Lending club case study

Rajeshraj A K
Mohit Mehta

Problem statement

- Lending club is a consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant’s profile. Two types of risks are associated with the bank’s decision.

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
2. 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.

Objective

- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. Using exploratory data analysis (EDA) this objective needs to be achieved with the help of dataset which has information about past loan applicants.
- 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.

Understanding the dataset

- The Loan dataset contains information about past loan applicants and whether they ‘defaulted’ or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. The dataset does not contain the rejection criteria details. The loan process involves three steps.

1. Borrower requests a loan amount
2. The approver decides to approve the loan based on past history and verification (funded_amnt).
3. The investor determines the final loan amount to be offered (funded_amnt_inv).

Data Analysis

- **Attributes**
  - **Primary attributes**
    - **Loan Status :** This is the primary attribute and has 3 possible scenarios
      - Fully-Paid: Applicant has fully paid the loan (the principal and the interest rate).
      - Charged-Off: Applicant is in the process of paying the installments, i.e. the tenure of the loan is not yet completed.
      - Current: Applicant has not paid the installments in due time for a long period, i.e. he/she has defaulted on the loan
  - **Key columns**

  The following columns, often referred to as predictors, are crucial. These attributes, available during the loan application process, provide insights into whether the loan can be approved or rejected.
    - **Customer Demographics**
      - **Annual income** (annual_inc) :- The self-reported annual income provided by the borrower during registration.
      - **Home Ownership** (home_ownership) :- The home ownership status provided by the borrower during registration.
      - **Employee length of service** (emp_length) :- Employment length of employee in years.
      - **Debt to income** (dti) :- A ratio calculated using the borrower's total monthly debt payments against income. Lower dti will always have higher chance of loan approval.
      - **State** (addr_state) :- The state provided by the borrower in the loan application

Data Analysis contd.

- **Loan characteristics**
  - **Loan Amount** (loan_amnt) :- The listed amount of the loan applied for by the borrower
  - **Funded amount** (funded_amnt) :- The total amount committed to that loan at that point in time.
  - **Funded amount by investor** (funded_amnt_inv) :- The total amount committed by investors for that loan at that point in time.
  - **Term** (term) :- The number of payments on the loan.
  - **Rate of interest** (int_rate) :- Interest Rate on the loan
  - **Monthly installment** (installment) :- The monthly payment owed by the borrower if the loan originates.
  - **Grade** (grade) :- LC assigned loan grade.
  - **Sub grade** (sub_grade) :- LC assigned loan subgrade.
  - **verification_status** (verification_status) :- Loan is Verified / Not Verified
  - **Date loan issued** (issue_d) :-The month which the loan was funded.
  - **Purpose of taking loan** (purpose) :-A category provided by the borrower for the loan request.

- **Excluded Columns** Certain columns are excluded from the analysis as they do not provide insights into whether the loan can be approved or rejected, nor do they help in predicting potential defaults.
  1. **Post approval activities** : This does not provide any details or insights relevant to our goal and will not be included in our analysis.
  2. **Granular data** : Columns providing highly detailed information which is not required for our analysis.

Dataset row Analysis

- **Header & Footer Rows:** No header or footer rows in the dataset.
- **Extra Rows:** No column numbers, indicators, etc., found in the dataset.
- **Summary Rows:** No summary rows in the dataset.
- **Dropping Loan Status Rows with "Current" status:** Rows with a "loan_status" of "Current" will be dropped as they represent loans in progress.
- **Identifying and Removing Duplicate Rows:** Duplicate rows in the dataset will be identified and removed if present.

Dataset Column Analysis

- **Dropping Columns**
  - 54 columns had null values which will be dropped.
  - 9 columns had Single value which will be dropped.
  - 3 Columns had more than 60% of missing data will be dropped.
  - Columns also dropped which do not contribute to our analysis.

- **Converting Column Format**
  - The term column will have the "months" text stripped and will be converted to an integer.
  - Percentage columns like (int_rate) are currently in object format. These columns will have the "%" character stripped and will be converted to float.
  - The issue_d column will be converted to datetime format with date in YYYY-mm-dd format.
  - The emp_length column will be converted to an integer. The columns have "Years ,+ & <". This will be stripped and converted to integer

Dataset Column Analysis contd.

