

Lending Club Case Study

Shweta Kumari

Sri Sakthi

Summary

- Problem Statement
- Input Overview
- Data Cleaning
- Exploratory Data Analysis
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Correlation Analysis
- Conclusions
- Recommendations

Problem Statement

This case study aims to analyse loan data from Lending Club to identify patterns that indicate the likelihood of default by borrowers.

The analysis focuses on understanding how various consumer and loan attributes influence loan default. The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

The company can utilise this knowledge for its portfolio and risk assessment.

Input Overview

```
In [3]: #Finding the rows and columns of the dataframe
dataframe.shape
```

```
Out[3]: (39717, 111)
```

```
In [4]: #To check the summary of the data
dataframe.describe() #87 columns are of numeric-type
```

```
Out[4]:
```

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	installment	annual_inc	dti	delinq_2yrs	inq_last_6mths	...
count	3.971700e+04	3.971700e+04	39717.000000	39717.000000	39717.000000	39717.000000	3.971700e+04	39717.000000	39717.000000	39717.000000	...
mean	6.831319e+05	8.504636e+05	11219.443815	10947.713196	10397.448868	324.561922	6.896893e+04	13.315130	0.146512	0.869200	...
std	2.106941e+05	2.656783e+05	7456.670694	7187.238670	7128.450439	208.874874	6.379377e+04	6.678594	0.491812	1.070219	...
min	5.473400e+04	7.069900e+04	500.000000	500.000000	0.000000	15.690000	4.000000e+03	0.000000	0.000000	0.000000	...
25%	5.162210e+05	6.667800e+05	5500.000000	5400.000000	5000.000000	167.020000	4.040400e+04	8.170000	0.000000	0.000000	...
50%	6.656650e+05	8.508120e+05	10000.000000	9600.000000	8975.000000	280.220000	5.900000e+04	13.400000	0.000000	1.000000	...
75%	8.377550e+05	1.047339e+06	15000.000000	15000.000000	14400.000000	430.780000	8.230000e+04	18.600000	0.000000	1.000000	...
max	1.077501e+06	1.314167e+06	35000.000000	35000.000000	35000.000000	1305.190000	6.000000e+06	29.990000	11.000000	8.000000	...

8 rows x 87 columns

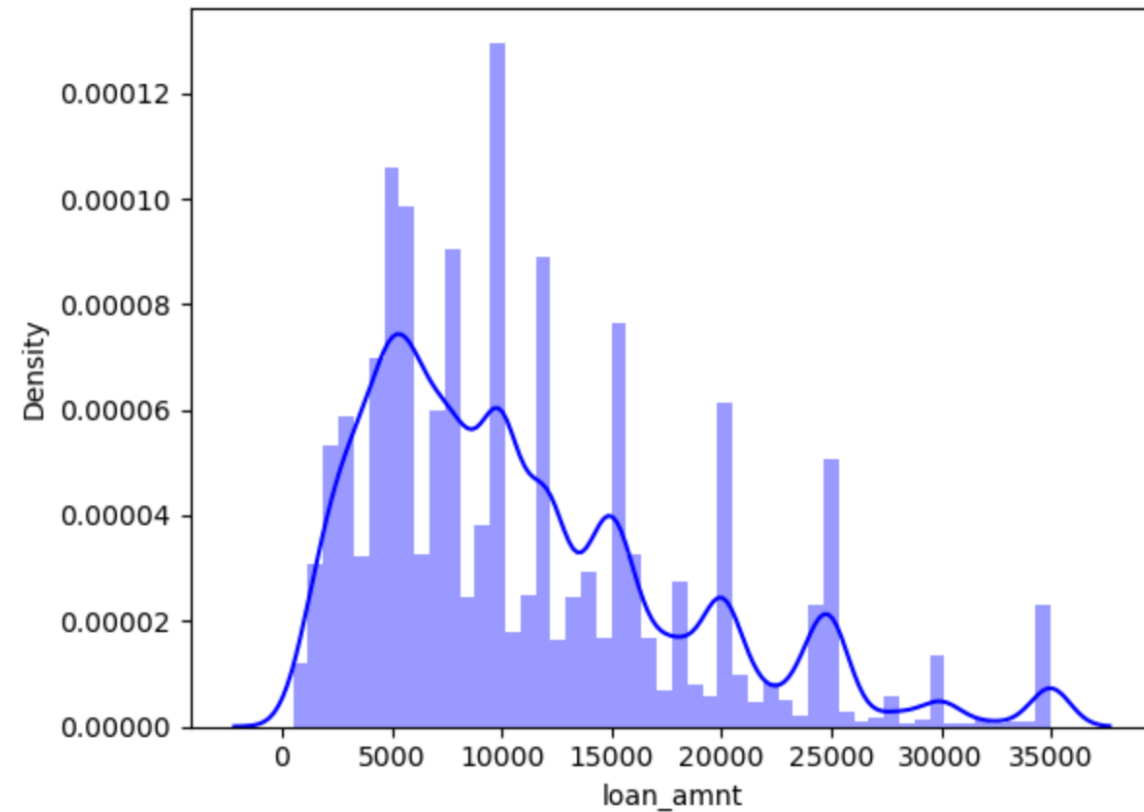
Data Cleaning Steps

1. Removed records with 'Current' loan status as the tenure is not completed.
2. Removed columns with 100% null values
3. Dropped columns with only one unique value as they don't contribute to analysis.
4. Removed columns irrelevant to loan approval process (post-approval behavioral columns).
5. Converted data types of `int_rate`, `term`, `loan_amnt`, `funded_amnt`, and `issue_d`.
6. Handled missing values in `emp_length` and `pub_rec_bankruptcies` columns by dropping rows.

