Lending Club Case Study

Group Members:
Somrik Banerjee
Dhirendra Kumar Suman

Assignment Details

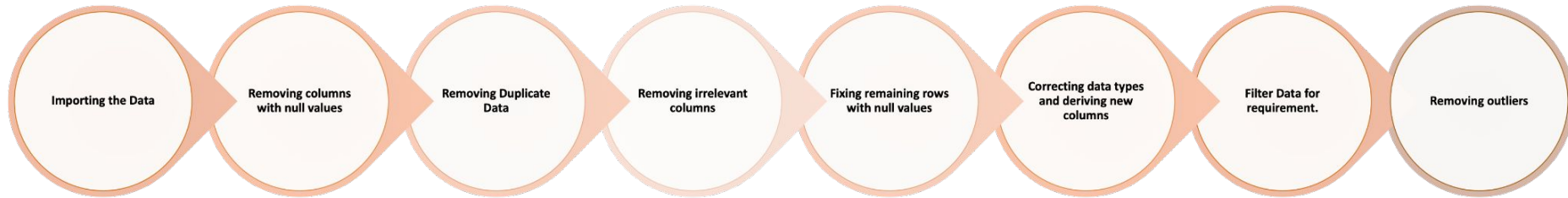| Module	Title	Objective
Exploratory Data Analysis (EDA)	Lending Club Case Study	The aim of this case study is to use Exploratory Data Analysis to understand the key driving factors which indicate whether a customer is likely to default on a loan. To achieve this, we will make use of various data cleaning, exploration as well as visualisation techniques using Python.

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

Analysis Approach

| Data Cleaning | Univariate Analysis | Bivariate Analysis | Conclusion |

Data Cleaning

| Importing the Data | Removing columns with null values | Removing Duplicate Data | Removing irrelevant columns | Fixing remaining rows with 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.
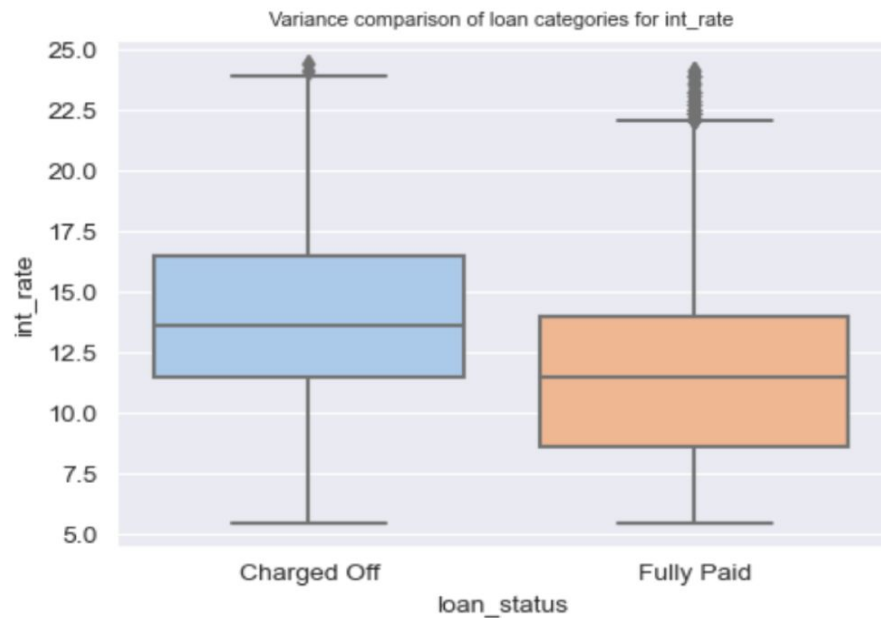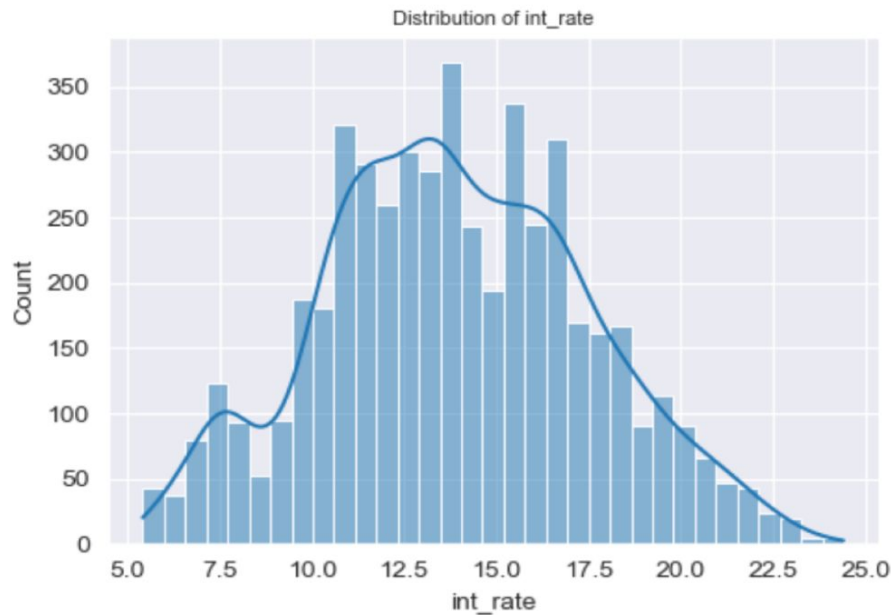
5

