# Lending Club Case Study

**Group Members:**

Somrik Banerjee

Dhirendra Kumar Suman

| | |
|---|---|
| **Assignment Module** | Exploratory Data Analysis(EDA) |
| **Assignment Title** | Lending Club Case Study |
| **Assignment Objective** | Right lending decisions based on the likelihood of :<br>1. Applicants likely to default (Reduce credit loss)<br>2. Creditworthy applicants (Increased business opportunities) |

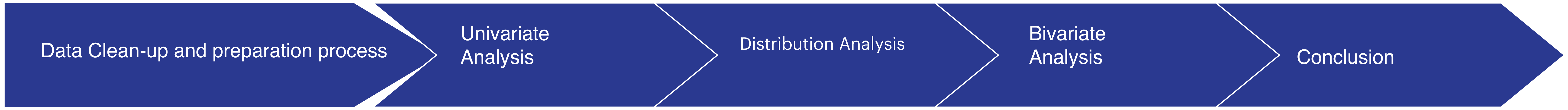
## Introduction

Lending Club Inc., as the largest online loan marketplace, offers a variety of loans including personal, business, and medical procedure financing through an efficient online platform. A major challenge they face, common in the lending industry, is credit loss, which occurs when borrowers default on their loans. These defaulters, referred to as 'charged-off' customers, represent the largest source of financial loss. To mitigate this, the company aims to identify these high-risk applicants. The objective is to understand the key factors that indicate the likelihood of a loan default, allowing the company to refine its loan portfolio and enhance its risk assessment strategies

## Problem Statement

The aim of this analysis is to use Exploratory Data Analysis to understand the key driving factors which indicate whether a customer is likely to default on a loan. To achieve this, we will make use of various data cleaning, exploration as well as visualisation techniques using Python.

# Analysis Approach



Data Clean-up and preparation process

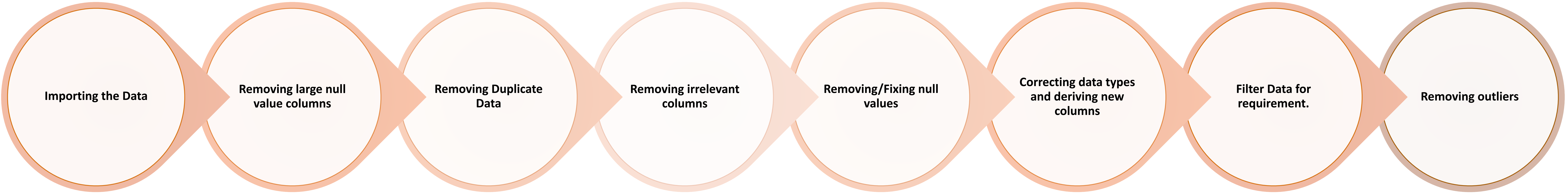
Univariate Analysis

Distribution Analysis

Bivariate Analysis

Conclusion

# Data Clean-up and preparation process:



Importing the Data

Removing large null value columns

Removing Duplicate Data

Removing irrelevant columns

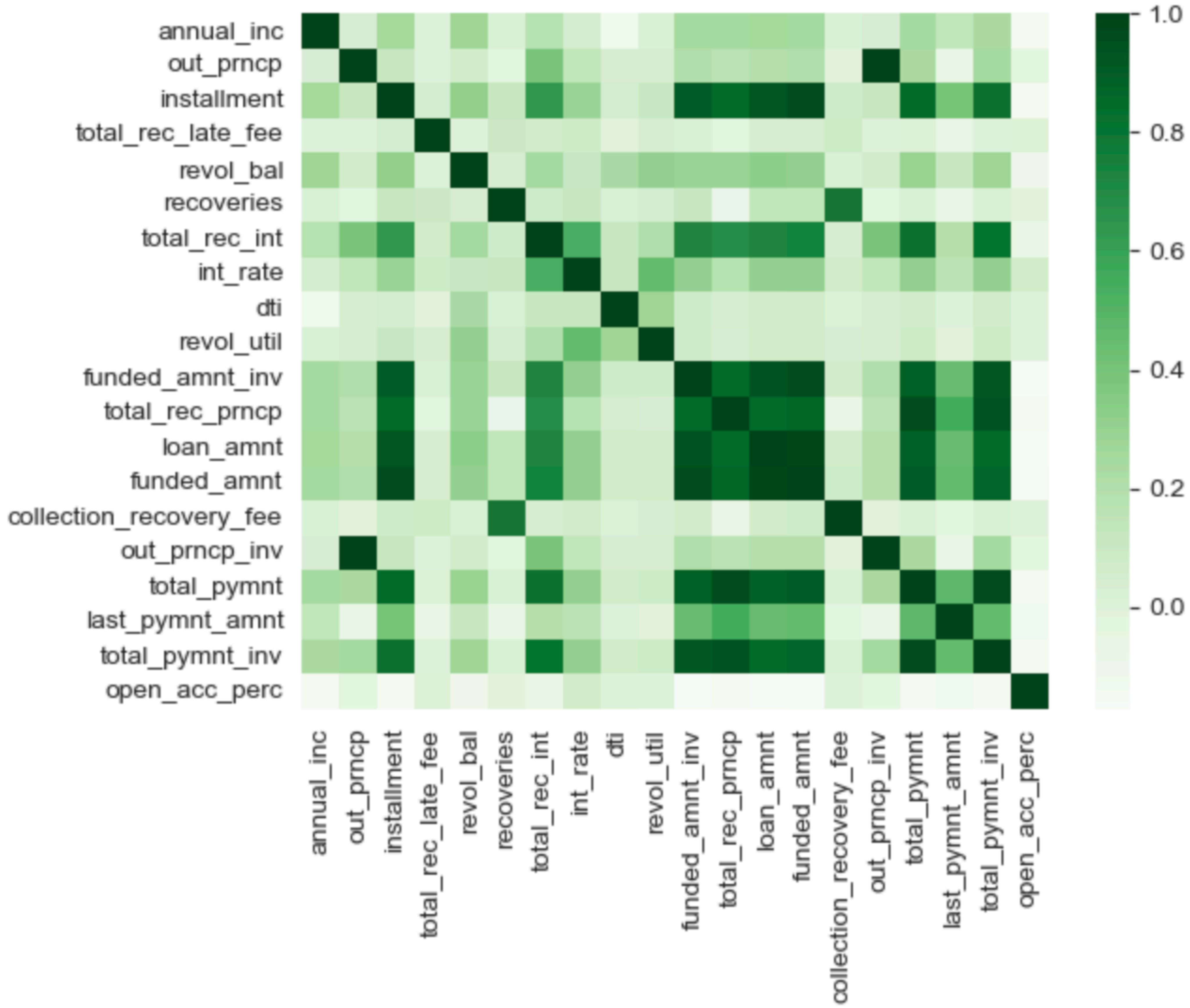
Removing/Fixing null values

Correcting data types and deriving new columns

Filter Data for requirement.