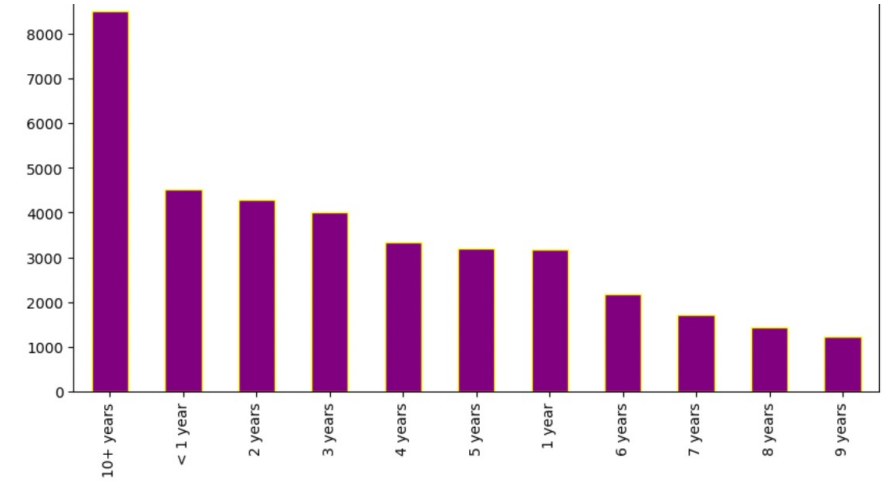
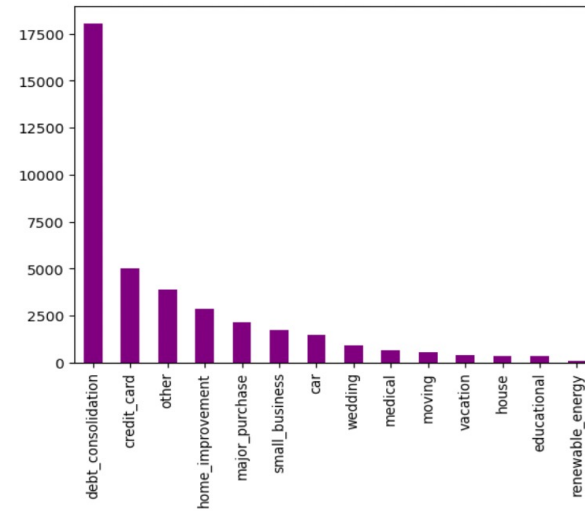
Exploratory Data Analysis

UNIVARIATE ANALYSIS

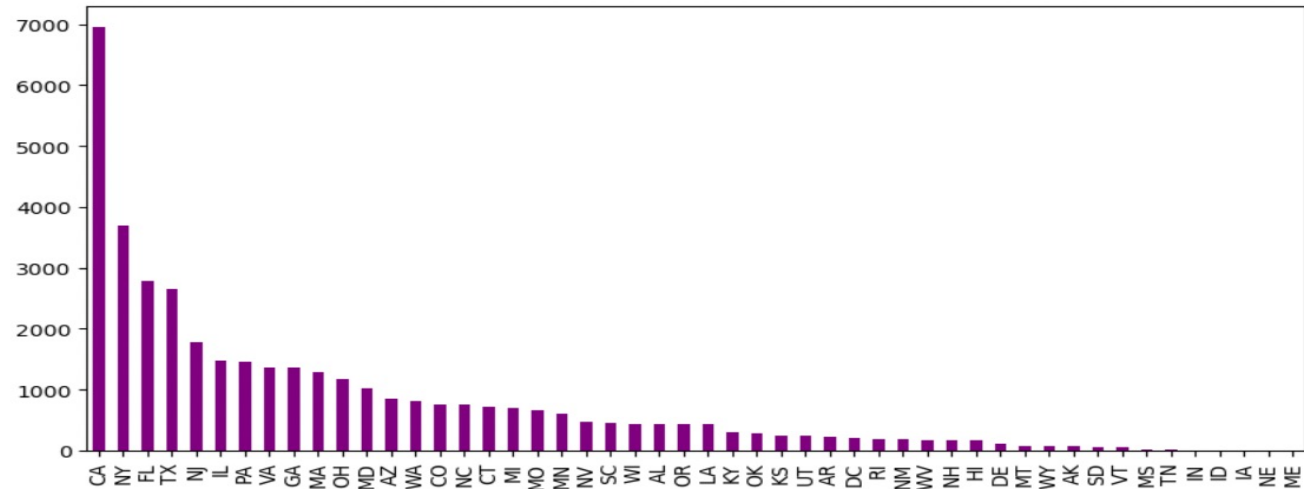
1. Distribution plot for Annual income and Loan amount