Correlated features
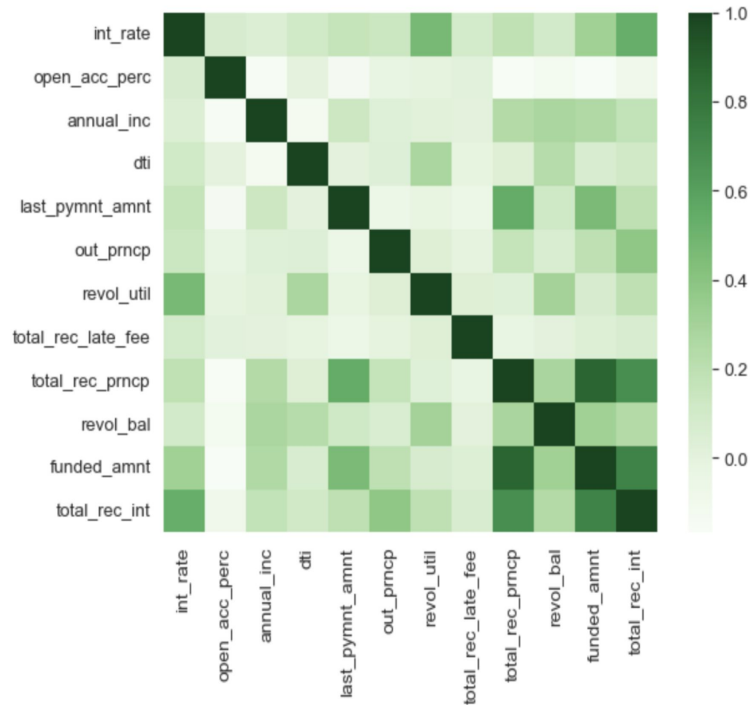
`out_prncp` and `out_prncp_inv` are two other variables of which only one is useful.

Here, we have dropped 'out_prncp_inv' .

# Distribution Analysis -Continuous Variables **(Interest Rate)**

Interest Rates: Interest rates in the 13%-17% range are associated with defaults. The company should review its interest rate determination process and consider adjusting rates based on DTI ratios to better align with the borrower's ability to repay.

