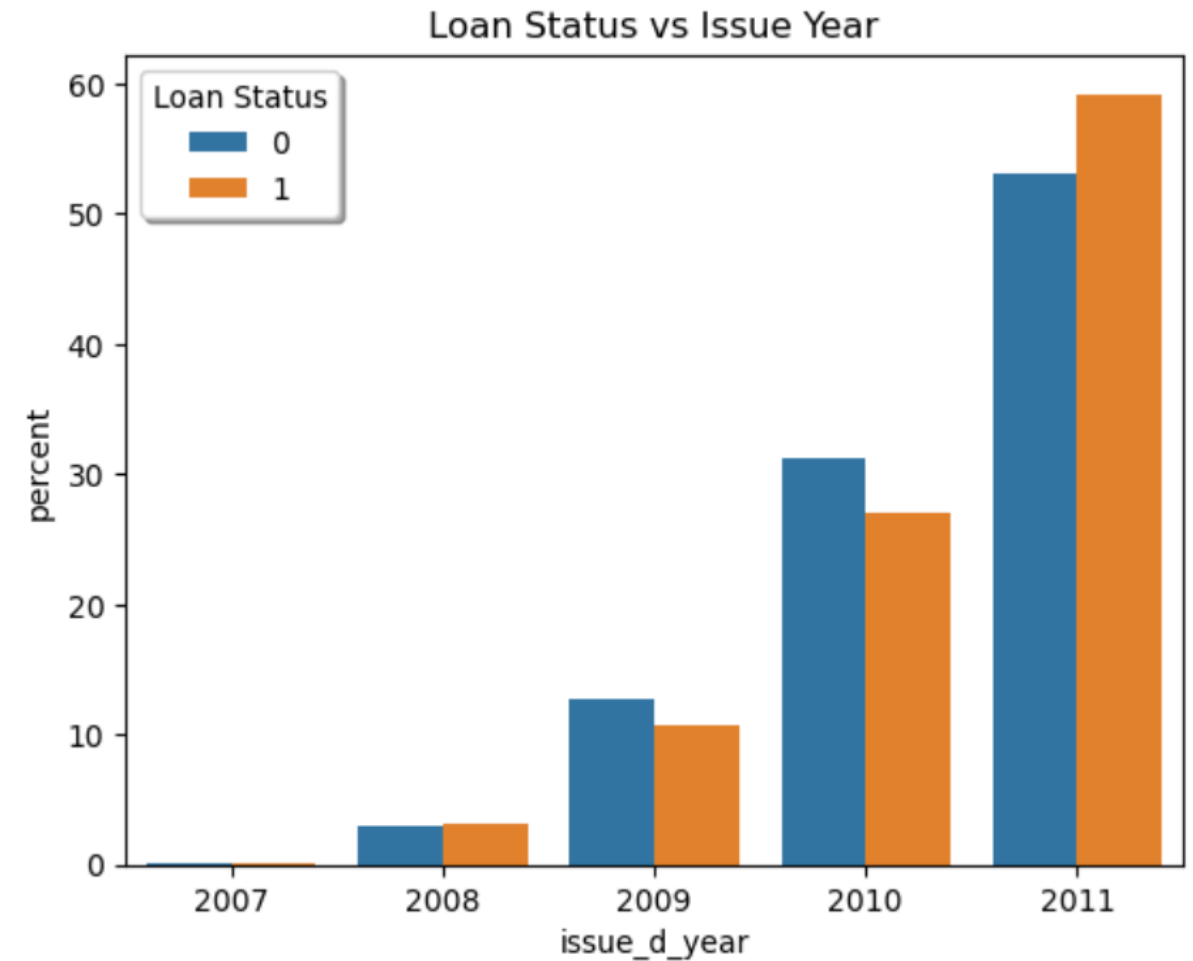

Loan Status vs Loan Amount
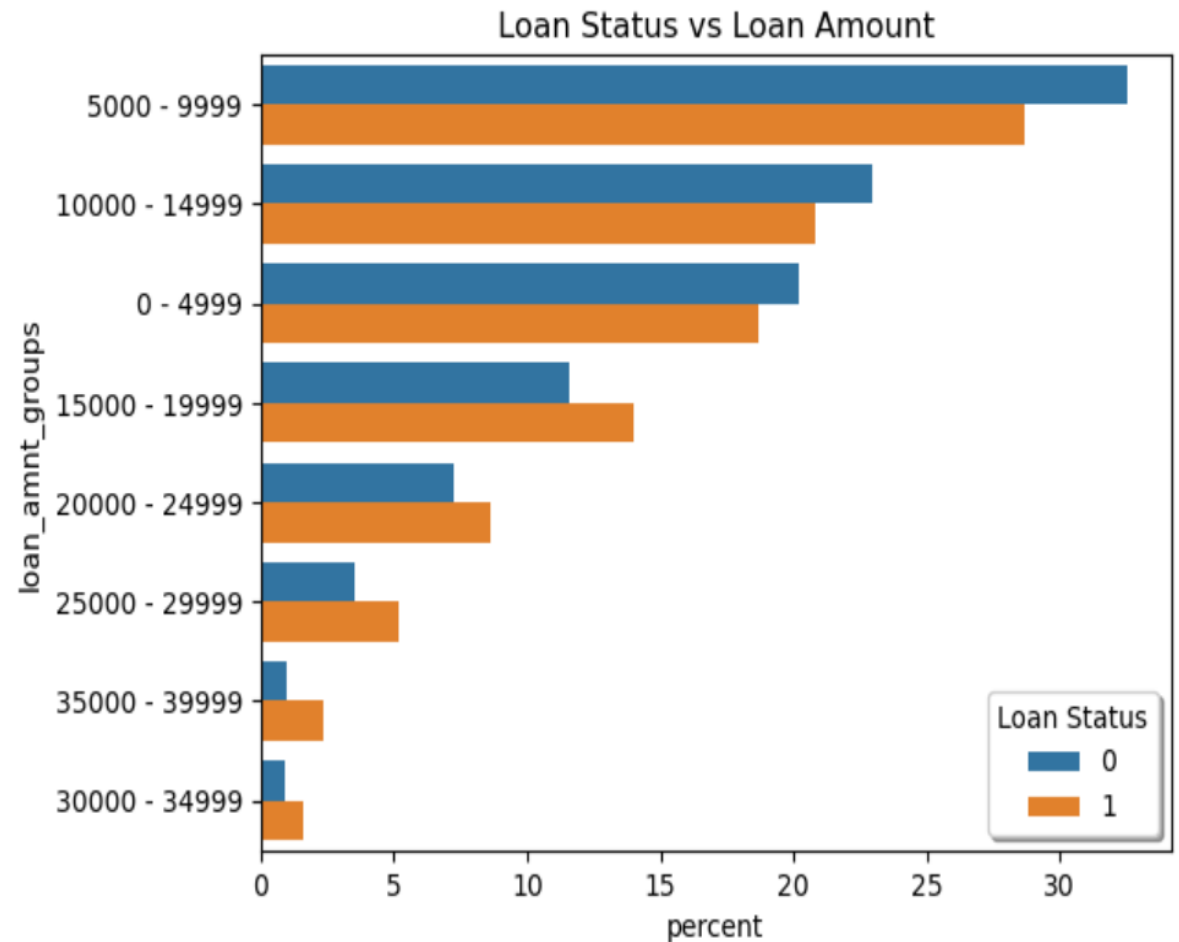
Loan Status and Employee Experience

INFERENCE:

❑Percentage of Charge Off accounts is highest when the employee experience is 10+