Univariate Analysis categorical variable

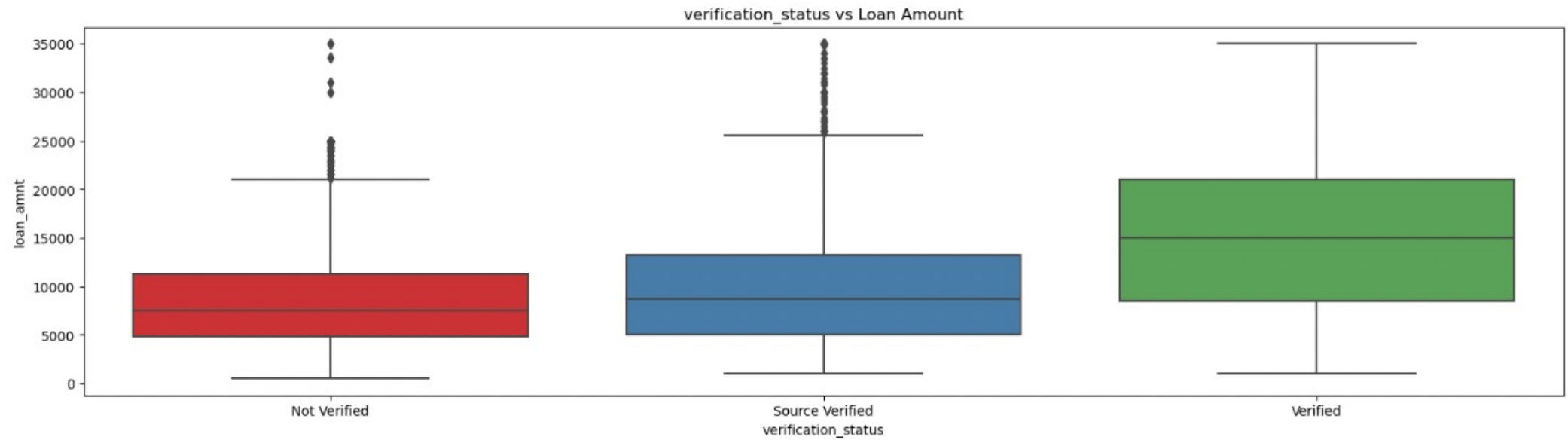


1. This analysis shows that highest number of people took loan for the purpose of debt consolidation
2. The Bar Plot shows that maximum people who takes loan is 10+ years Experienced
3. The Bar Plot shows that Maximum people who is taking Loan is from CA



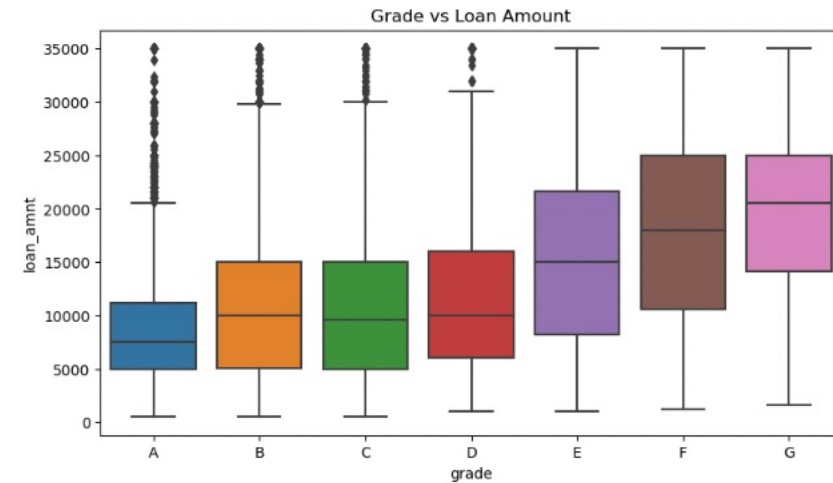
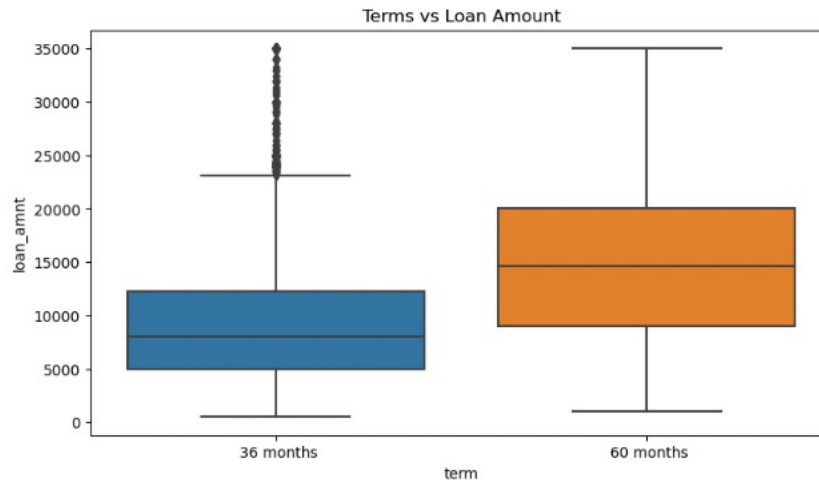
Segmented Univariate Analysis