# Distribution Analysis -Continuous Variables**(Annual Income)**
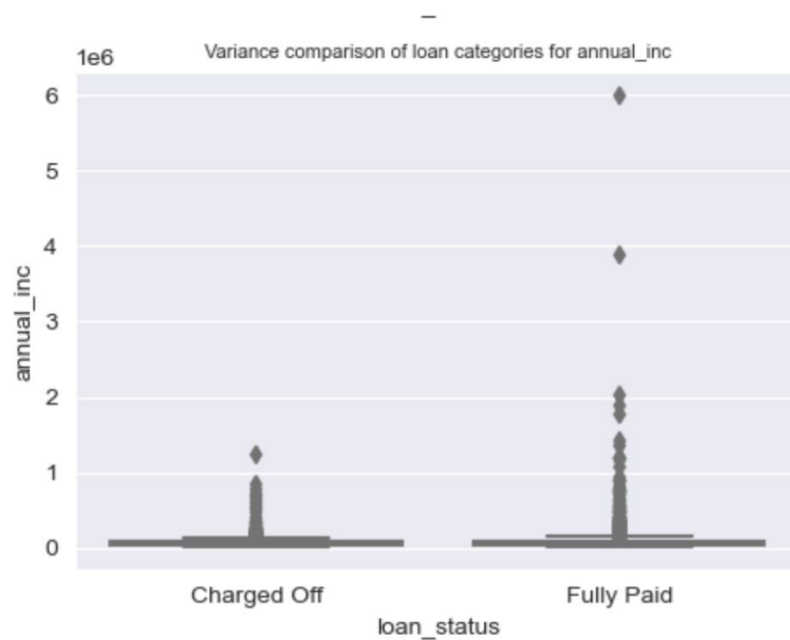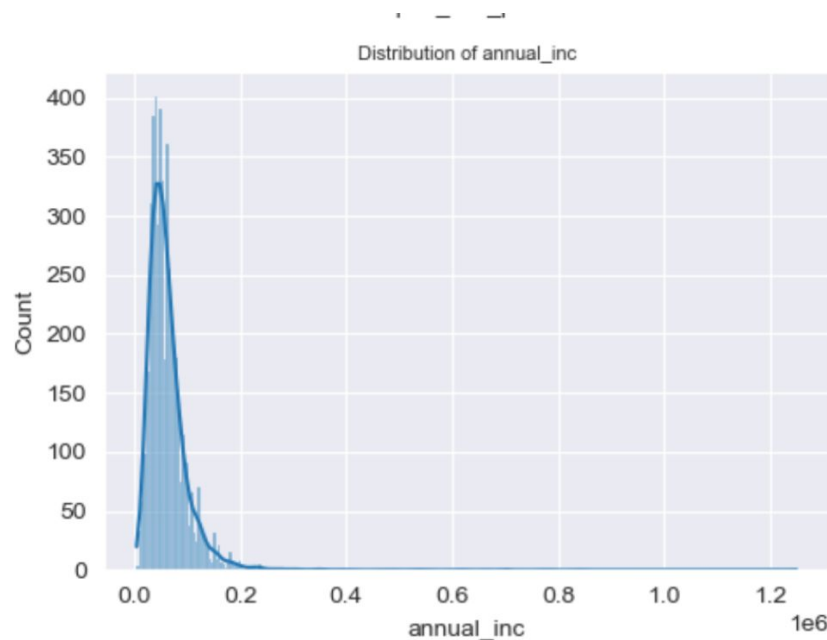
Low Annual Income: Applicants with annual incomes less than $40,000 have a higher likelihood of defaulting. The company should consider offering financial education resources or setting maximum loan amounts based on income levels to ensure affordability for borrowers.

# Distribution Analysis - **Amount of Revolving Credit**

Distribution of revol_util

Variance comparison of loan categories for revol_util
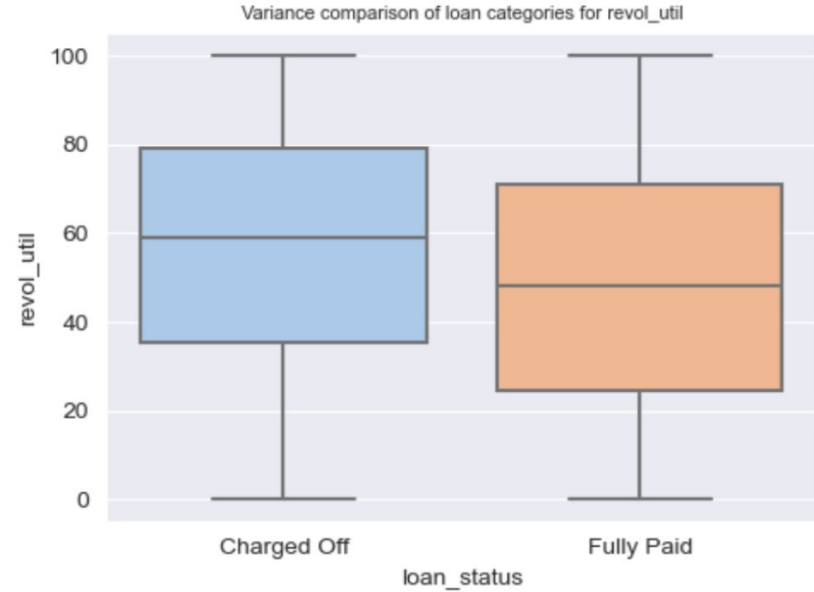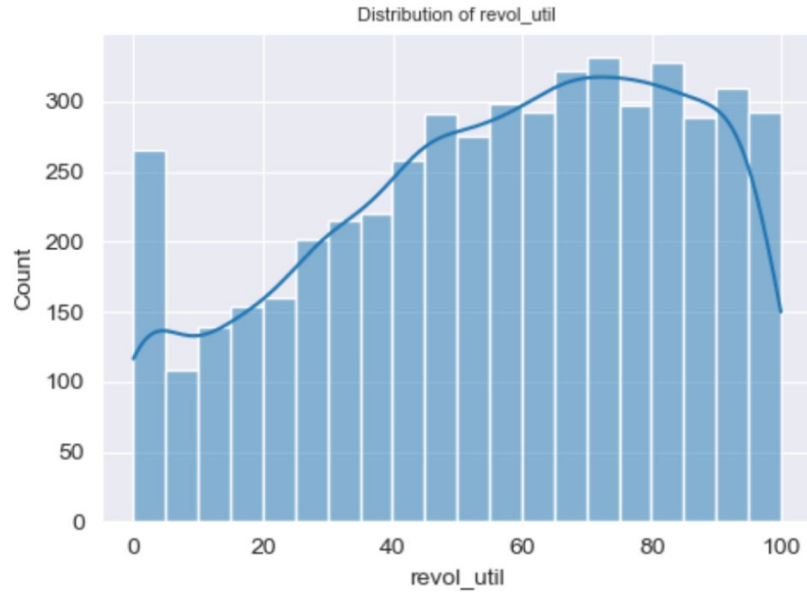
Charged off loans are directly proportional to High amount of Revolving credit.