- **Standardizing Values**
  - Similar values for home_ownership like NONE and OTHER has been standardized as OTHER.
  - Similar values for verification_status like Source Verified and Verified has been standardized as Verified
  - Imputing Missing Values: Columns with a lower percentage of missing values will undergo imputation to fill in the missing data.

- **Handling Outliers**
  - Approach here is to identify outliers and remove them.
  - Here the outliers are verified on following columns loan_amnt , funded_amnt,funded_amnt_inv,annual_inc, int_rate,installment and dti

Data cleaning

- Load Loan Dataset
- Import necessary libraries
- Identifying and removing columns having complete null values.
- Verifying and remove columns had single values
- Verifying and remove columns that having missing data more than 60%
- Checking and dropping Duplicate records (No records found duplicate)
- Deleting loan records that has 'CURRENT' Status.
- Removing columns which not directly add value to analysis.
- Identifying and imputing columns thar had null values.
  - emp_length & pub_rec_bankruptcies has null values.
  - For emp_length , NULL values replaced with most commonly used value (10+ Years)
  - For pub_rec_bankruptcies , Records with Null values deleted

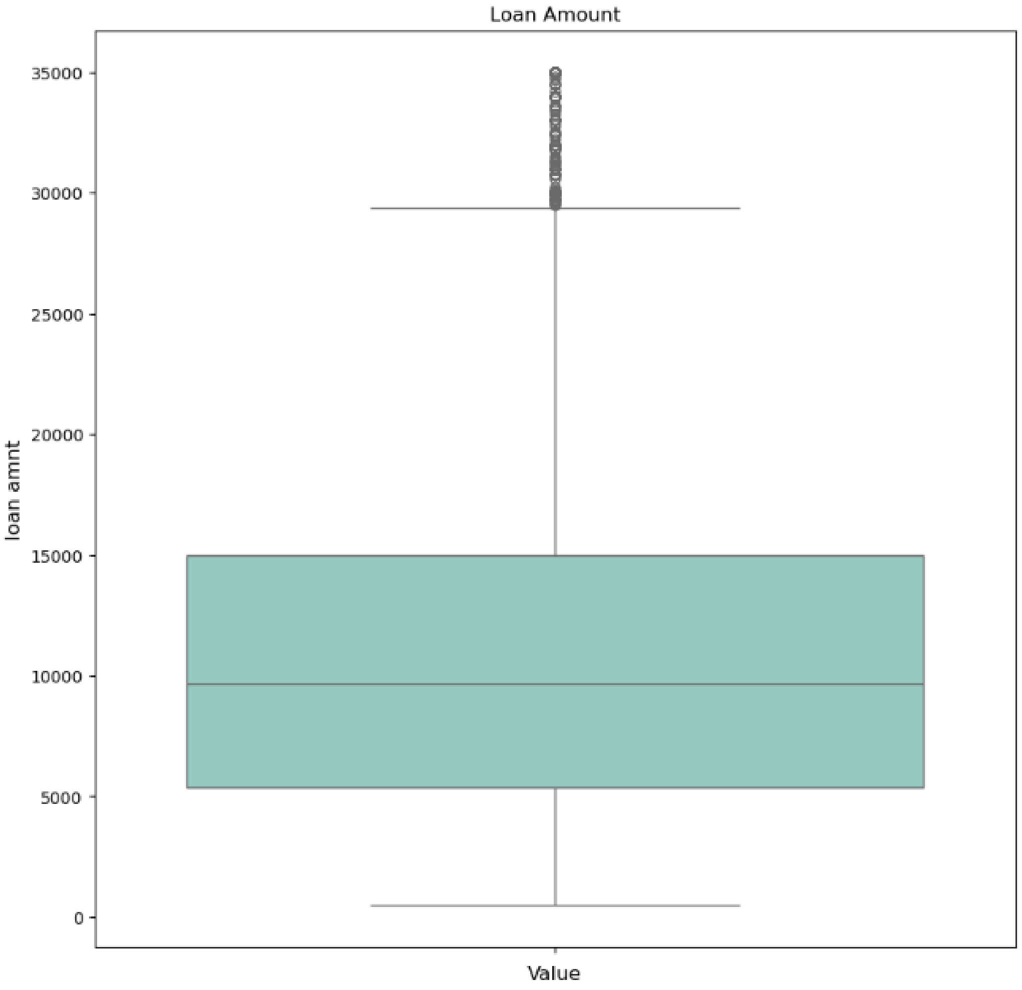
Outliers

- Approach here is to identify outliers and remove all outliers
- Here the outliers are verified on following columns loan_amnt , funded_amnt,funded_amnt_inv,annual_inc, int_rate,installment and dti

Verifying Outliers for Loan amount

```
# verifying outliers for Loan amount
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['loan_amnt'],orient='v' , palette ='Set3')
plt.title('Loan Amount')
plt.xlabel('Value' ,fontsize = 12)
plt.ylabel('loan amnt', fontsize = 12)
plt.show()
```

Majority of Loan applicants loan amount is between 5000 to 15000



Outliers – Contd.