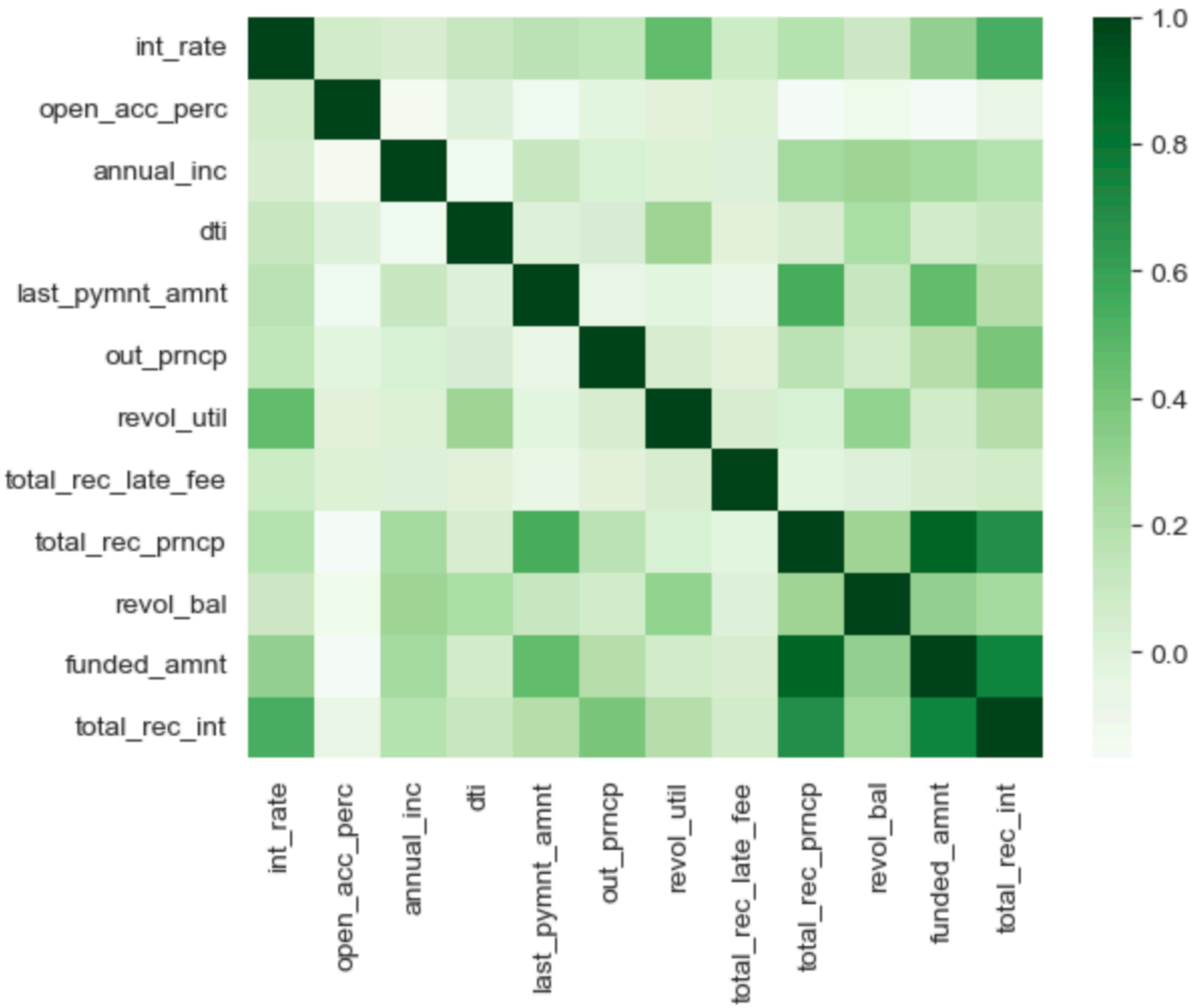
Removing outliers

# Correlated features within continuous variables to identify possible proxies



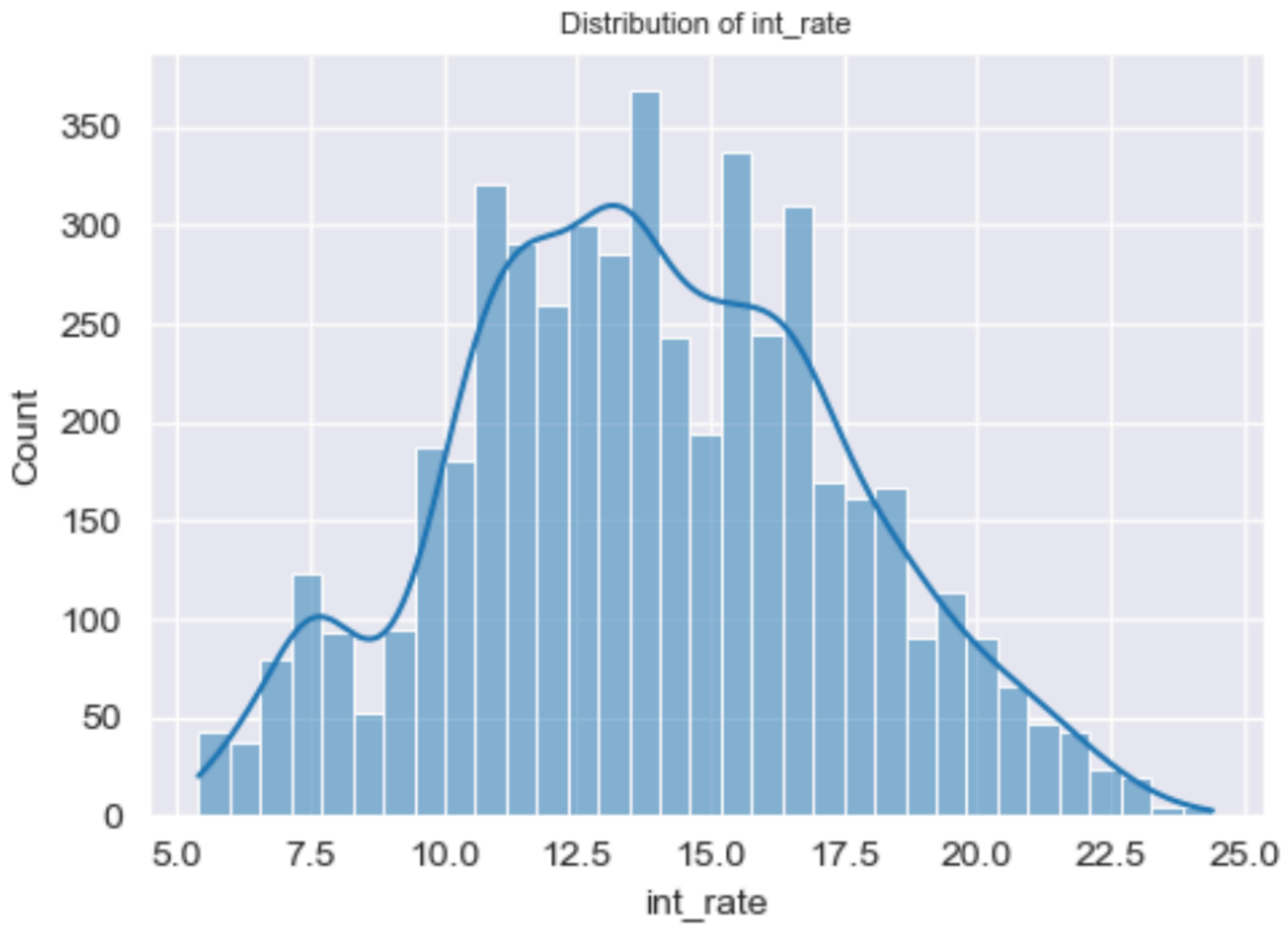
We immediately see some very obvious patterns. `loan_amnt`, `funded_amnt`, `funded_amnt_inv` and `installment` are a very tightly correlated group that all quantify one thing i.e. the amount of money loaned to the borrower. We can keep `funded_amnt` and drop the rest.

