LENDING CLUB CASE STUDY
UPGRAD

PROBLEM STATEMENT

Lending Club Case Study

Risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending various types of loans to urban customers.

Risks Involved

- 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

DECISIONS TAKEN BY THE COMPANY BASED ON APPLICANT

Loan Accepted

Fully paid: Applicant has fully paid the loan.
Current: Applicant is in the process of paying the instalments.
Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted**

Loan Rejected

Loan Rejected:
The company had rejected the loan for the candidates who does not meet the Loan Processes requirements so no data available

IDENTIFYING RISKS OF CREDIT LOSS

Using EDA To Identify the Reasons for defaulters

- Identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- In other words, 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 utilize this knowledge for its portfolio and risk assessment.

NECESSARY LIBRARIES

Numerical and Data Analysis

- ımporting numpy and pandas

Data Visualization

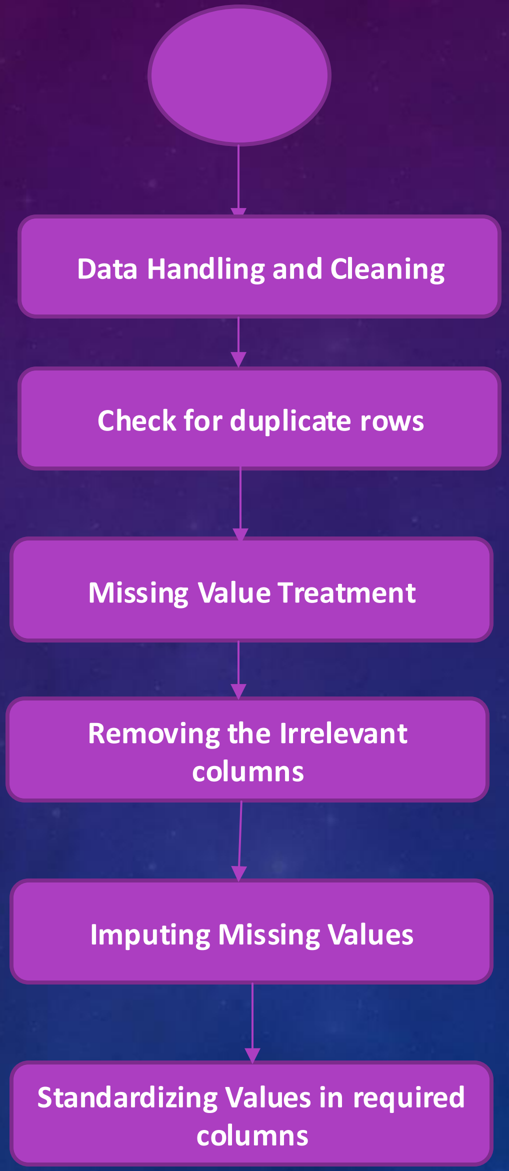
- Importing seaborn and matplot

Extra - To Suppress warnings

- Warnings

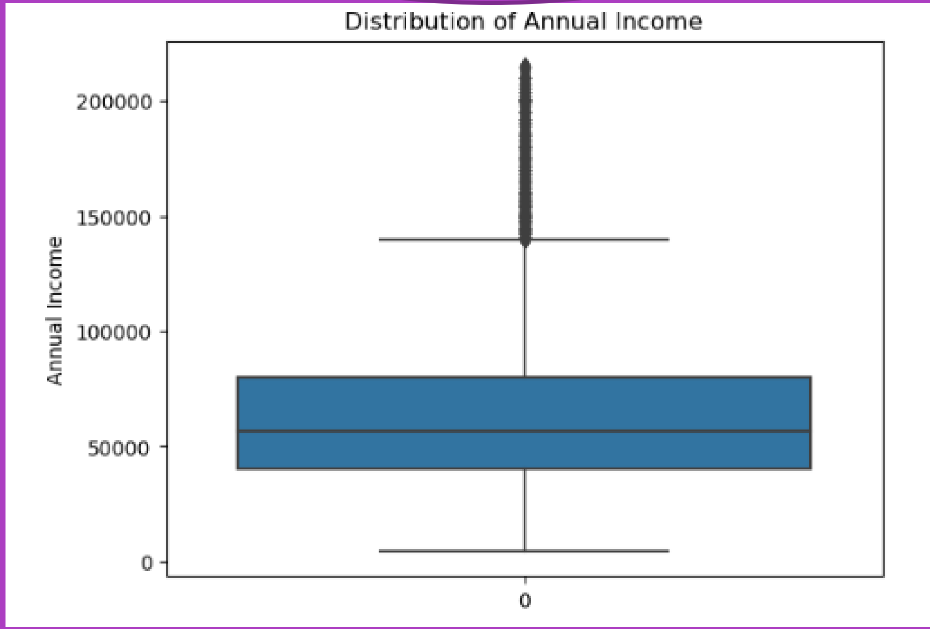
```
# Numerical and Data Analysis
import numpy as np
import pandas as pd

# Data Visualization
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

# Extra - To Suppress warnings
import warnings
warnings.filterwarnings('ignore')

# Display max columns
pd.set_option('display.max_columns', None)
```