#This plot shows that verified customers are taking more Loan amount



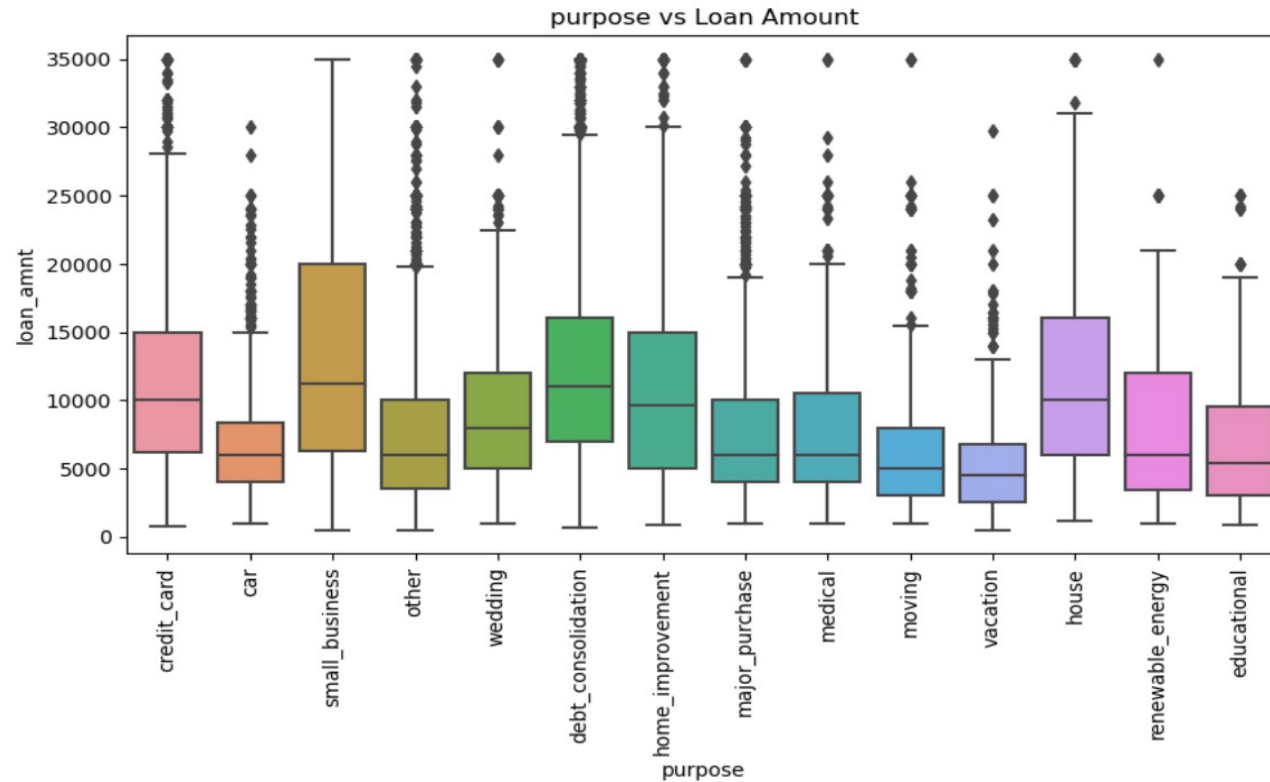
Segmented Univariate Analysis

#The first graph shows that the people with 60 months of tenure is taking more Loan
#The second graph shows the people with increasing grade has more Loan Amount



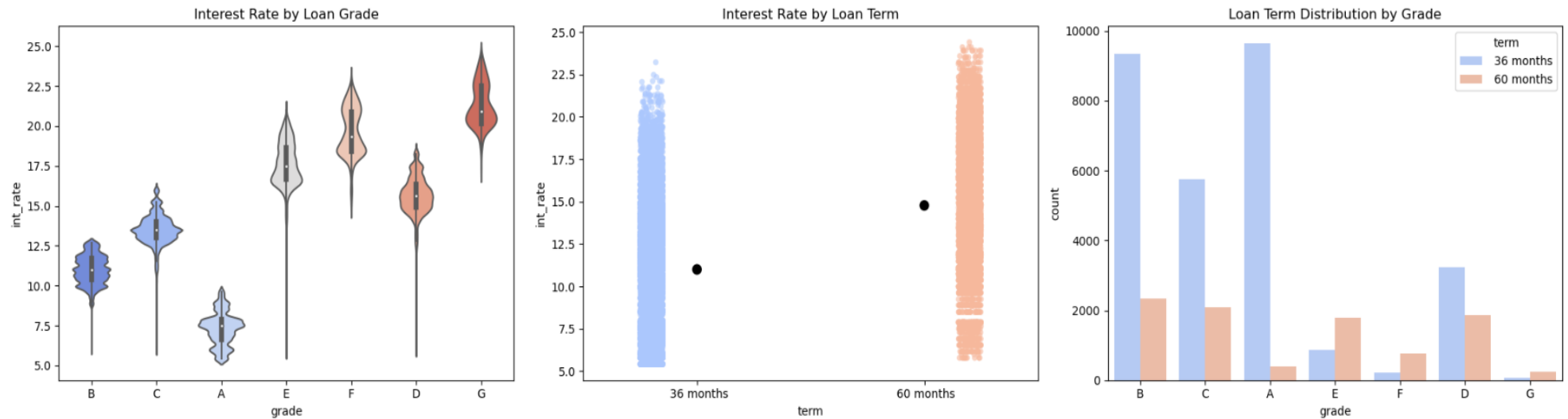
Segmented Univariate Analysis

#The below Graph shows that people are taking more loan for the purpose of Credit card payment or small business or debt consolidation or home improvement or house