Distribution Analysis - **Loan Amount Funded**

Distribution of funded_amnt


Variance comparison of loan categories for funded_amnt
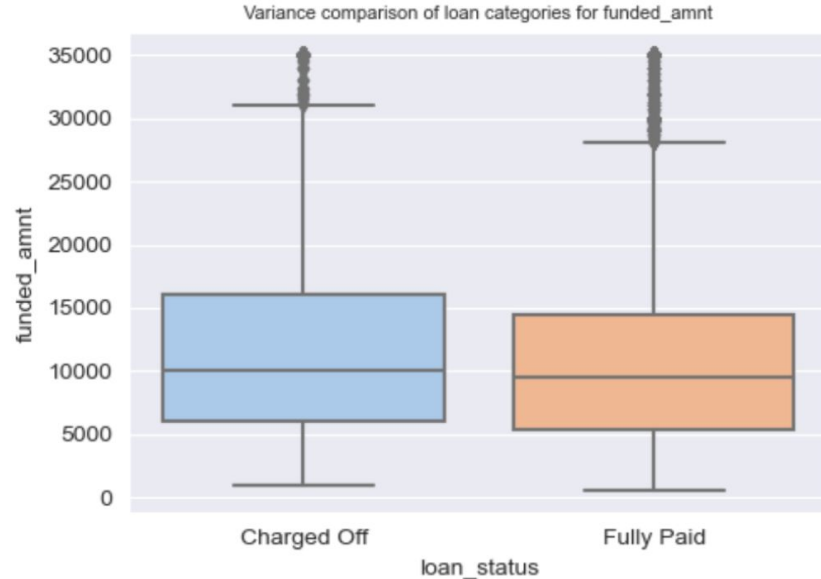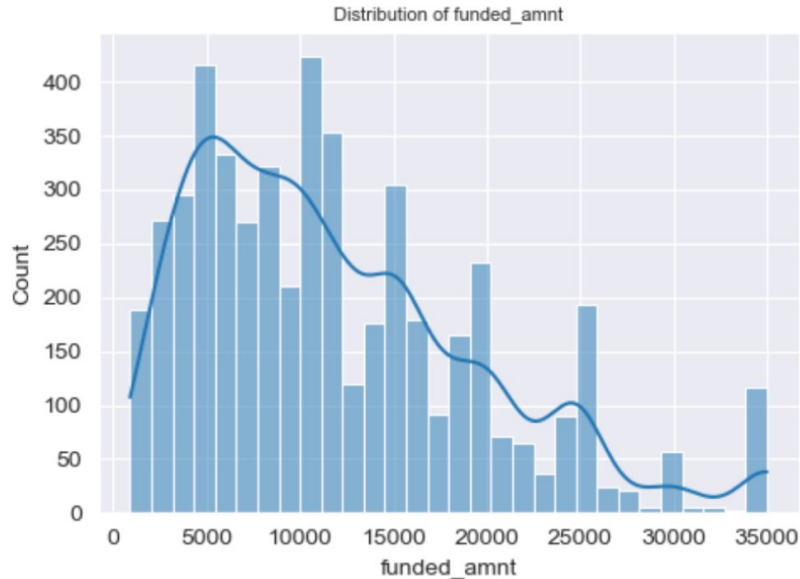
High Loan Amounts : Applicants receiving loan amounts of $15,000 or higher are more likely to default. The company can mitigate this risk by conducting more thorough assessments for larger loan requests