Loan Status vs Employee Experience
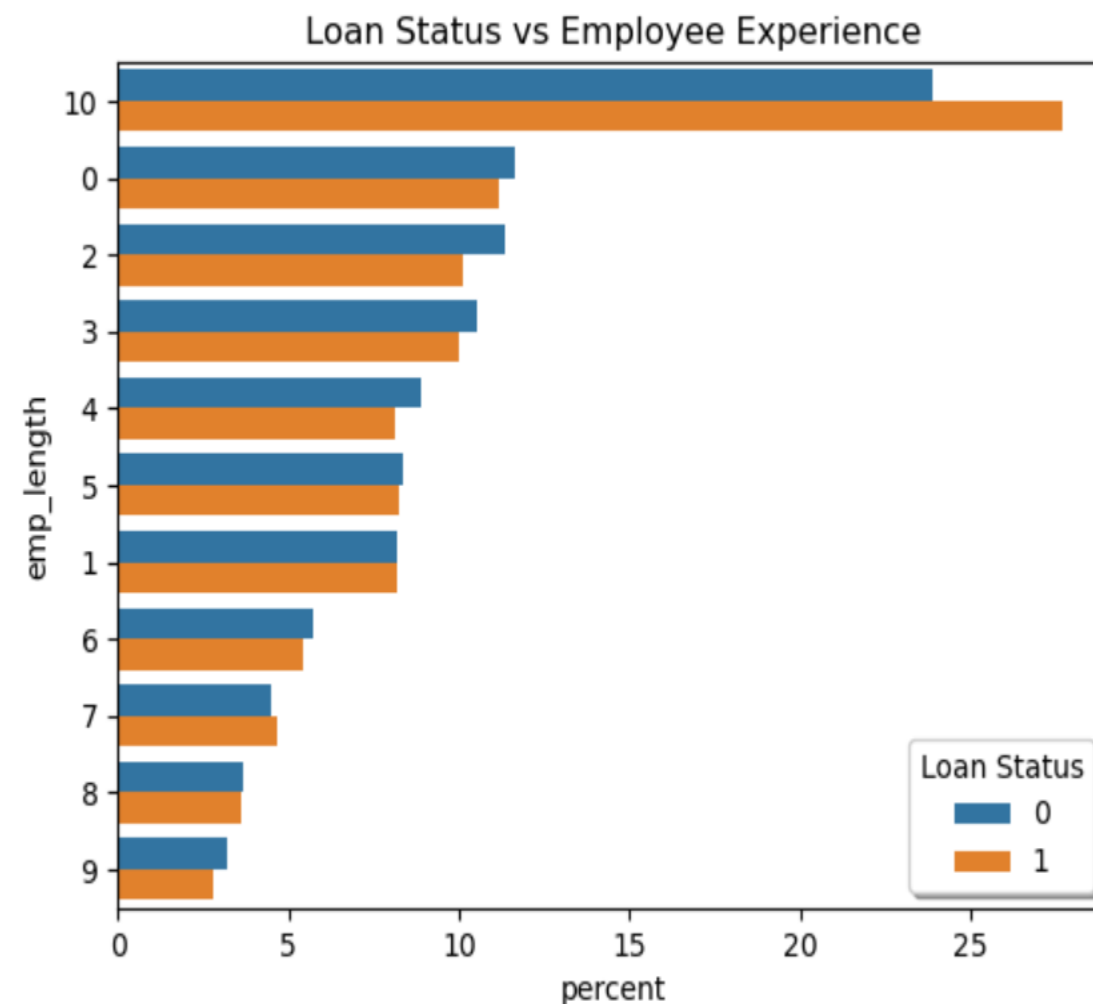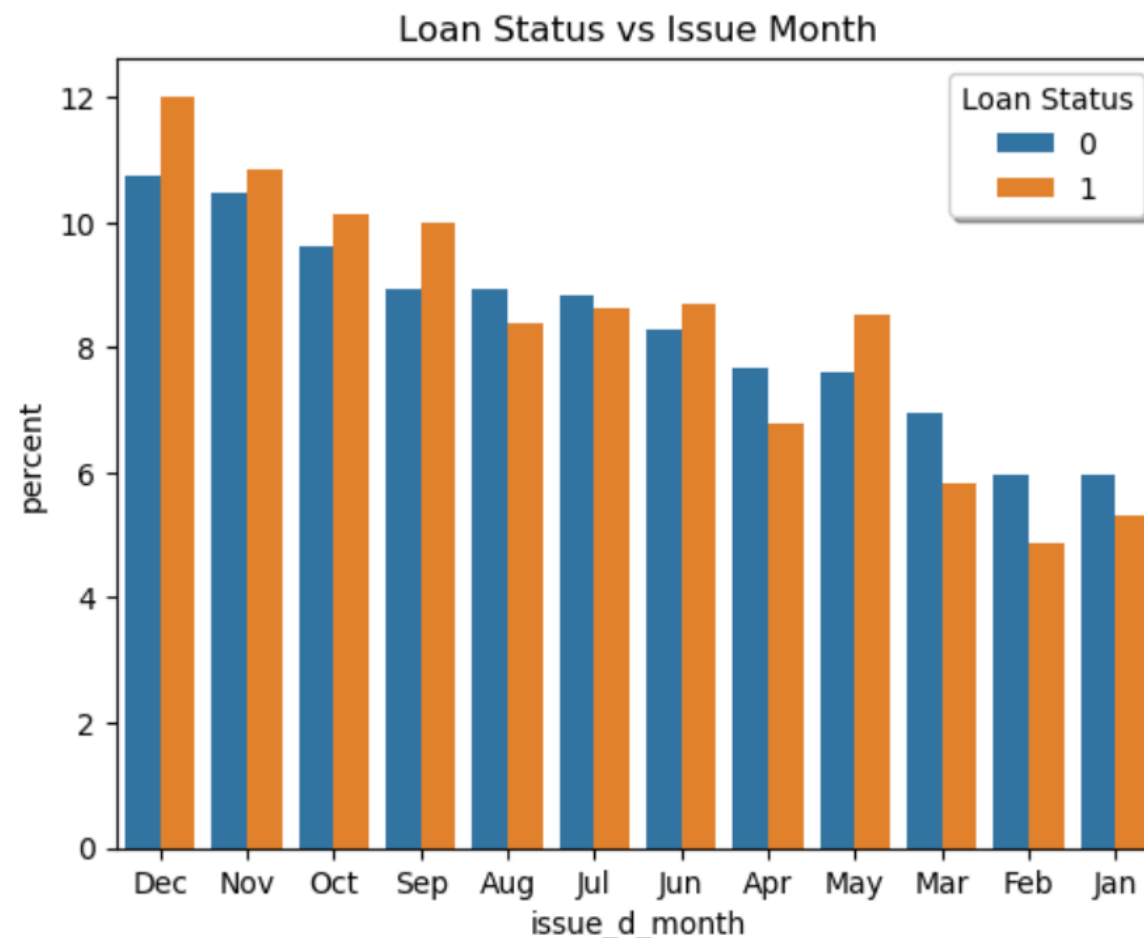
Loan Status and Issue Year

INFERENCE:

❑ Percentage of Charge Off accounts is slightly high in 2011

Loan Status vs Issue Year
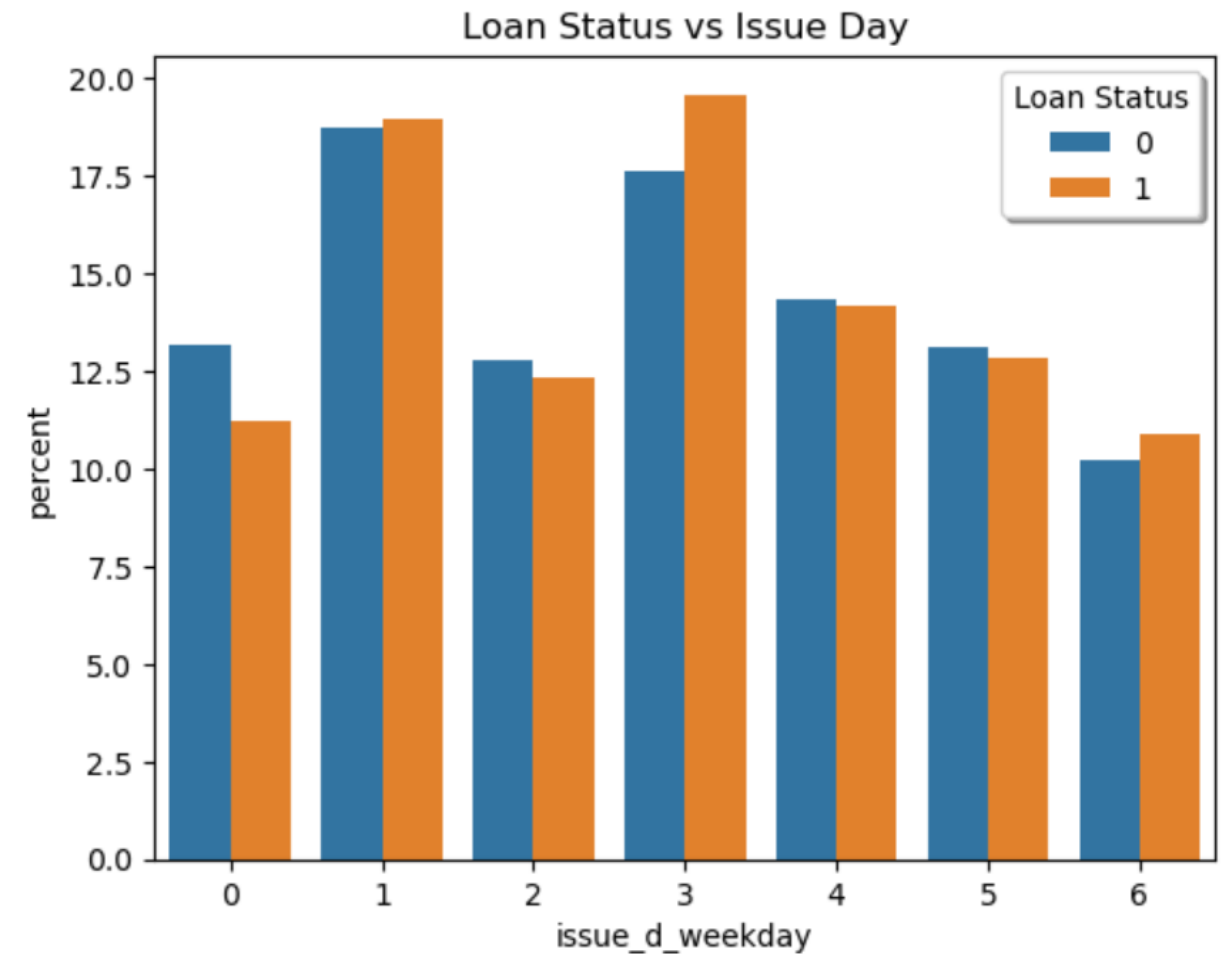
Loan Status and Issue Month

INFERENCE:

❏Percentage of Charge Off accounts is slightly high in May, Jun, Sep, Oct, Nov and Dec compared to Fully Paid
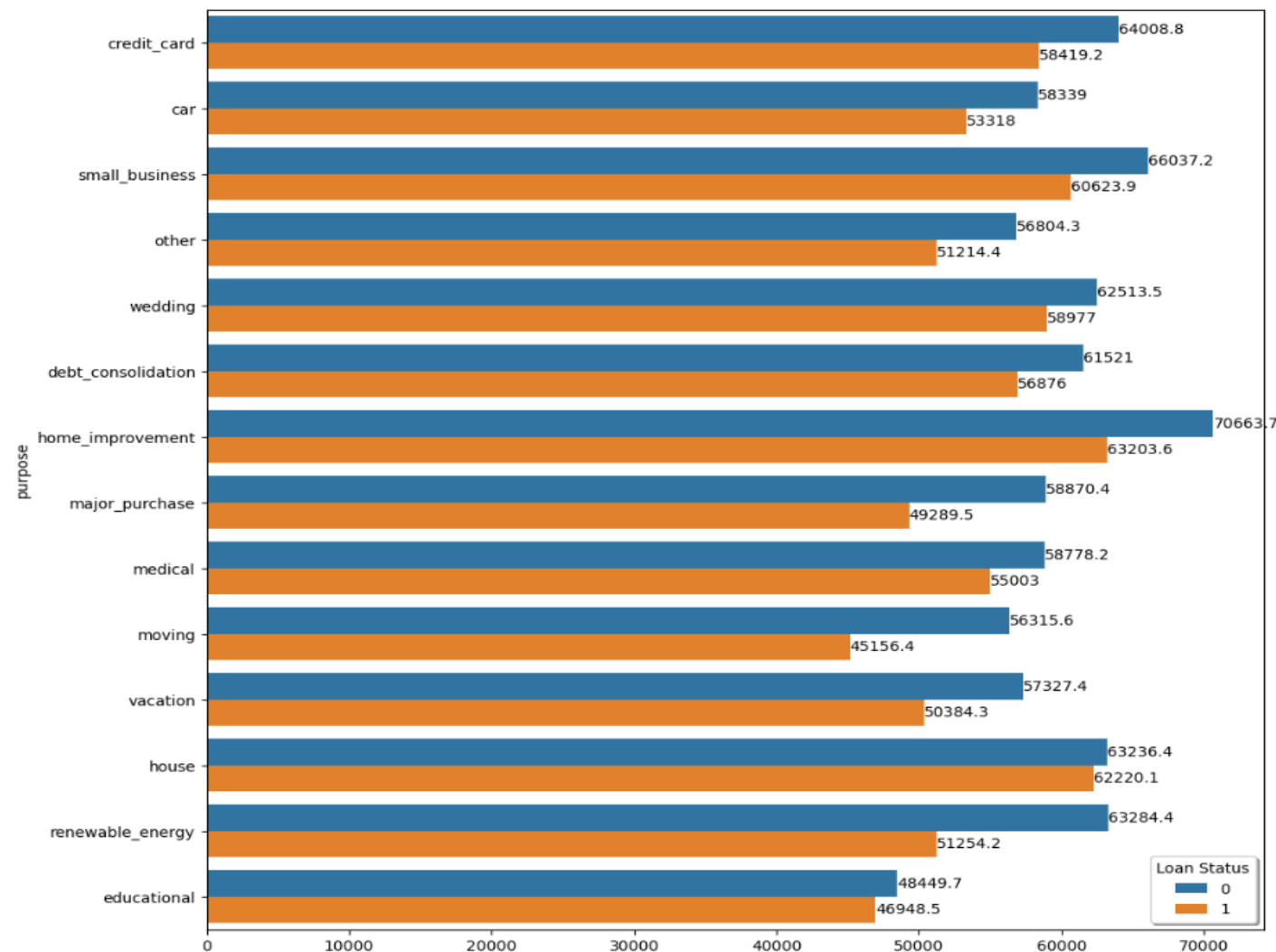
Loan Status and Issue Day