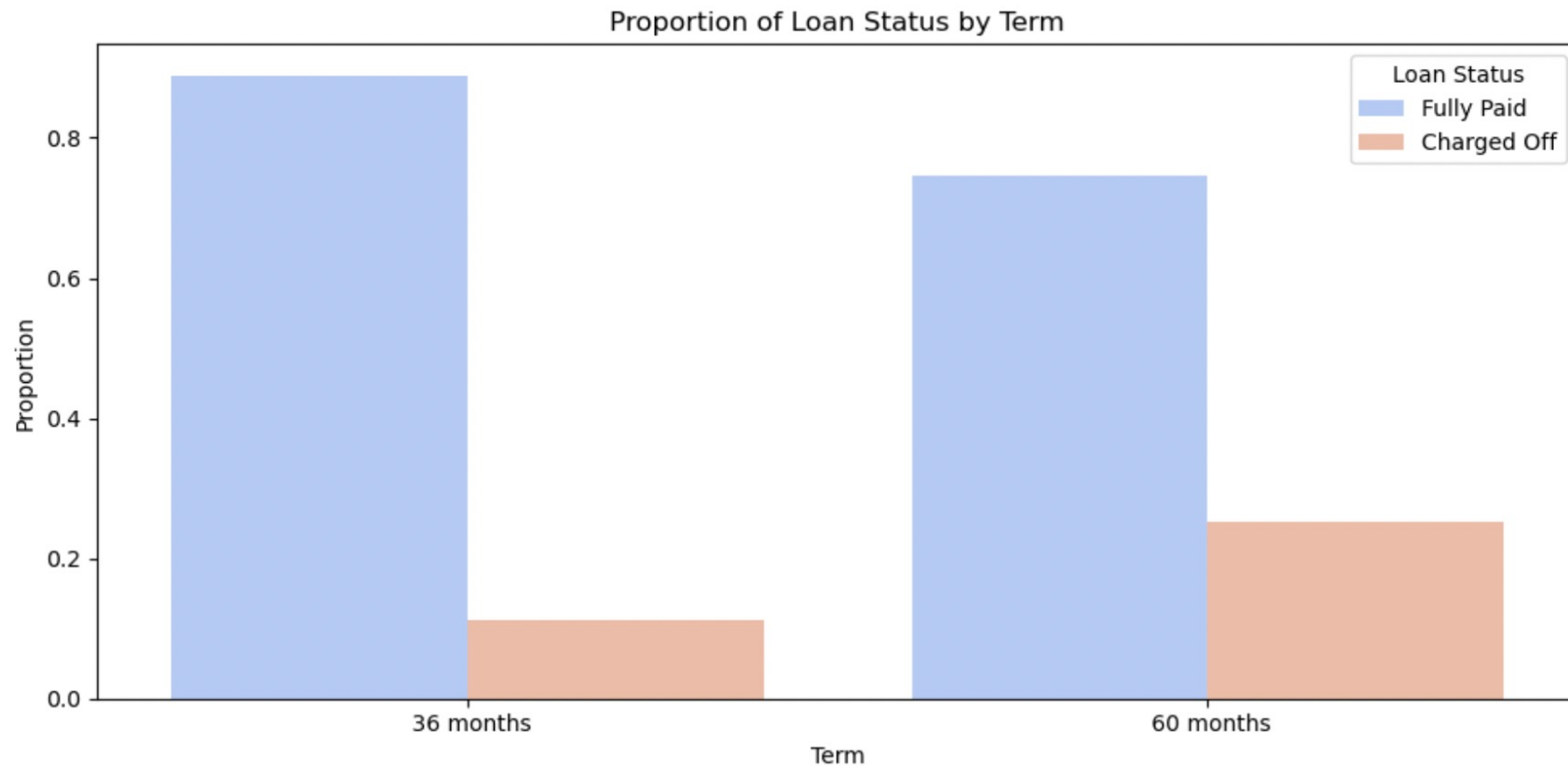
Bivariate Analysis

#Plot 1 shows interest rate is higher for higher Grade people
#plot 2 Shows that interest rate is higher for long term Loan
#Plot 3 shows People with A and B grade takes more Loan but the tenure of Loan is comparatively small which is 36 Mo.