# Correlated features

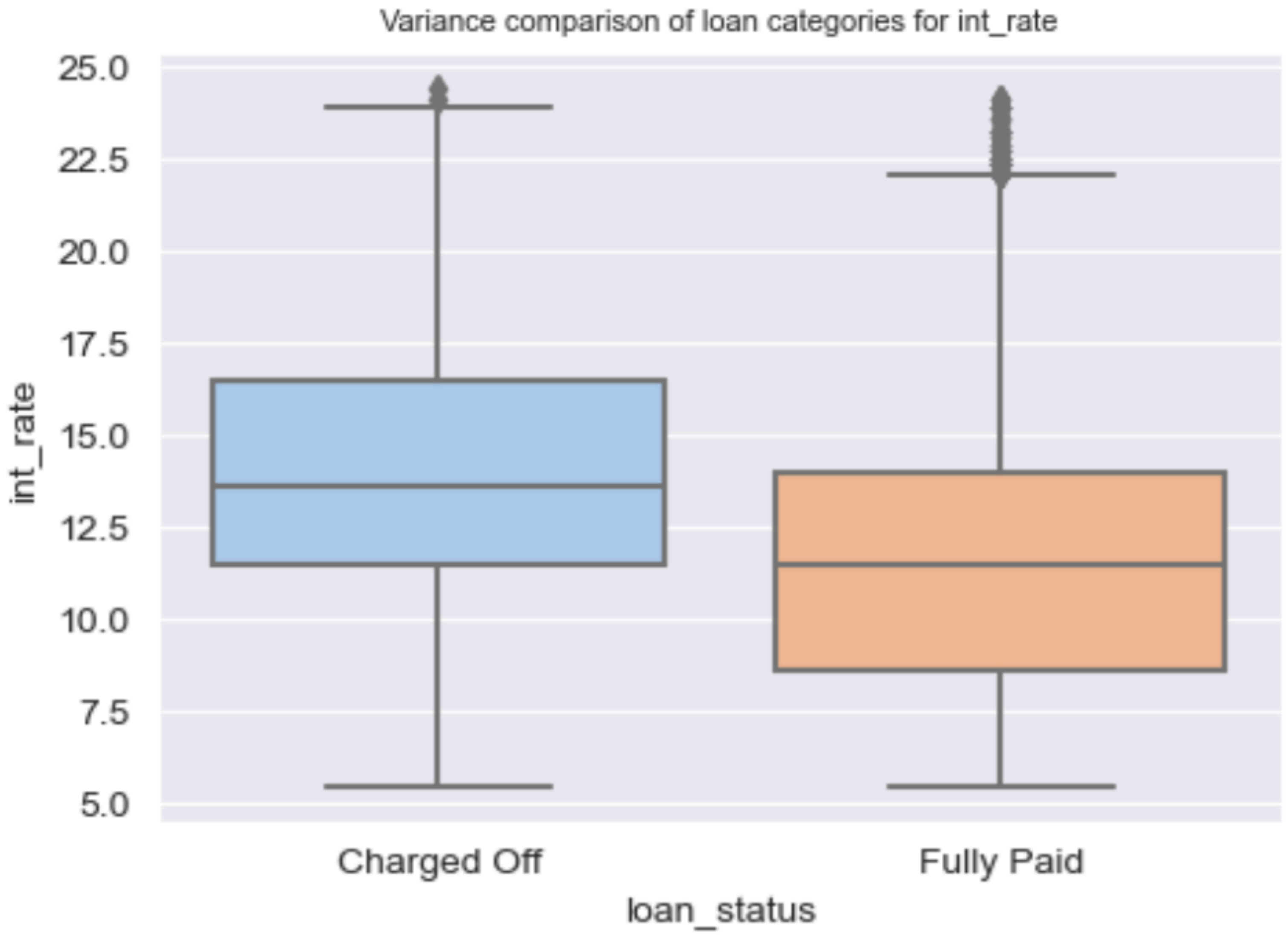


# Distribution Analysis - Continuous Variables

## Intrest Rate

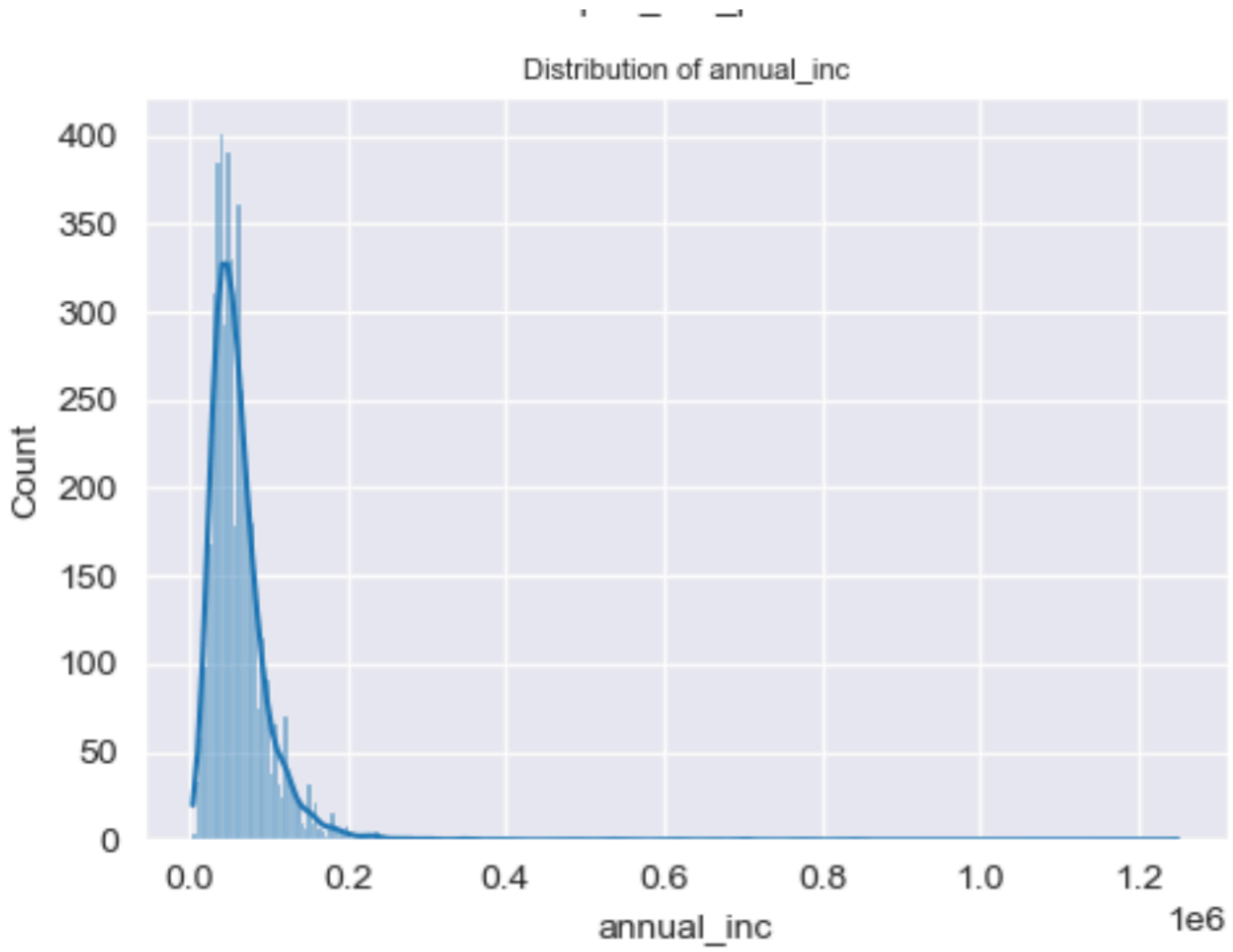