INFERENCE:

❑ percentage of Charge Off accounts is slightly high on 3 and 6 week day
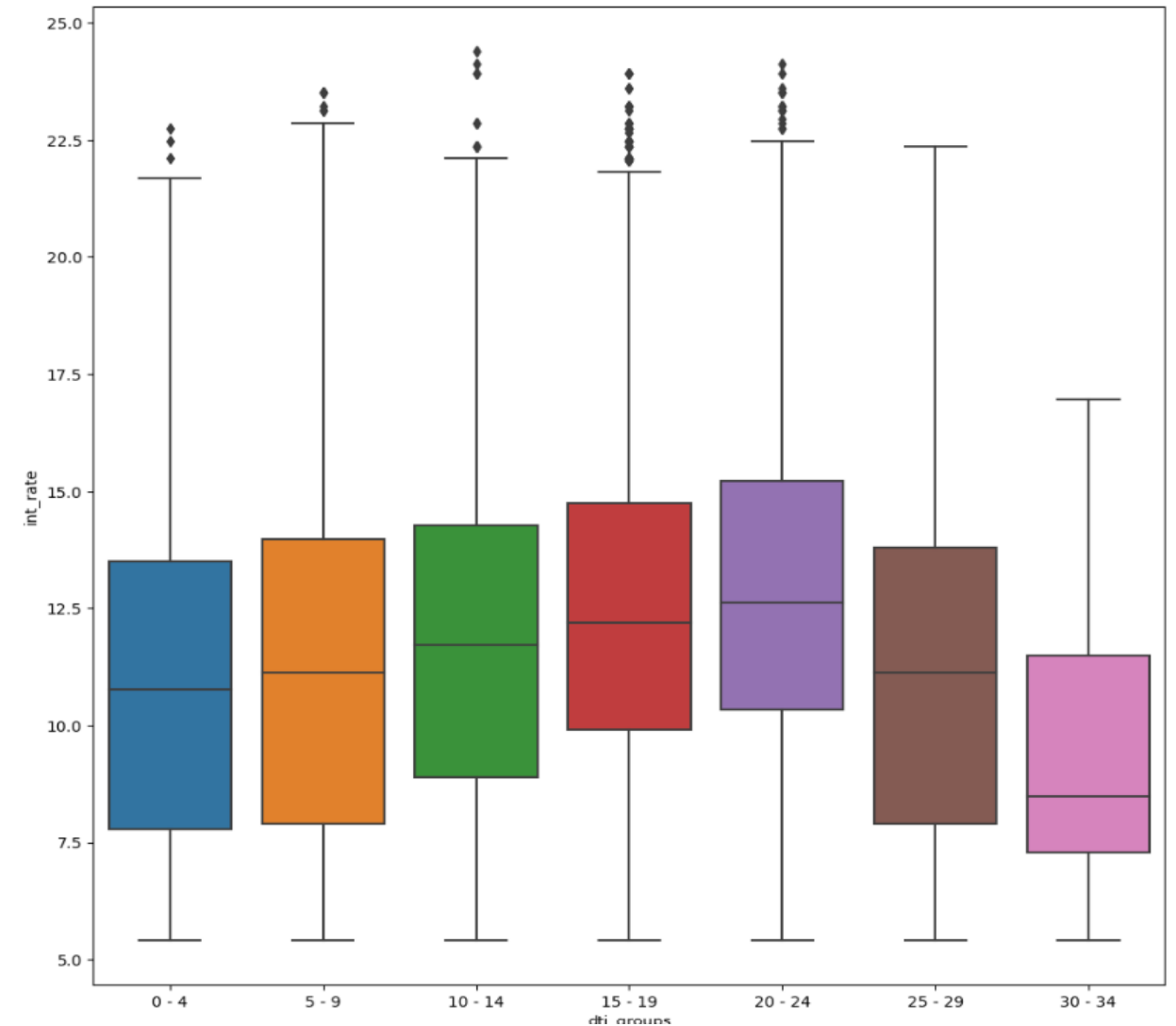
Annual Income and Purpose