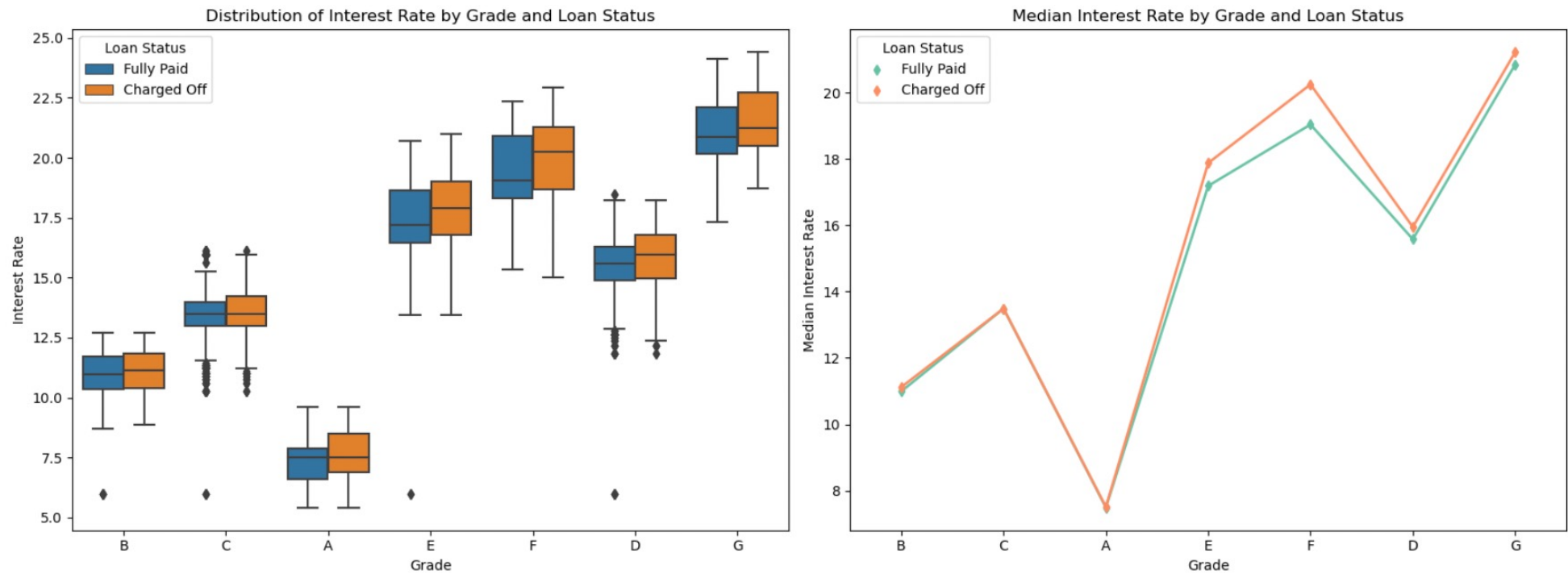
Analysis on Proportion of Loans Status vs Term