and potentially capping loan amounts for higher-risk applicants.
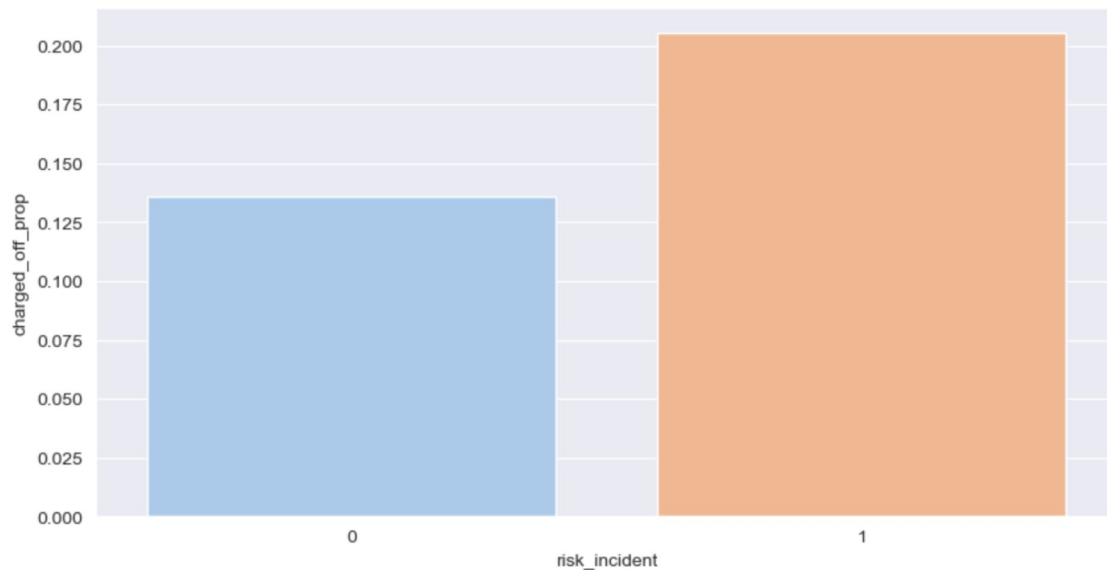
Distribution Analysis - **Term of Loan**
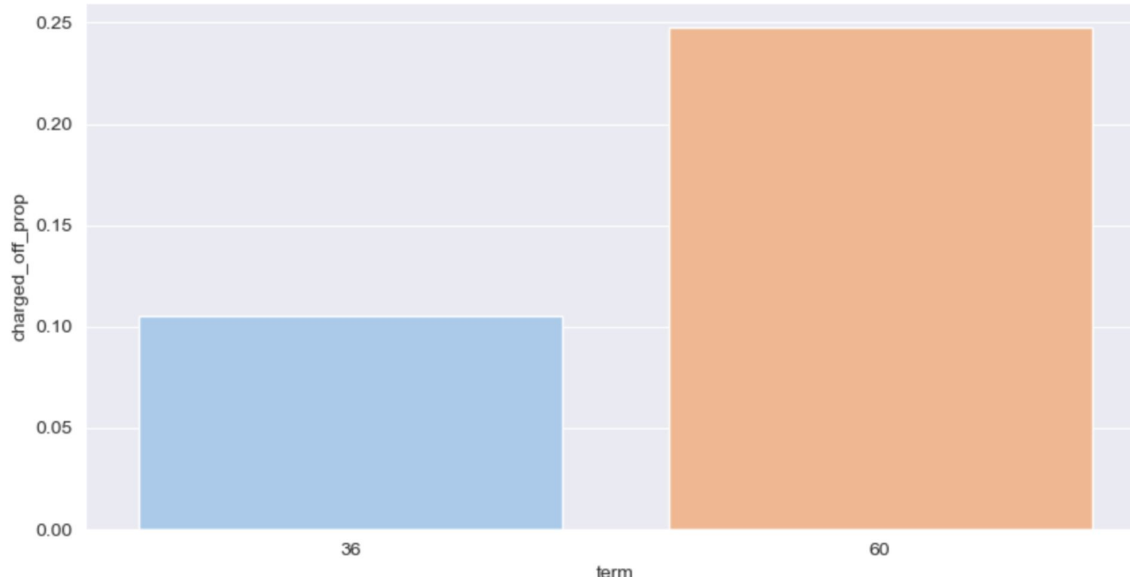
Term Length: Given that applicants opting for 60-month loans are more likely to default, the company should consider evaluating the risk associated with longer-term loans and potentially limiting the maximum term or adjusting interest rates accordingly