INFERENCE:

❑Applicants with higher salary mostly applied loans for "home_improvment", "house", "renewable_energy" and "small_businesses"
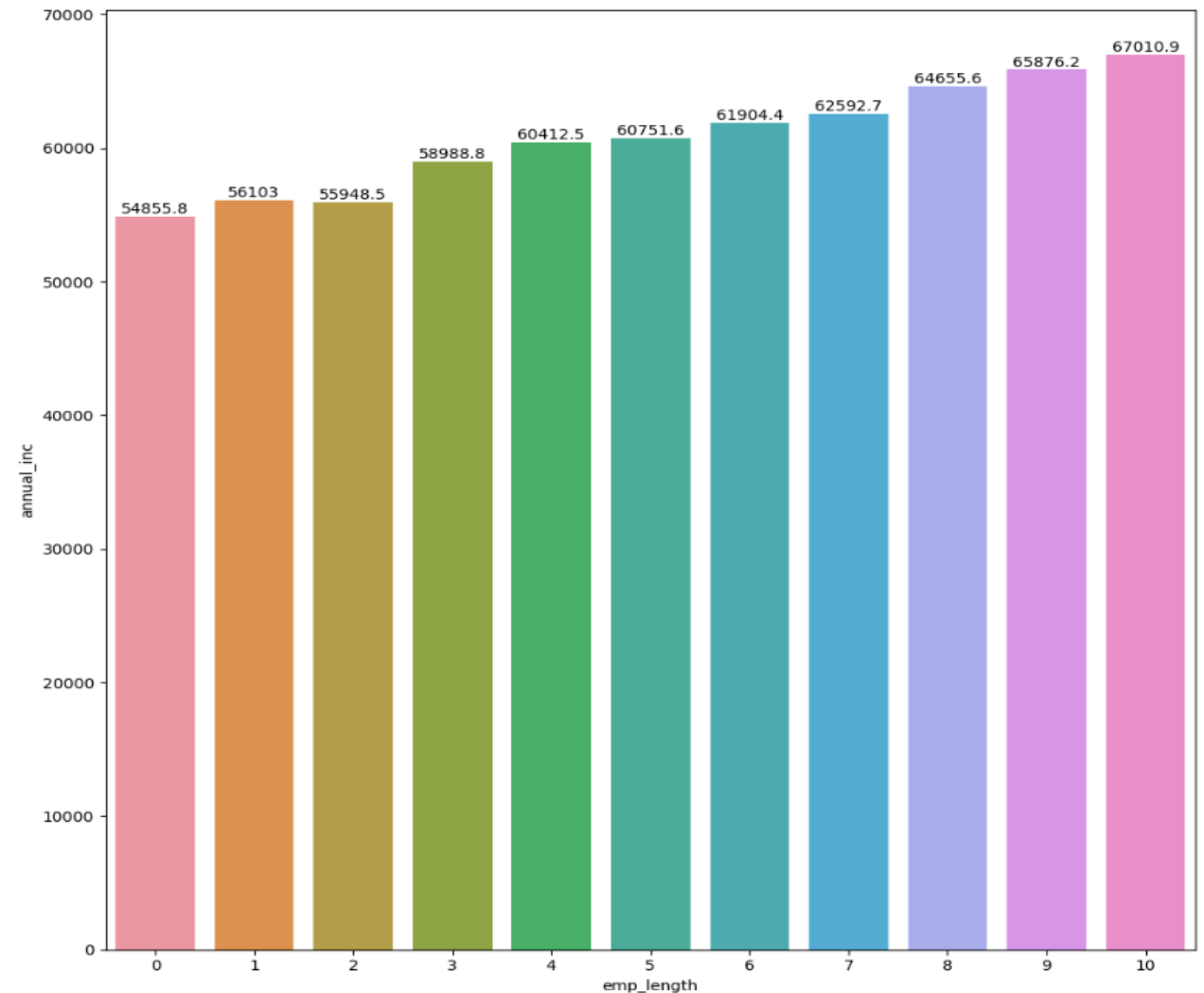
DTI and Interest Rate