ANALYSIS APPROACH

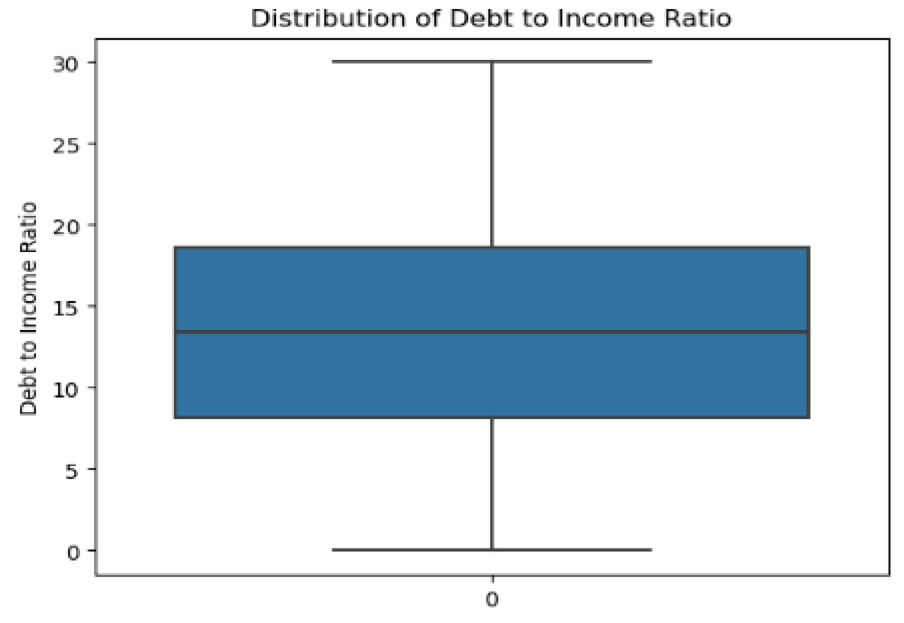


DISTRIBUTION OF ANNUAL INCOME

Distribution of Annual Income

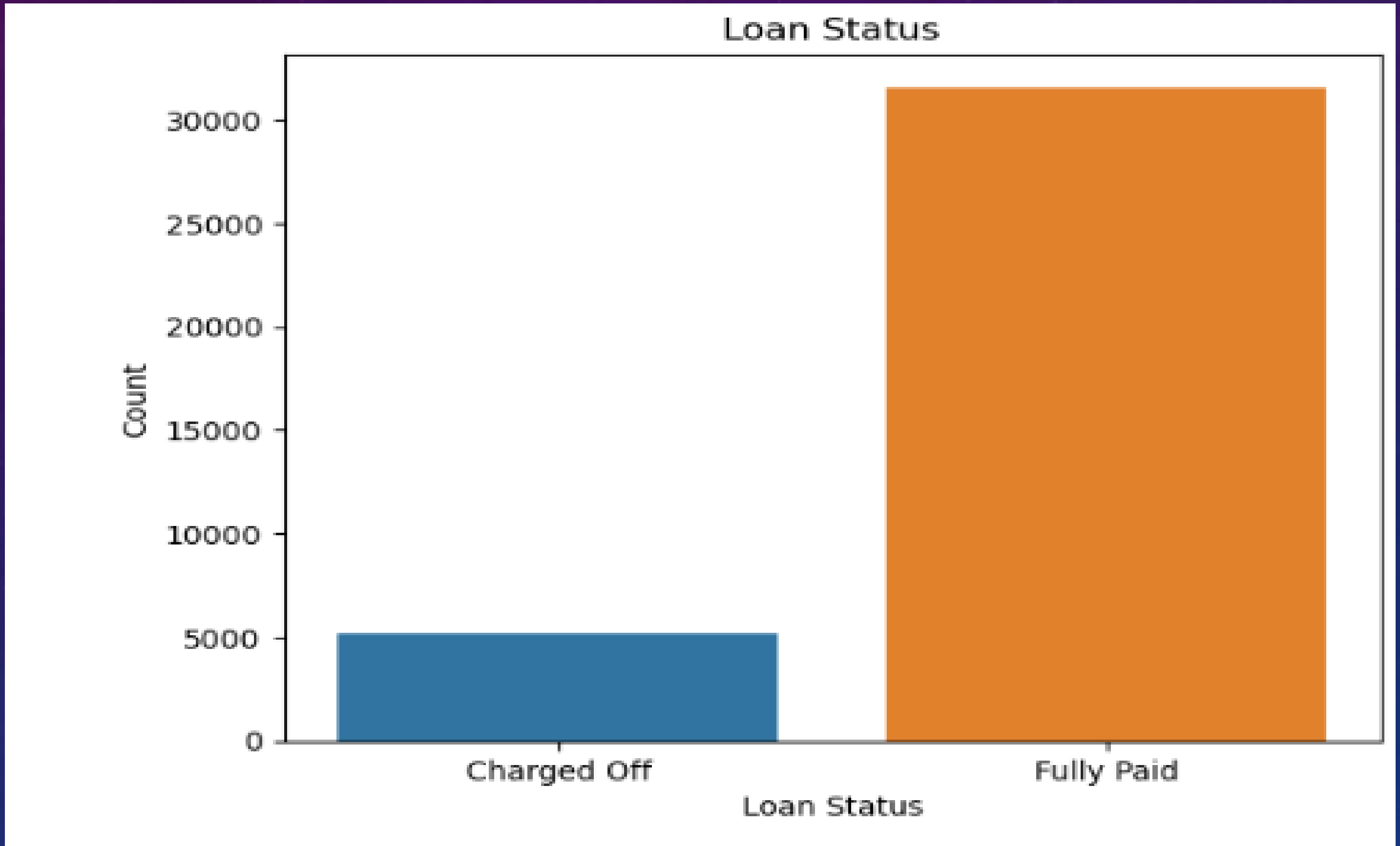


Distribution of Debt to Income Ratio