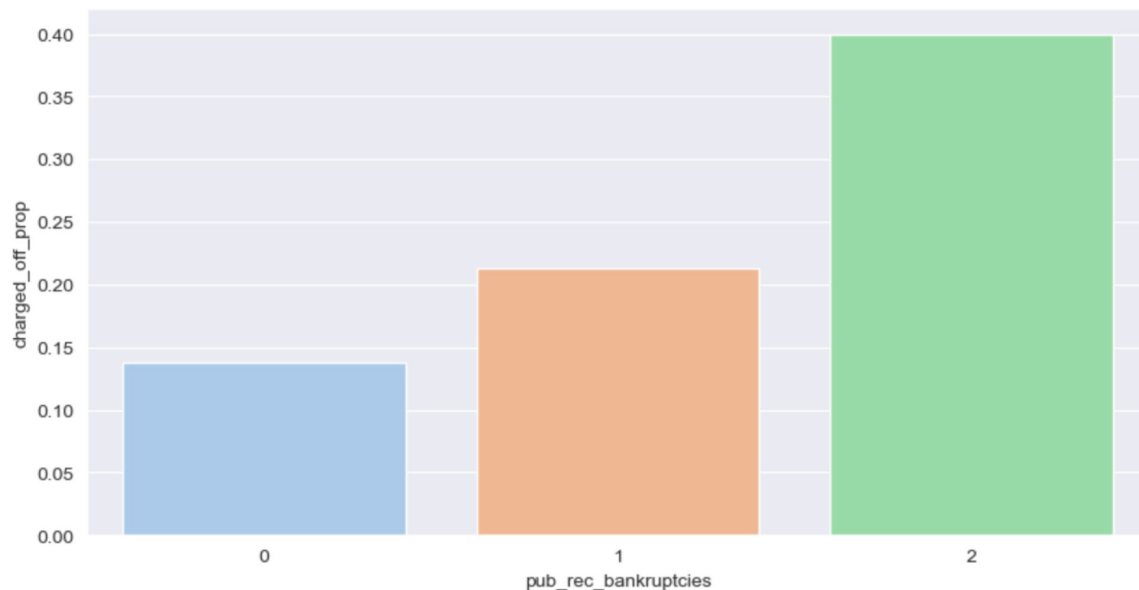
Distribution Analysis - **Loan Grade**
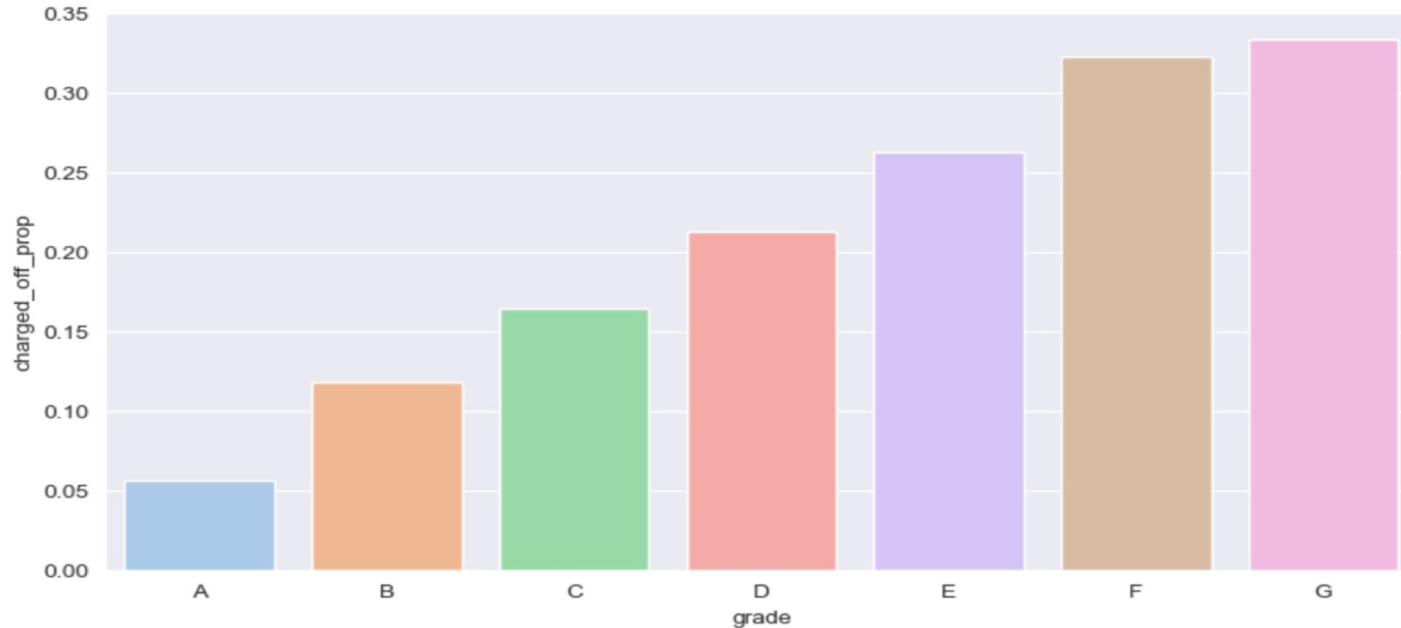
Risk Assessment for Grades B, C, and D: Since loan applicants from Grades B, C, and D contribute to most of the "Charged Off" loans, the company should consider implementing stricter risk assessment and underwriting criteria for applicants falling into these grades.