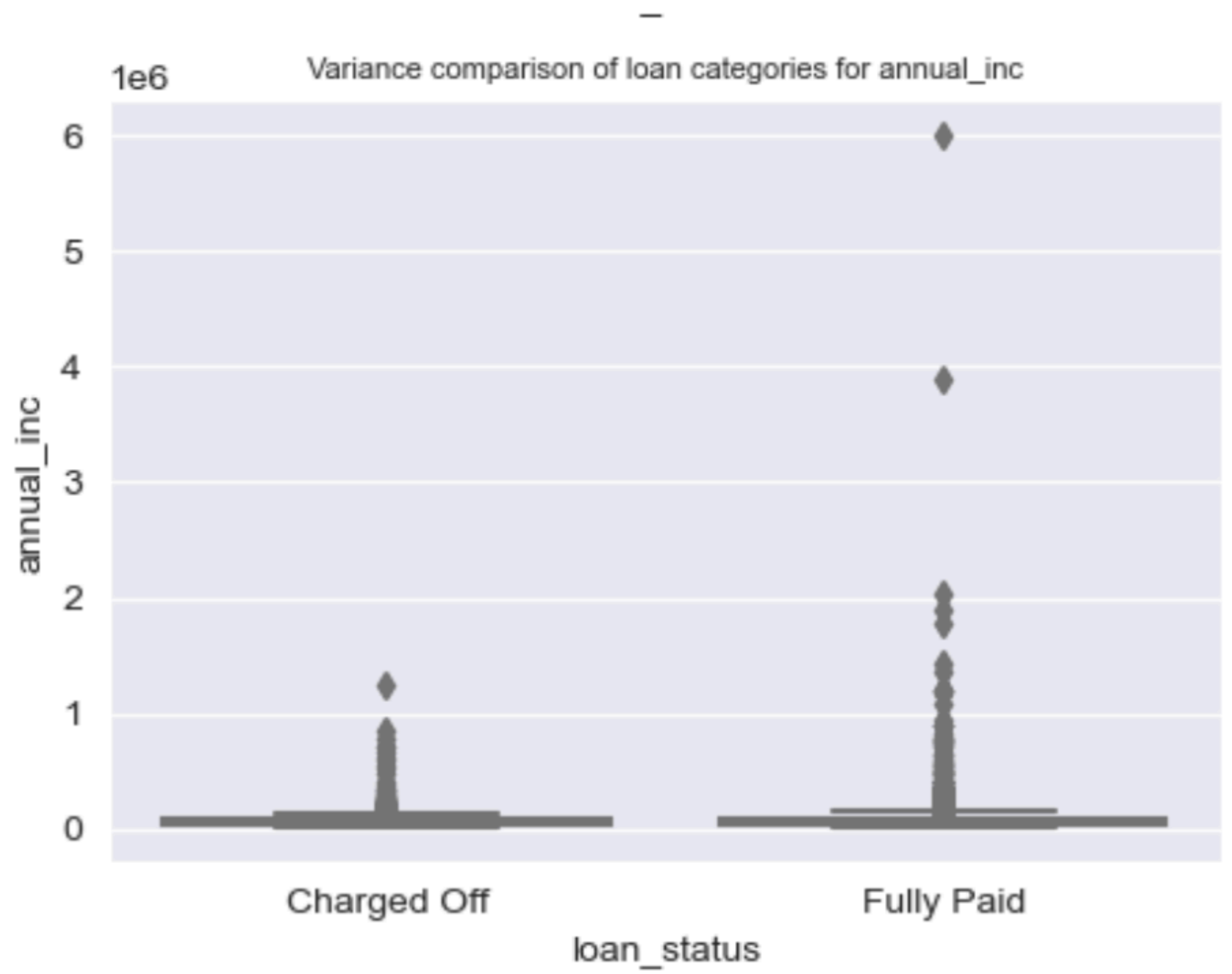
Distribution of int_rate



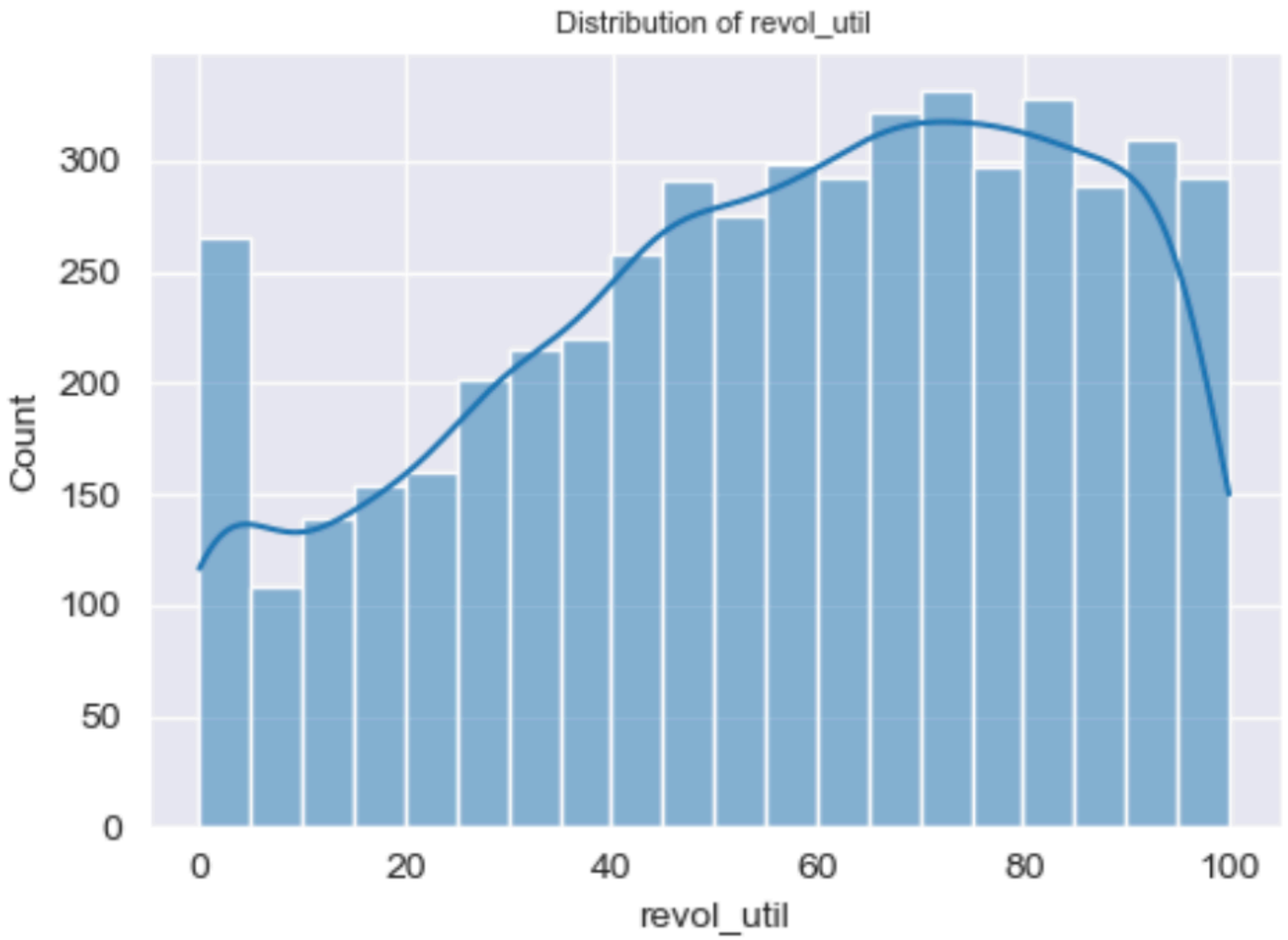
Variance comparison of loan categories for int_rate