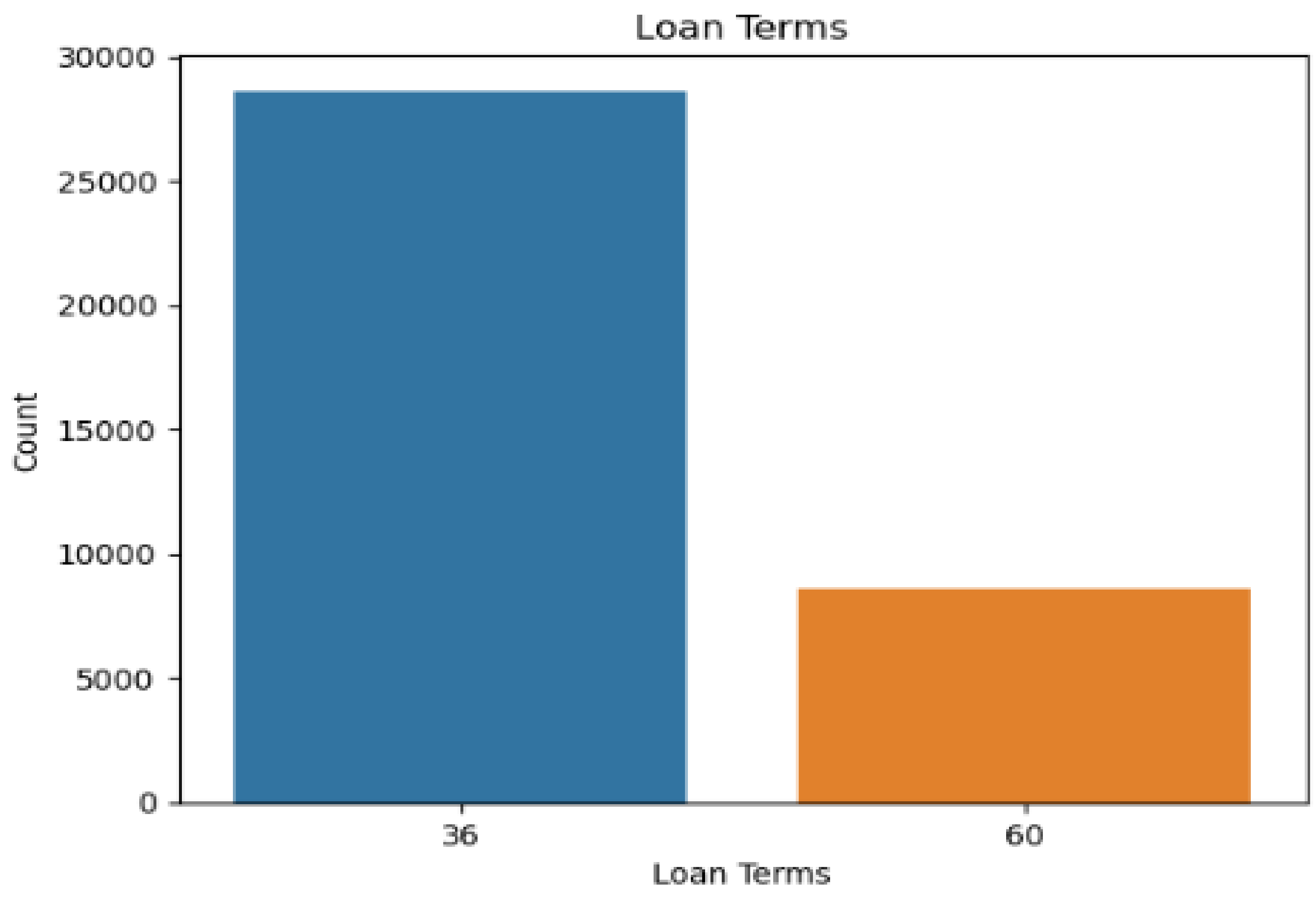
UNIVARIATE ANALYSIS

Univariate of Loan Status



Analysis shows that 85.90 % loans are fully paid and 14.09 % loans are charged off. It means defaulted loans are lower than fully paid loans