#This shows 60 months Loans has high risk of defaulter



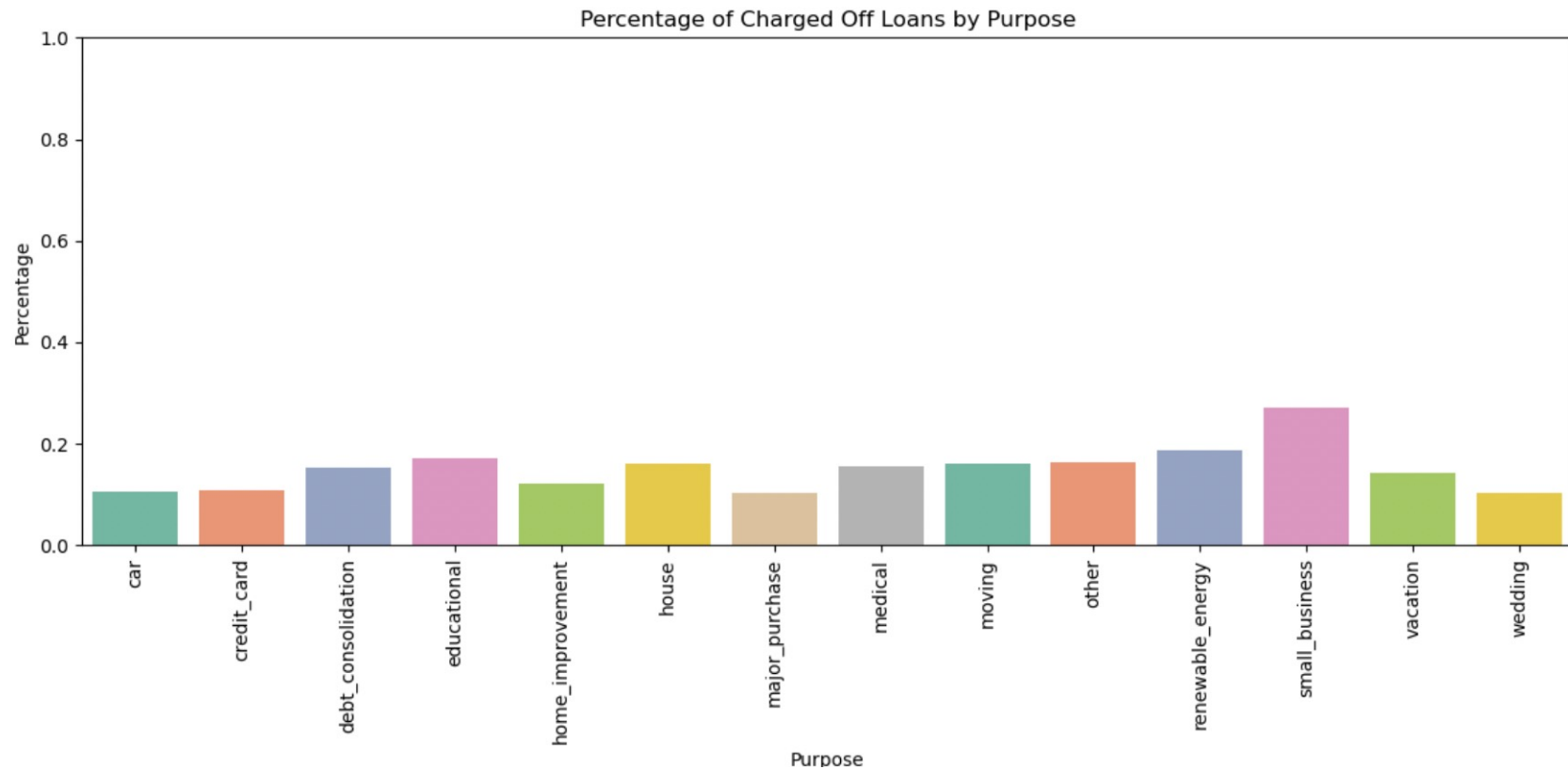
Analysis on interest rate vs grade

#This graph shows that people with Grade E,F,G has high interest rate and this people are more defaulters as well.



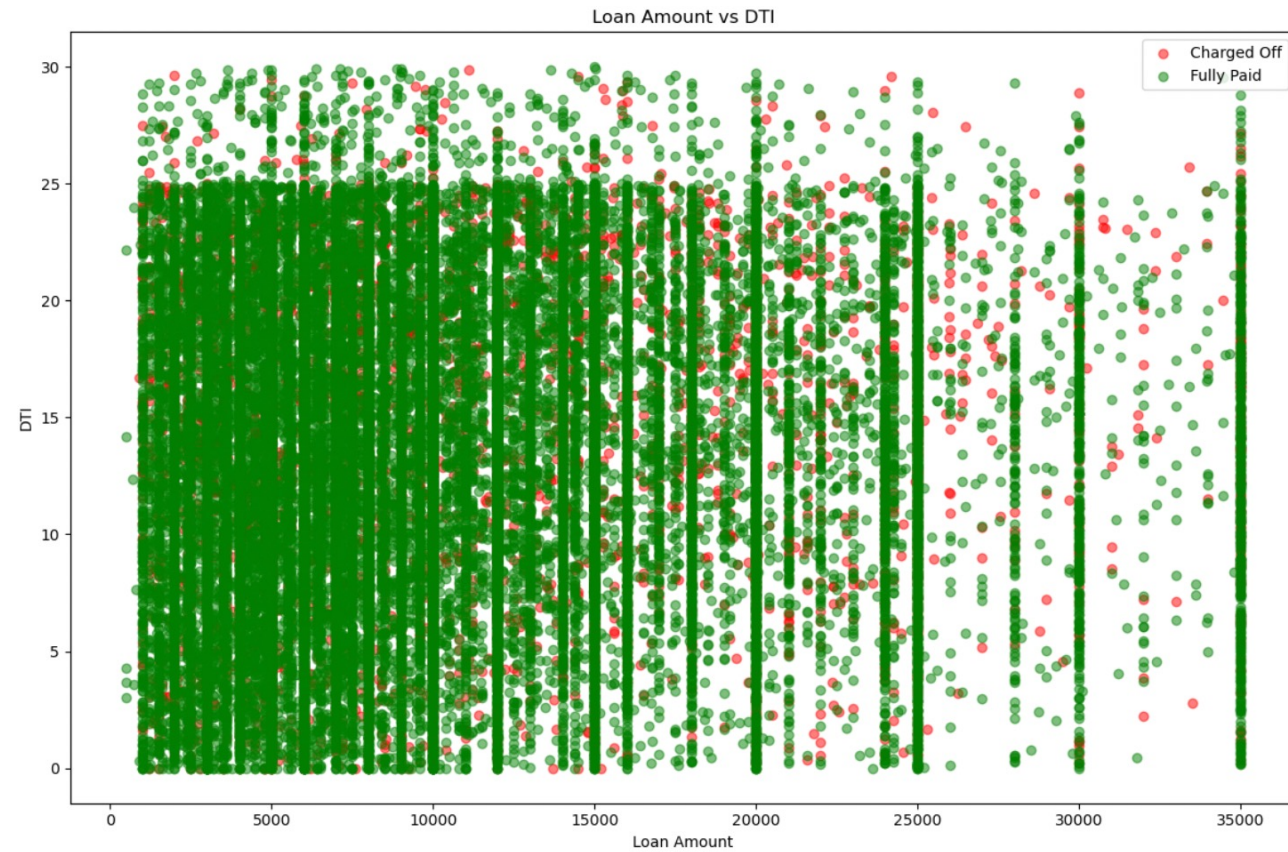
Analysis on charged off Loans Vs Purpose

#This analysis shows that small business purpose has highest number of defaulters