Distribution Analysis- **Risk_incident**(Derived Metric from mths_since_last_record)
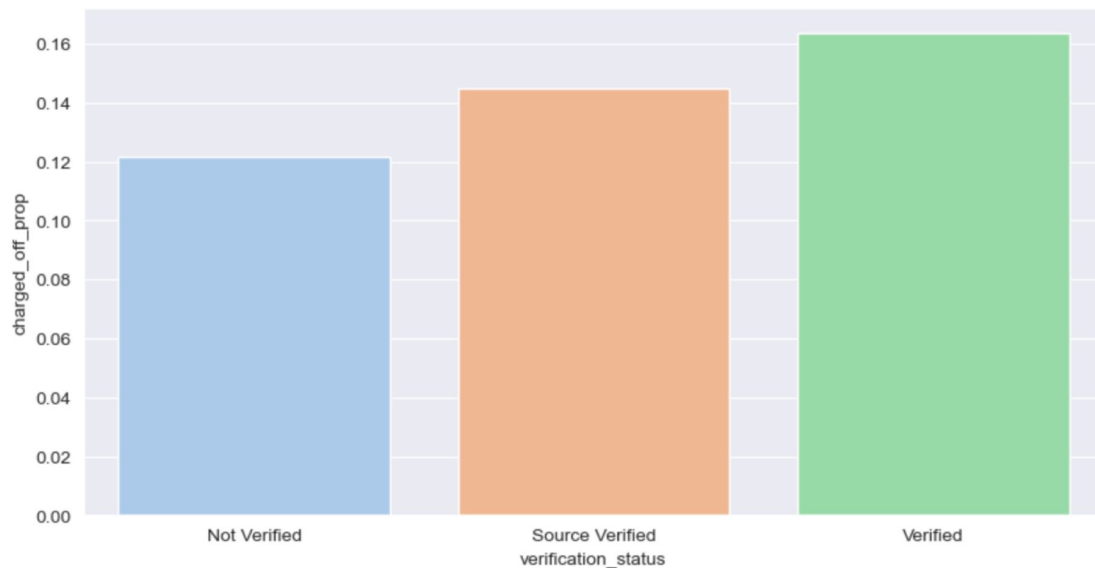

Charged off loans are high for positive risk incident.

Distribution Analysis - **Number of publicly recorded bankruptcies**

Charged off loans increases for increase in publicly recorded bankruptcies, highest for 2 publicly recorded bankruptcies.

Distribution Analysis - **Verification Status**

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