ANALYSIS OF LOAN TERMS

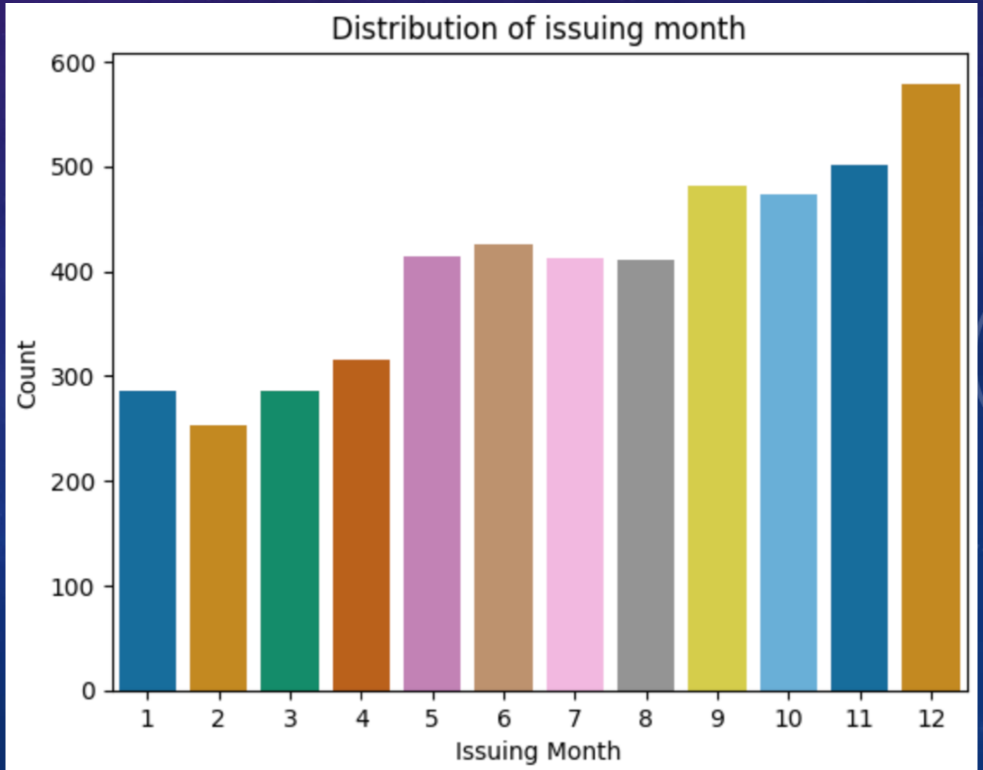


Analysis shows that 77.45 % loan terms are 36 months and 22.54 % loan terms are 60 months. It means people taking shorter loans than long duration loans.