# Annual Income

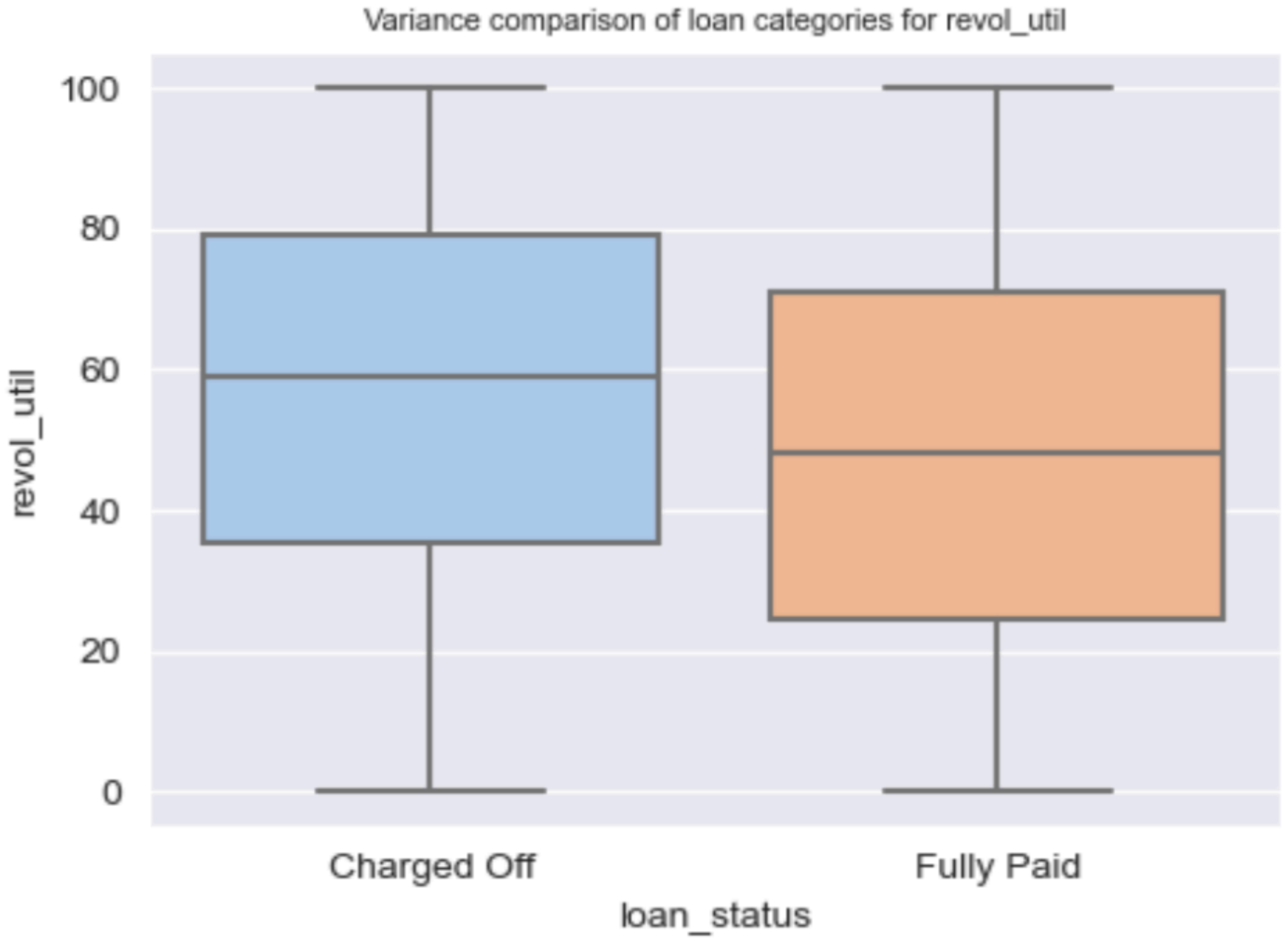

Distribution of annual_inc


Variance comparison of loan categories for annual_inc

# Amount of Revolving Credit



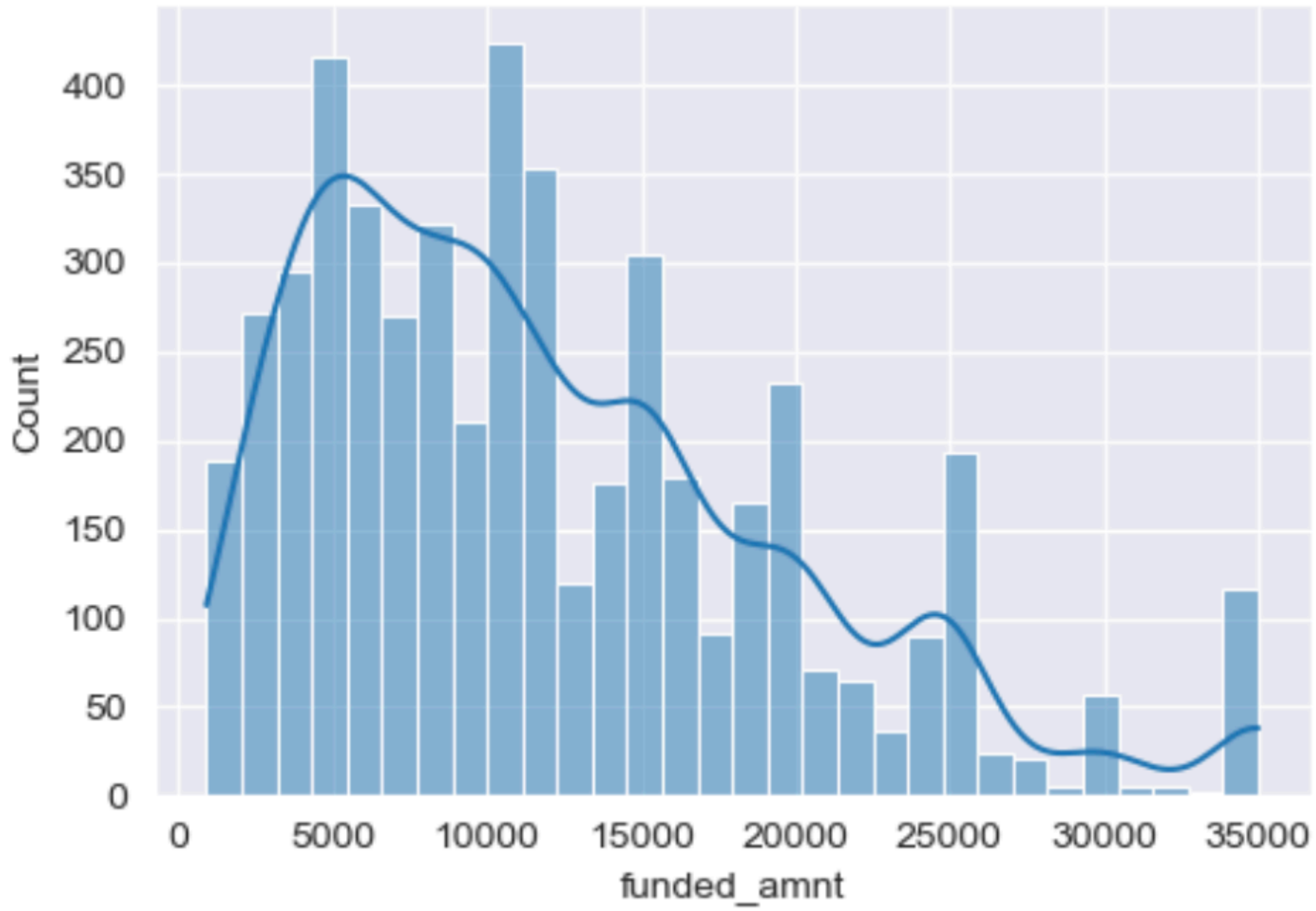
Distribution of revol_util



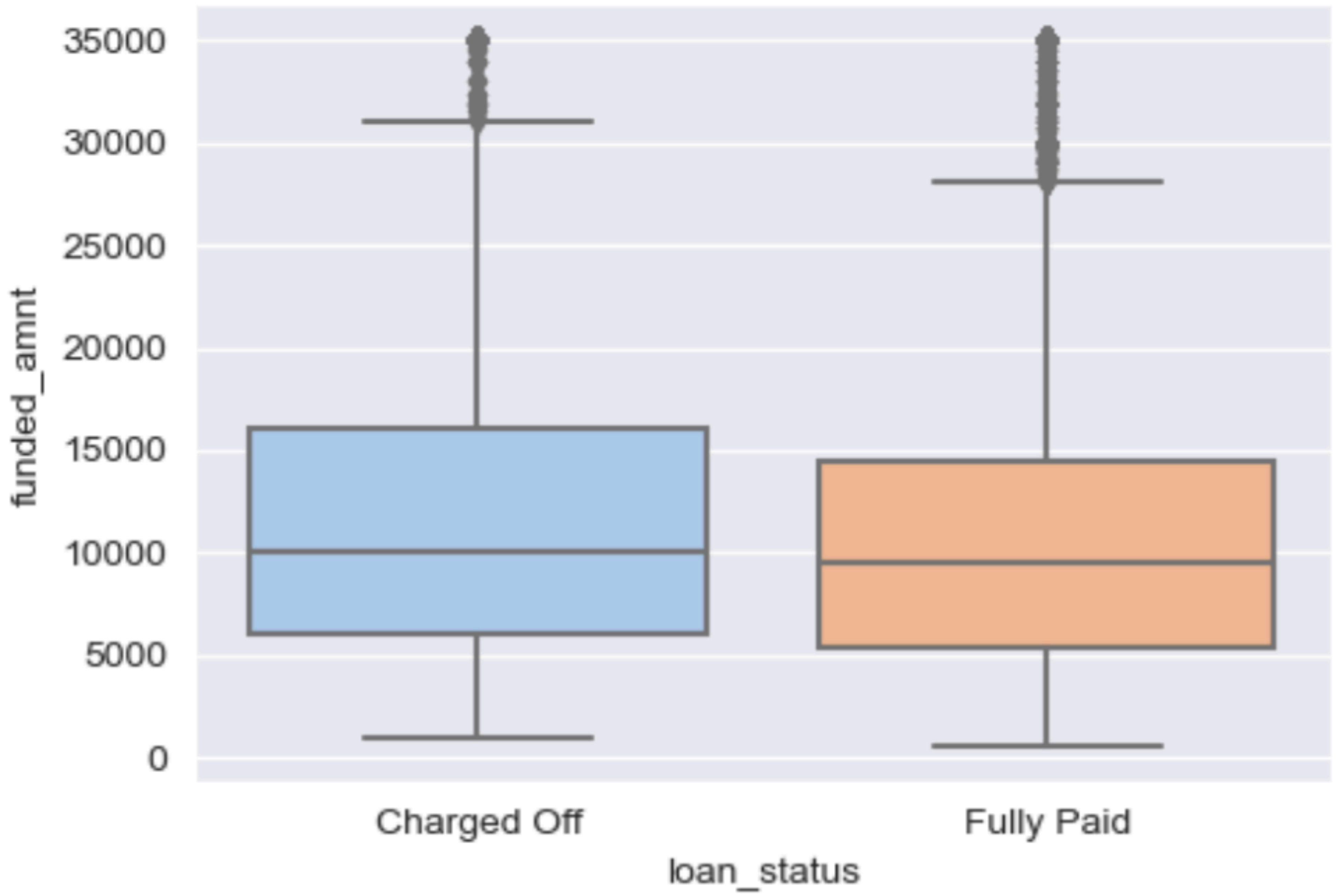
Variance comparison of loan categories for revol_util