INFERENCE:

❑Mostly DTI is increasing with increase in interest rate

Annual Income and Employee Experience
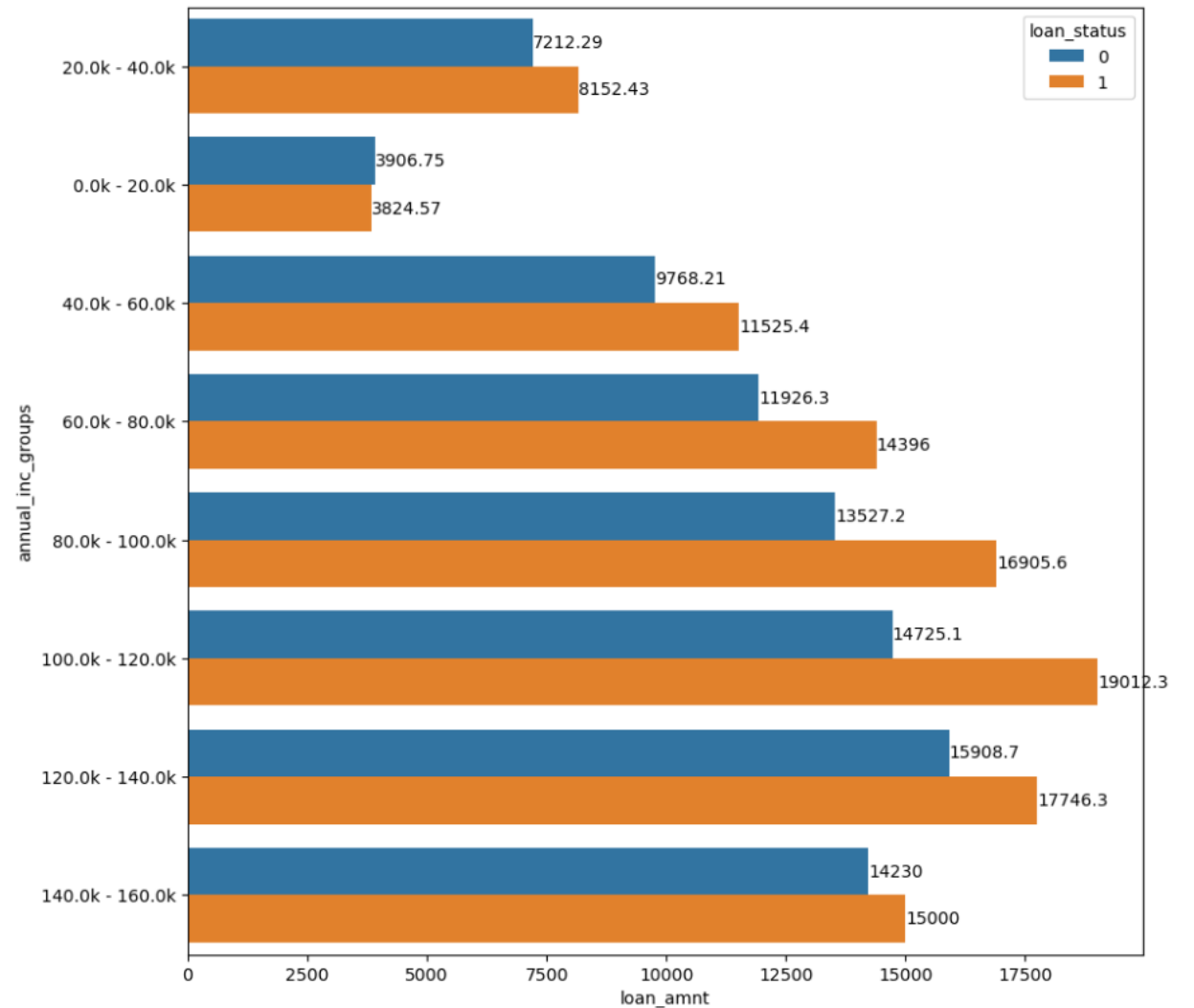
INFERENCE:

❑Annual Income is increasing with increase in Experience

Annual Income and Loan Amount
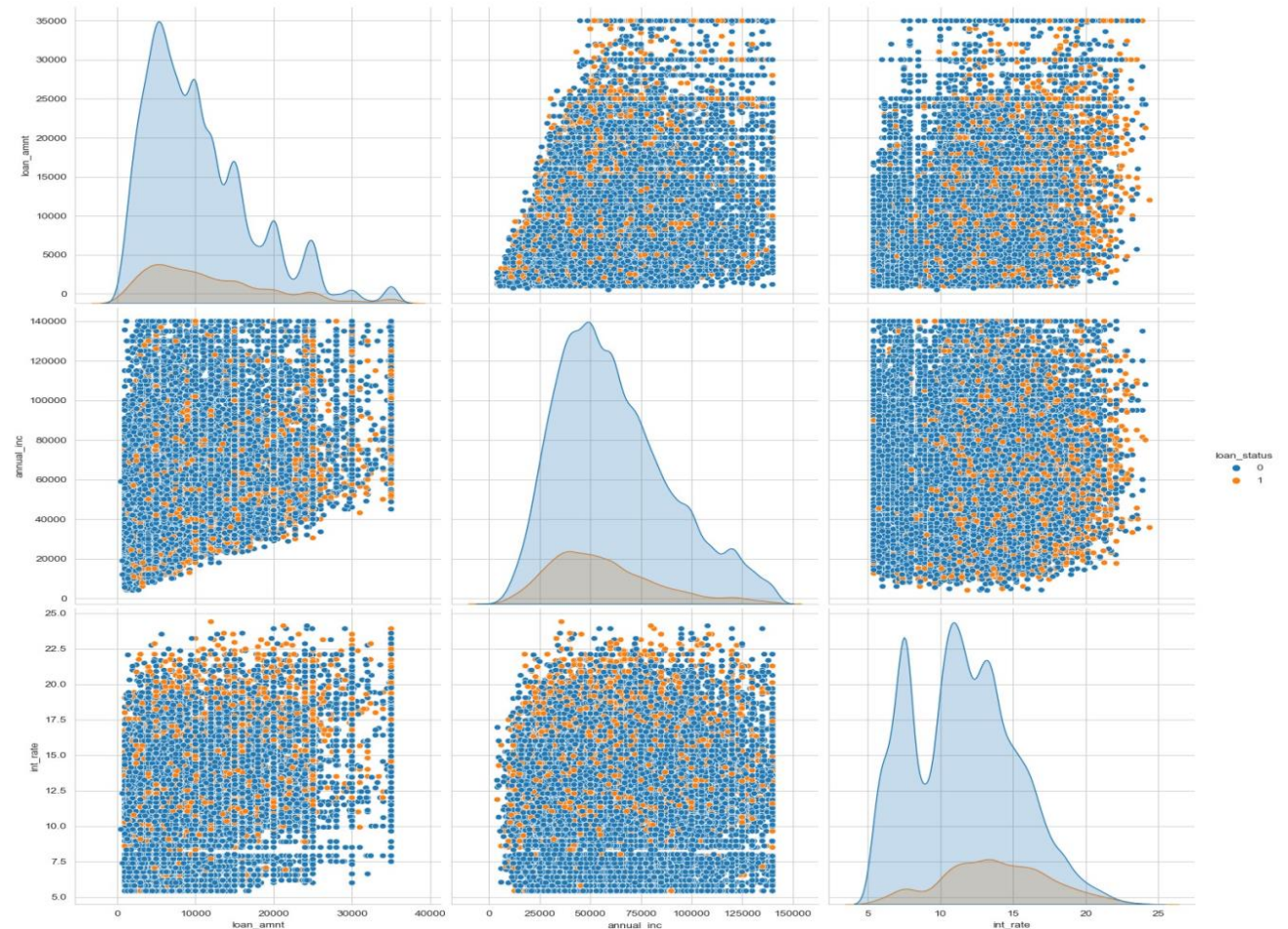
INFERENCE:

❑ Across all the income groups, the loan amount is higher for people who defaulted.

Multivariate Analysis using Pair Plots on Numeric Columns

INFERENCE:

❑ Higher the annual income higher the loan amount slightly.

❑ Interest rate is increasing with loan amount increase

Correlation Matrix

INFERENCE:

❑ Loan Amount & Funded Amount has highest positive correlation

❑ There are variables that have negative correlation also Ex: annual_inc vs dti

❑ There are variables that almost have zero correlation Eg:public_rec_bankruptcies vs dti

Recommendations

❑ Our analysis with respect to the charged off loans vs Full Paid Loans for each variable suggests the following. There is a more probability of defaulting when :

➢ Applicants having house ownership as 'RENT'

➢ Applicants who use the loan to clear other debts or who have applied for small business

➢ Applicants who receive interest at the rate of 15-25%

➢ Applicants who have an income of range 0k -60k

➢ Applicants with employment length of 10 +

➢ When funded amount by investor is between 15000-19999

➢ Loan amount is between 15000-29999

➢ DTI is between 15-19

➢ When monthly installments are between 400-600

➢ Term of 60 months

➢ When the loan status is Verified

➢ Grade is 'C','D','E','F','G'

➢ Issued on weekday no 3 and 6

➢ Issued in Dec month

➢ Issued in 2011 year