ANALYSIS OF ISSUING MONTH

```
sns.countplot(x='month',
              data=data[data['loan_status']=='Charged Off'],
              palette='colorblind')
plt.title('Distribution of issuing month')
plt.ylabel('Count')
plt.xlabel('Issuing Month')
plt.show()
```

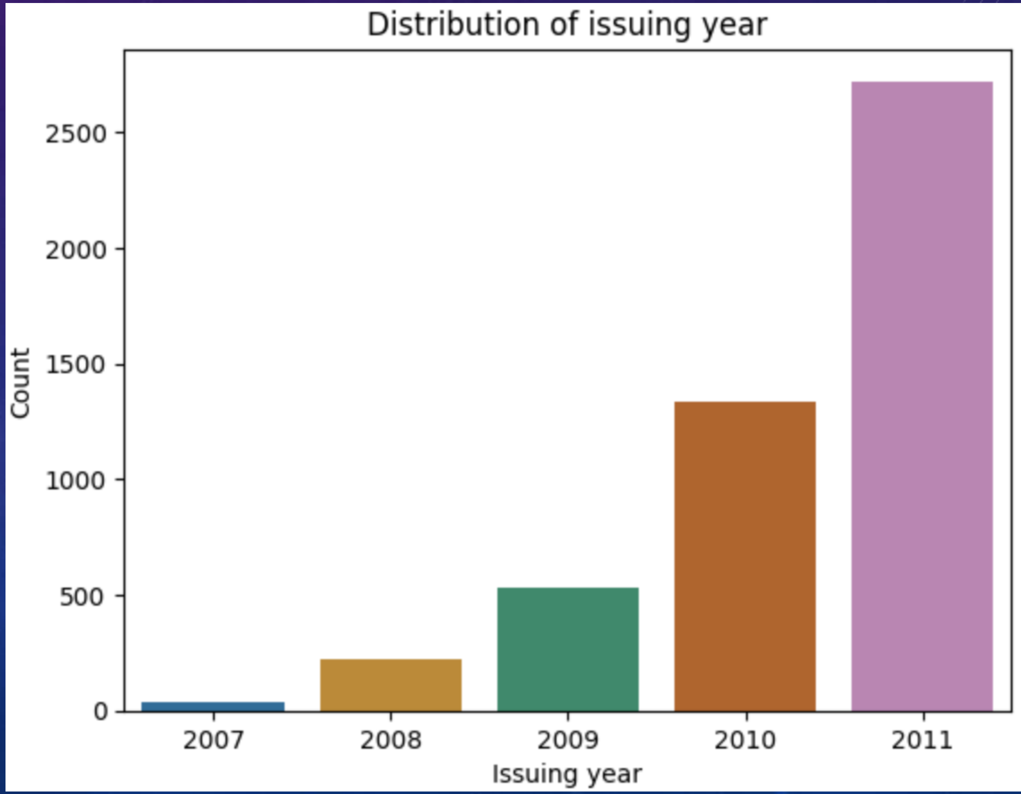
Analysis of Issuing Month: It is showing that loan count is increasing towards year ends. As year progress, loan counts increase in every month compare to last month.



ANALYSIS OF ISSUING YEAR

```
sns.countplot(x='year',
              data=data[data['loan_status']=='Charged Off'],
              palette='colorblind')
plt.title('Distribution of issuing year')
plt.ylabel('Count')
plt.xlabel('Issuing year')
plt.show()
```

Analysis of Issuing Year: It is showing that loan approval increasing exponential rate. Increasing every year compare to previous year.