## Loan Amount Funded

Distribution of funded_amnt



Variance comparison of loan categories for funded_amnt

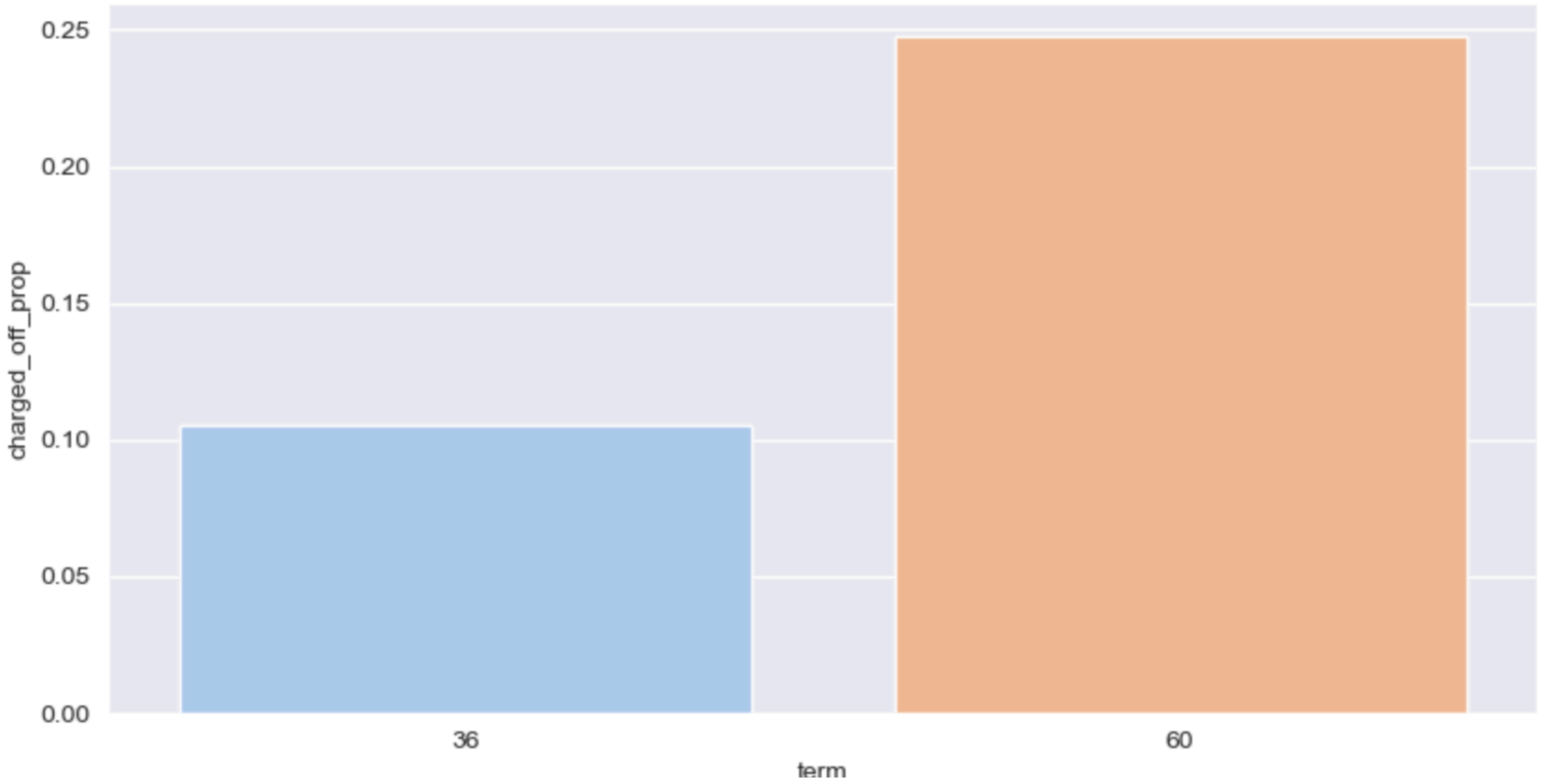


## Observation

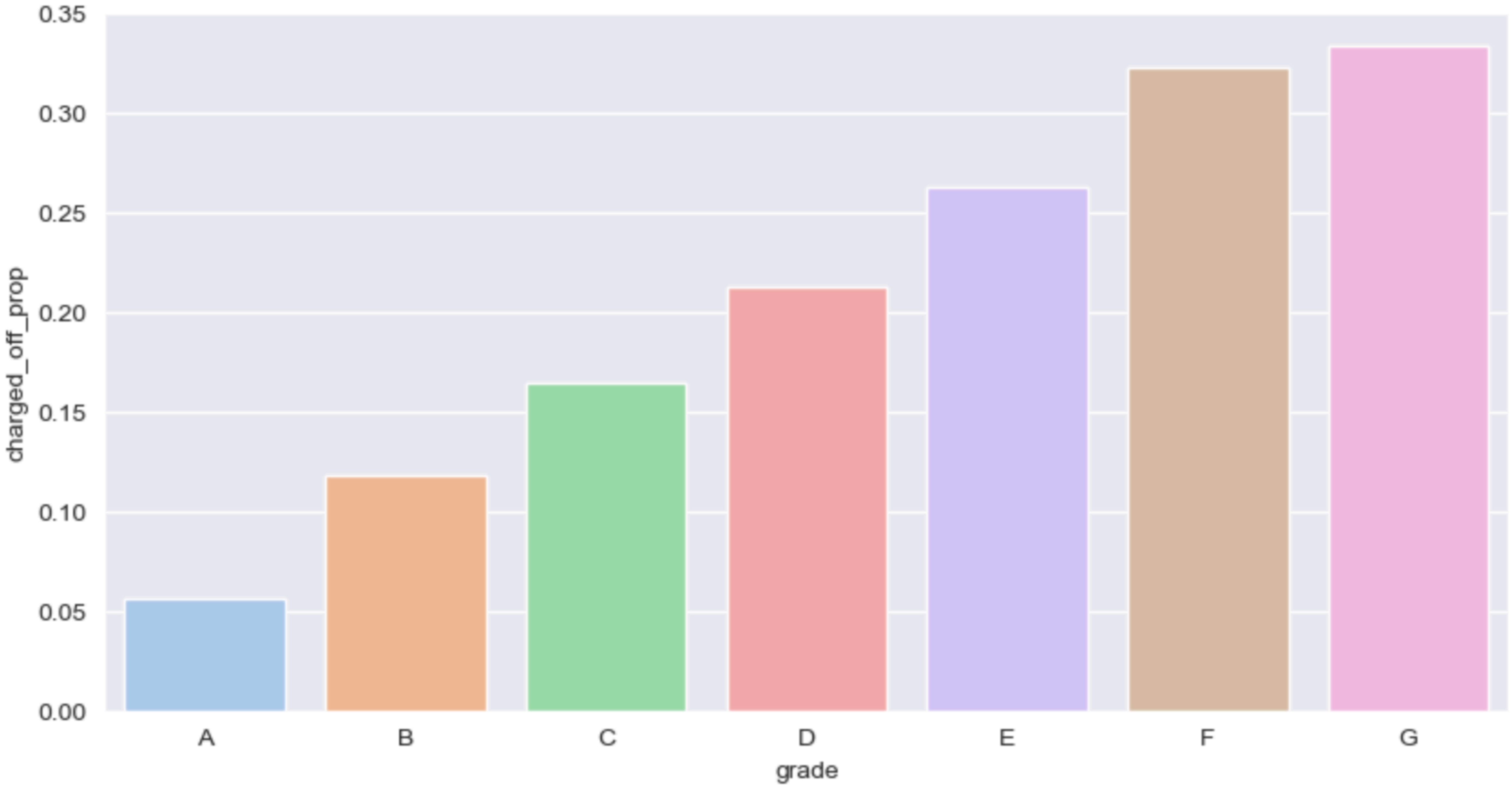
We can see from this analysis that `loan_status` is significantly dependent on variation in `int_rate`, `funded_amnt` and `revol_util`.