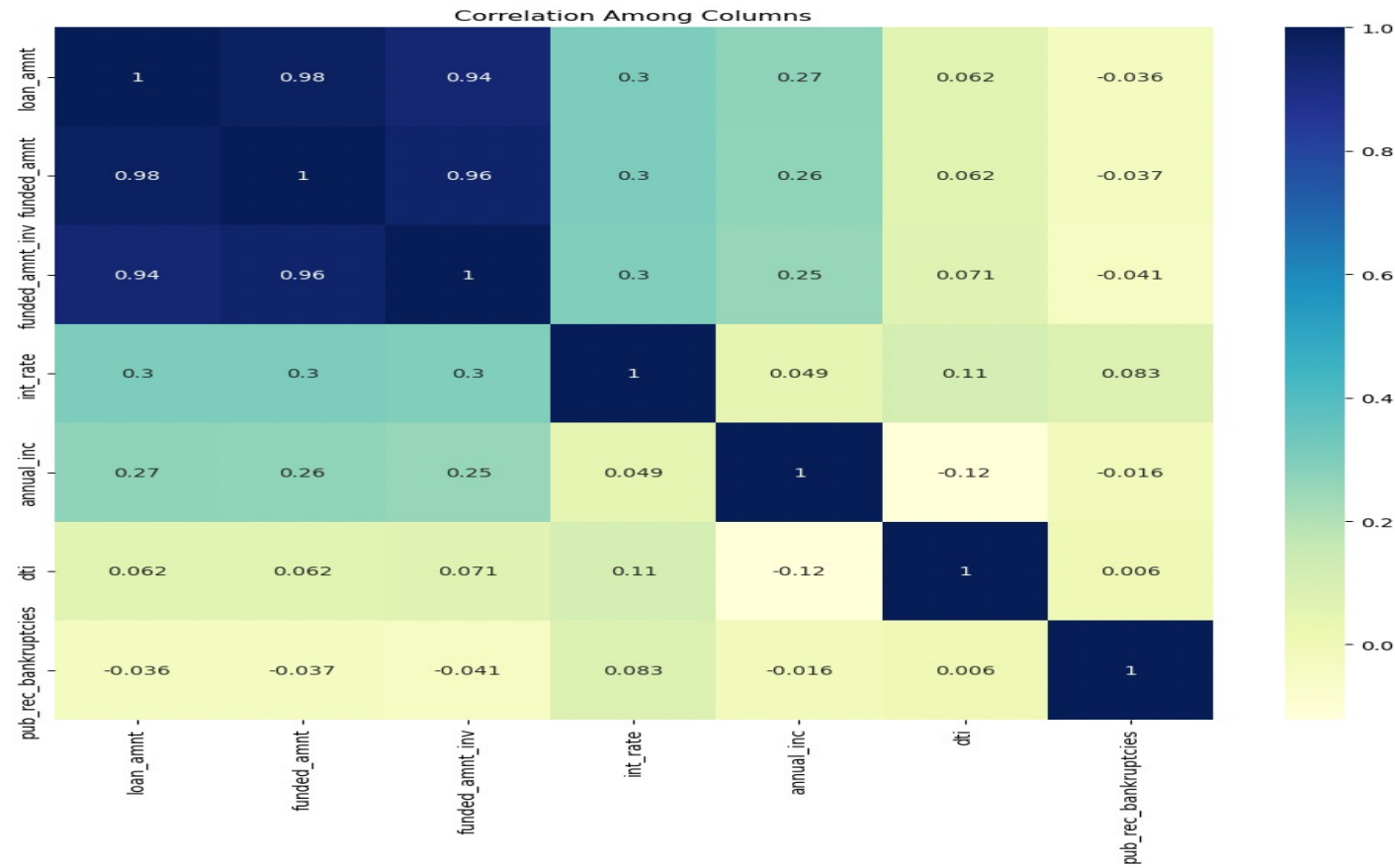


Analysis on Loan Amount Vs DTI

#This shows that more DTI and More the Loan Amount, There is higher chances of Default.



Correlation analysis



Inverse Relationships:

There is a negative correlation between the loan amounts requested (loan_amnt) and the incidences of public record bankruptcies (pub_rec_bankruptcies).

Similarly, the funded amounts (funded_amnt) and annual income exhibit negative correlations with debt-to-income ratio (dti).

Moderate Associations:

The size of the loan (loan_amnt) shows moderate positive correlations with the loan duration (term).

The loan duration (term) also moderately correlates with the interest rate charged (int_rate).

Strong Connections:

Strong positive correlations exist between the loan amounts (loan amount) and the actual funded amounts (funded amount).

Additionally, the funded amount from investors (funded_amnt_inv) demonstrates a robust correlation with the funded amount (funded amount).

Conclusion

- The analysis provides insights into factors influencing loan defaults. Key observations include:
- Higher loan amounts are associated with higher default risk.
- Interest rates vary significantly across loan grades and verification statuses.
- Certain loan purposes and borrower characteristics correlate with higher default rates.
- Maximum people who takes Loan is 10+ years experienced

Recommendations

- Based on the findings, recommendations for mitigating default risk include:
- Tighter scrutiny for higher loan amounts.
- Adjusting interest rates based on risk profiles identified.
- Monitoring loans issued during certain months or for specific purposes more closely.
- Tighter scrutiny for people with 10+ years of experience