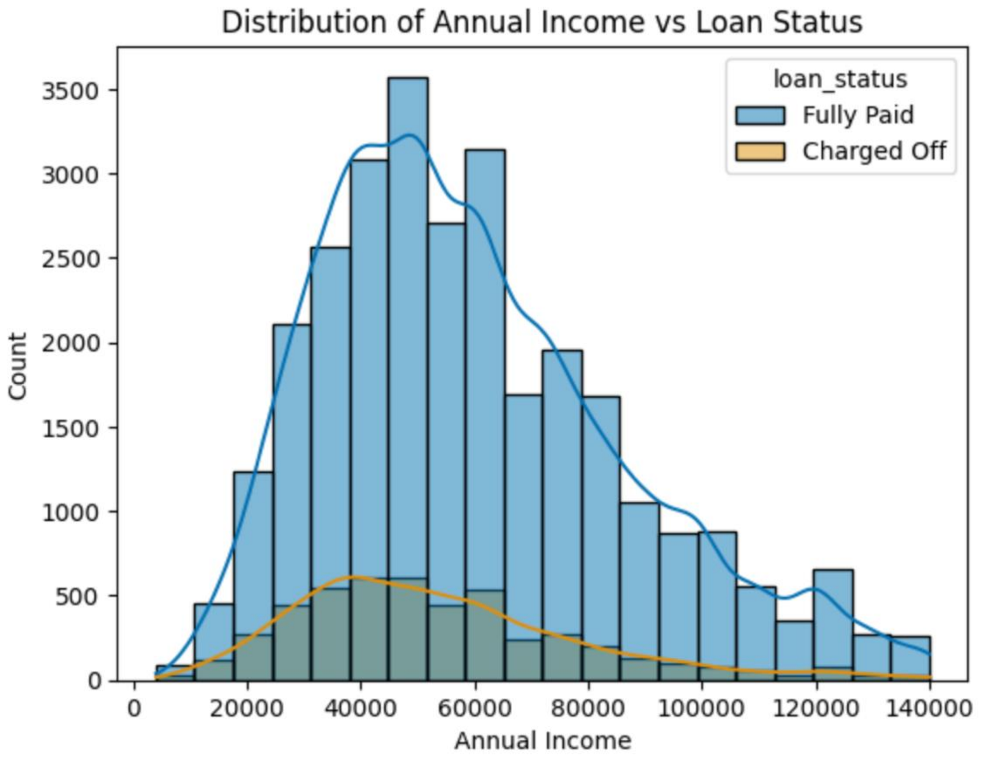


BIVARIATE ANALYSIS

- To find relationship between two sets of values.

```
sns.histplot(data =data, x='annual_inc',
             hue ='loan_status', kde=True,
             bins = 20,palette="colorblind")
plt.title('Distribution of Annual Income vs Loan Status')
plt.xlabel('Annual Income')
plt.ylabel('Count')
plt.show()
```

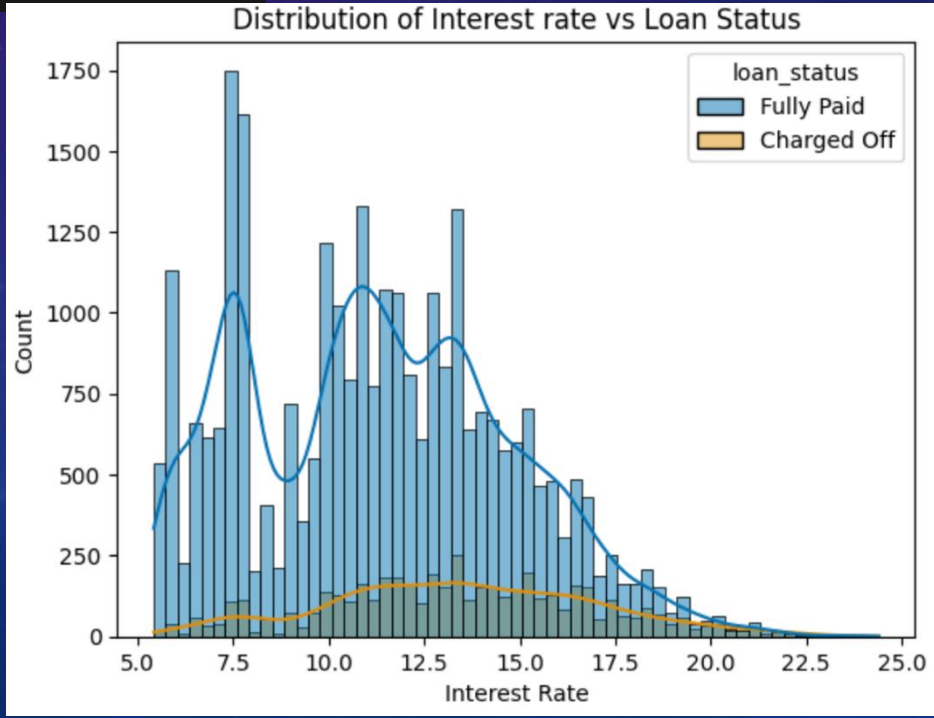
Analysis of Annual Income against Loan Status: Income group having income 50000 and less are more likely to default than higher income groups.