# Distribution Analysis - Categorical Variables

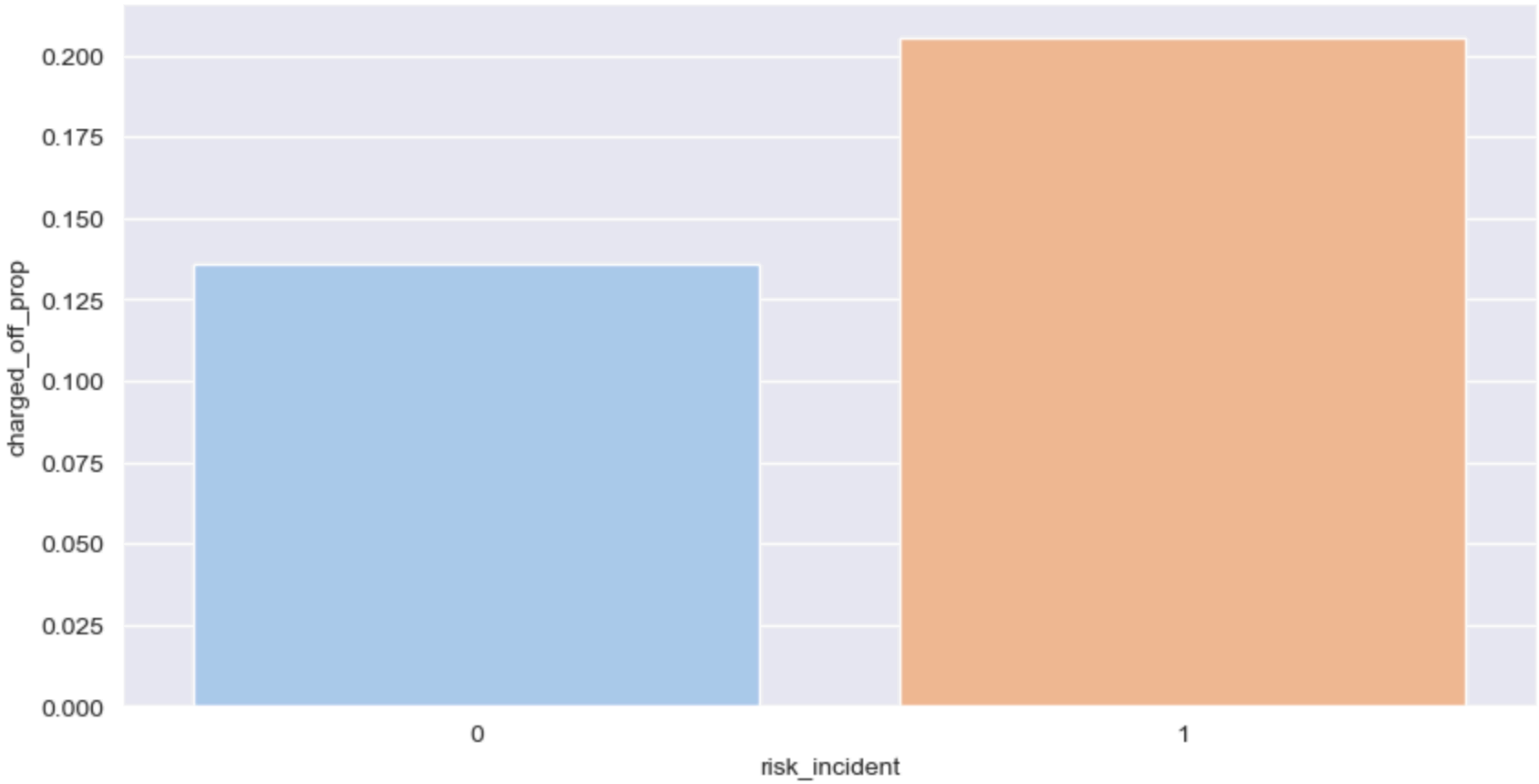
## Term of Loan



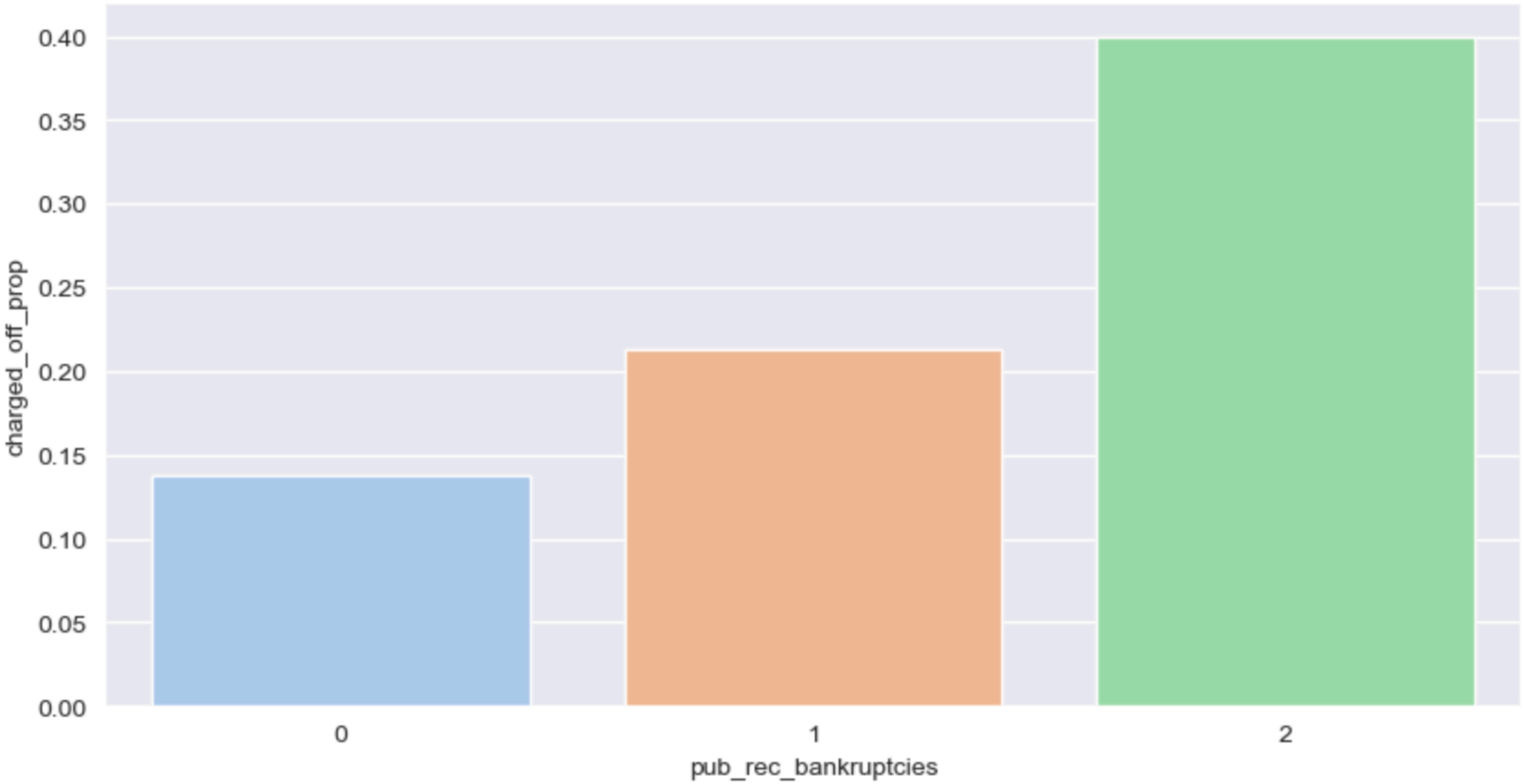
# Loan Grade


0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
charged_off_prop
A
B
C
D
E
F
G
grade

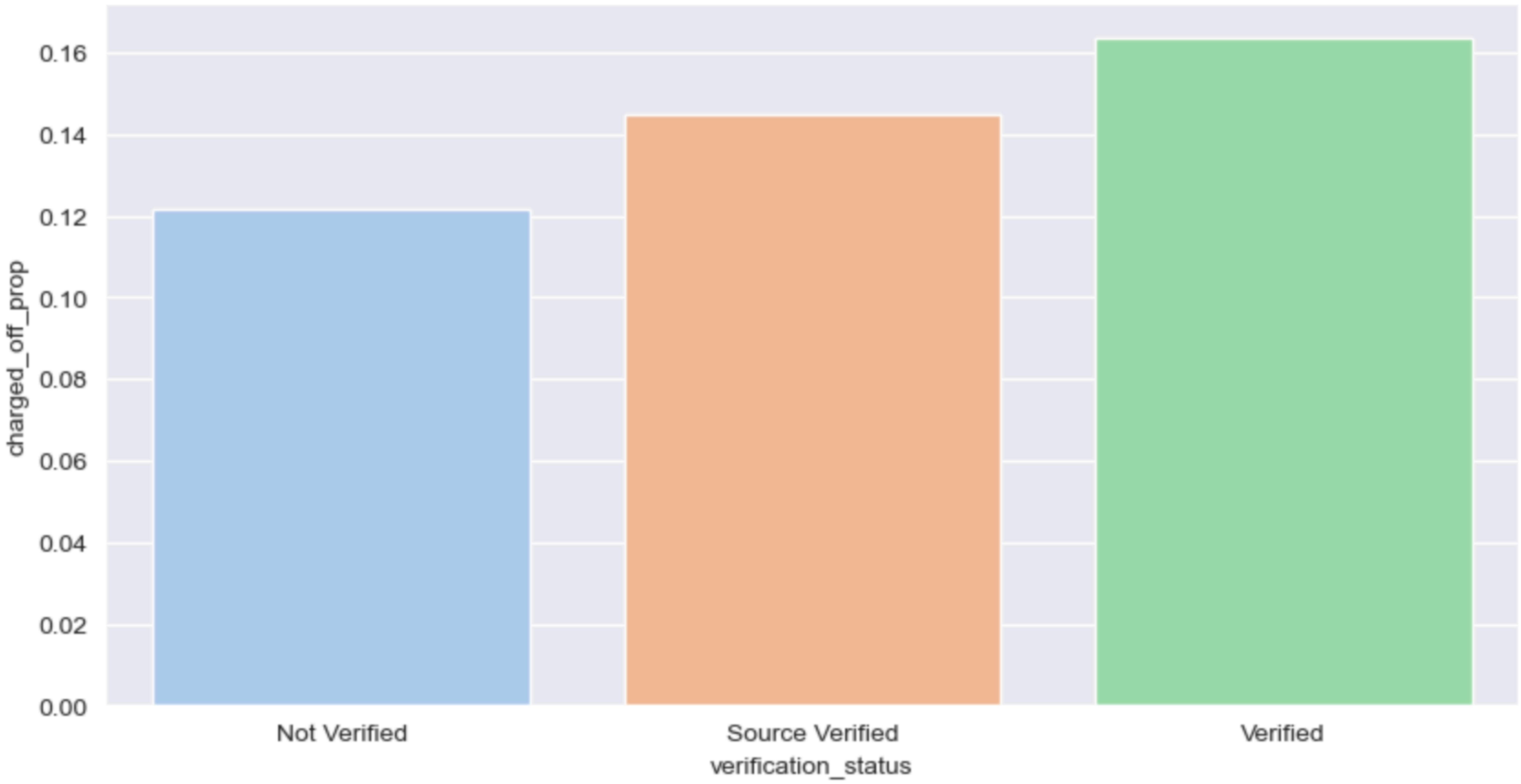
# Risk_incident(Derived Metric) from mths_since_last_record



# Number of publicly recorded bankruptcies



# verification_status



## Observation

From the above charts, it is obvious that among the categorical features, `term` is the most prominent indicator of a potential defaulter. Other prominent indicators are `grade`, `risk_incident` and `pub_rec_bankruptcies`.

Although `verification_status` shows a very prominent correlation with loan status, we cannot use it as an indicator because loans are more likely to be given out to verified sources so we will obviously have more charge-offs for higher verification levels.

## Conclusion:

The following are the list of prominent indicators of a loan defaulter as obtained from the above analysis.

- `int_rate` -> interest rate
- `annnual_inc` -> annual income
- `revol_util` -> amount of revolving credit
- `funded_amnt` -> loan amount funded
- `term` -> term of loan
- `grade` -> loan grade (higher grade points to higher risk of default)
- `risk_incident` -> derived metric indicating whether the borrower has had any publicly recorded credit risk incident in the past, based on `mths_since_last_record`
- `pub_rec_bankruptcies` -> number of publicly recorded bankruptcies

Thank you