BIVARIATE ANALYSIS OF INTEREST RATE AGAINST LOAN STATUS

```
sns.histplot(data =data,x='int_rate',
             hue ='loan_status'
             palette="colorblind",kde=True)
plt.title('Distribution of Interest rate vs Loan Status')
plt.ylabel('Count')
plt.xlabel('Interest Rate')
plt.show()
```

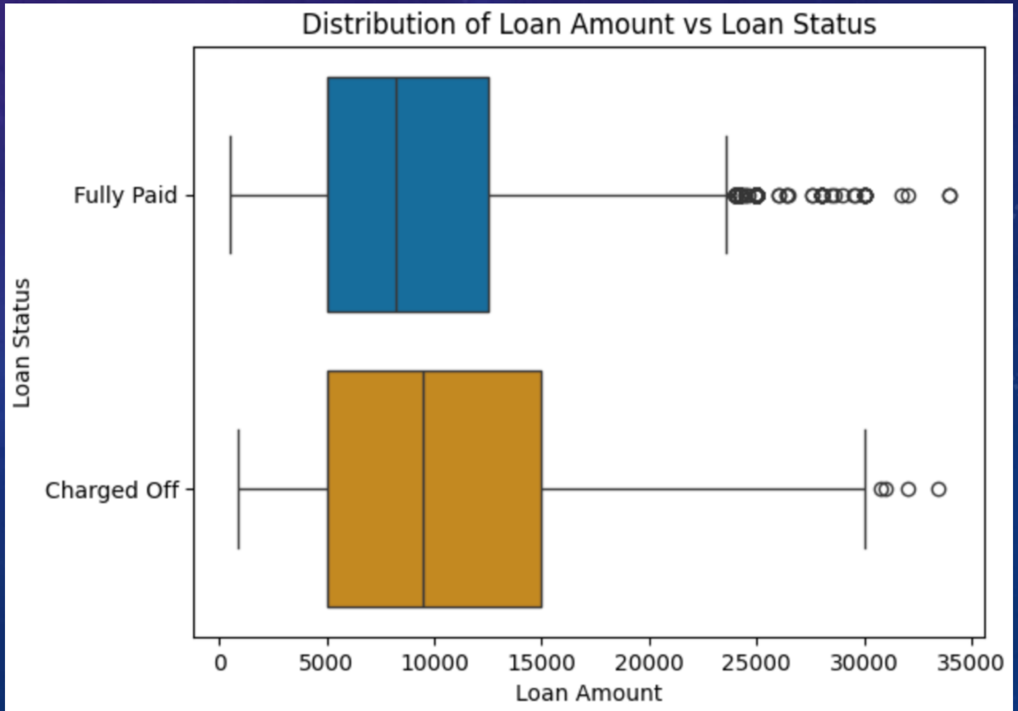
Analysis of Interest Rate against Loan Status: As Interest Rate of loan increases default ratio for loans is also increasing.



BIVARIATE ANALYSIS OF LOAN AMOUNT AGAINST LOAN STATUS

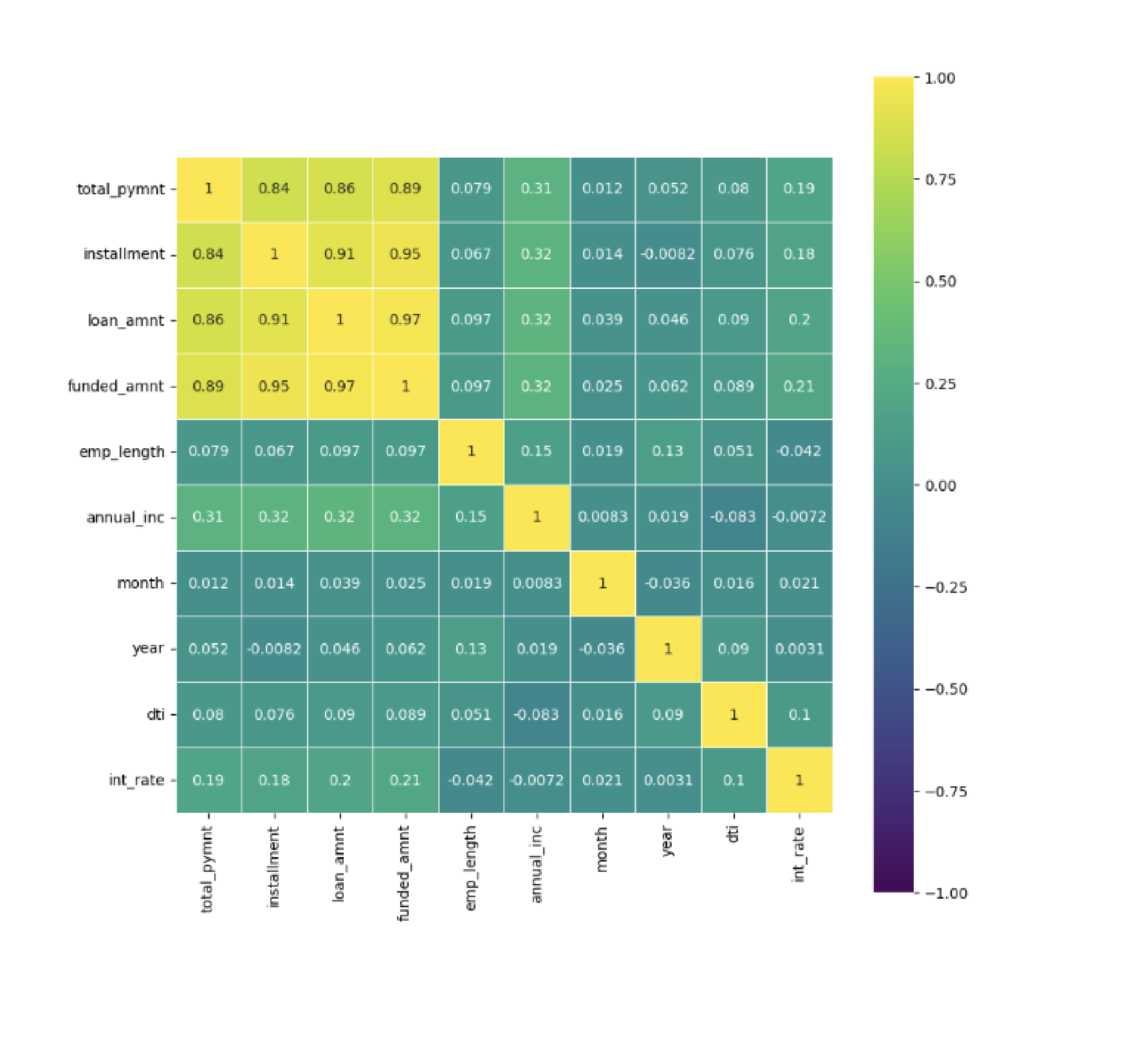
```
sns.boxplot(data =data,x='loan_amnt',
            y ='loan_status',palette="colorblind")
plt.title('Distribution of Loan Amount vs Loan Status')
plt.ylabel('Loan Status')
plt.xlabel('Loan Amount')
plt.show()
```

Analysis of Loan Amount against Loan Status: Fully Paid and Charged Off are almost same for 25% but Charged Off is getting higher after 50% and above than Fully Paid. It means higher the loan amount higher chances of getting defaulted.



VARIABLES OF LOAN DATASET.

```python
cols = ['total_pymnt', 'installment', 'loan_amnt', 'funded_amnt', 'emp_length',
        'annual_inc','month','year','dti','int_rate']
loan_correlation = data[cols].corr()
plt.figure(figsize=(8, 6))
sns.heatmap(loan_correlation,annot=True,
            cmap="viridis",
            vmin=-1,
            vmax=1, center=0, linewidths=0.5, square=True)
plt.show()
```



OBSERVATIONS

Univariate Analysis

- The number of defaulted loan is 6 times less than the number of fully paid loan.
- People have taken most of loan with 5 to 7.5 and 10 to 15 percent Interest Rate.
- People are mostly belongs to 10 years of employment. It shows that these people are most capable to take loans.
- Loans are taken by mostly lower income(20000 to 90000) group people.
- 50 % of people and above are verified by lending company or source verified.
- It is showing that loan count is increasing towards year ends. As year progress, loan counts increase in every month compare to last month.
- It is showing that loan approval increasing exponential rate. Increasing every year compare to previous year.

OBSERVATIONS

Bivariate Analysis

- Income group having income 50000 and less are more likely to default than higher income groups.
- As Interest Rate of loan increases default ratio for loans is also increasing.
- The employees having 10+ years of experience are less likely to default and higher chance to fully paid the loans.
- It shows that the Grade A(Sub Grade) has higher risk and the Grade(Sub Grade) G has lower risk. We can say higher grade(Sub Grade)
- People are making defaulter with Loan amount 10000 and less with None Home ownership. So, Bank should avoid giving loan these people.

OBSERVATIONS

Factors which can be used to predict about chances of loan default are:

- Income group
- Interest Rate
- Grade
- Home ownership
- DTI

THANK YOU !!