Verifying Outliers for Annual income

```
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['annual_inc'],orient='v' , palette ='Set3')
```

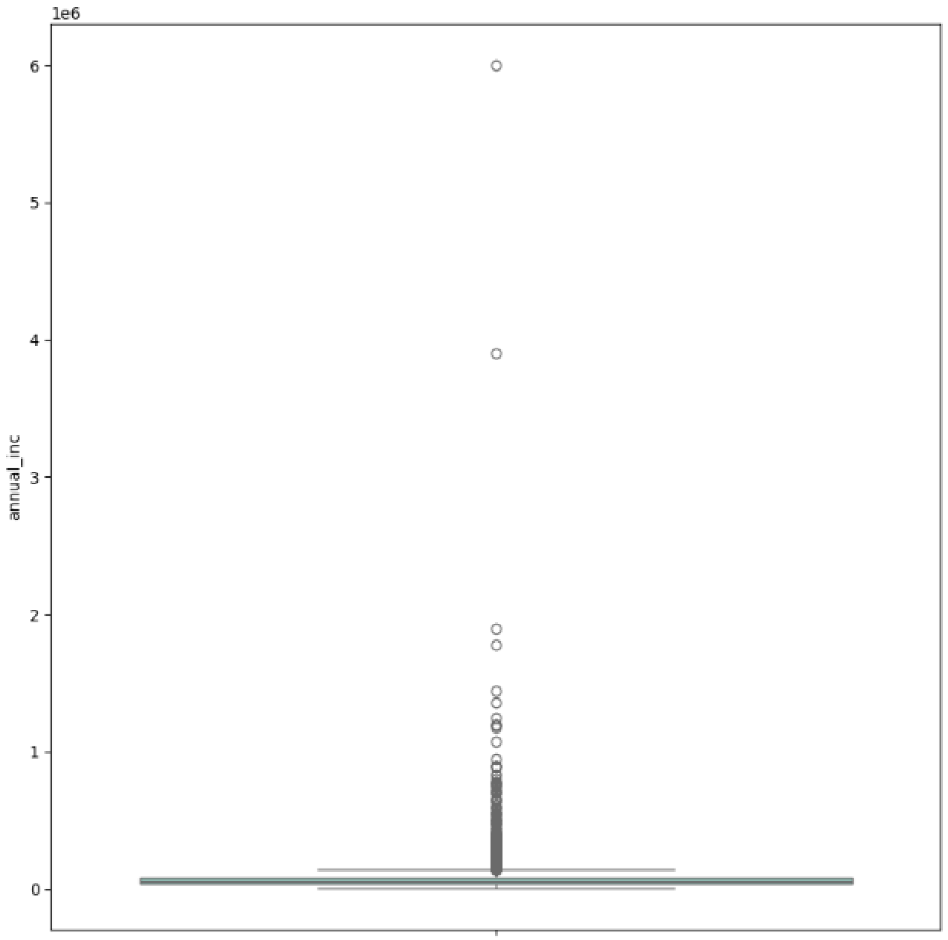
- Annual income has outliers.
- Next step is to Calculate the inter-quartile range (IQR) and filtering out the outliers outside of lower and upper bound

```
# calculating lower and upper bound

first_quantile = df_loan['annual_inc'].quantile(0.25)
third_quantile = df_loan['annual_inc'].quantile(0.75)
iqr = third_quantile-first_quantile
lbound = first_quantile - ( 1.5 * iqr )
ubound = third_quantile + ( 1.5 * iqr )
df_anual_inc_ll = df_loan.annual_inc > lbound
df_anual_inc_up = df_loan.annual_inc < ubound

df_net_income = df_anual_inc_ll & df_anual_inc_up

# filetering outliers and geting annual income within lower and upper bound range
df_loan=df_loan[df_net_income]
```

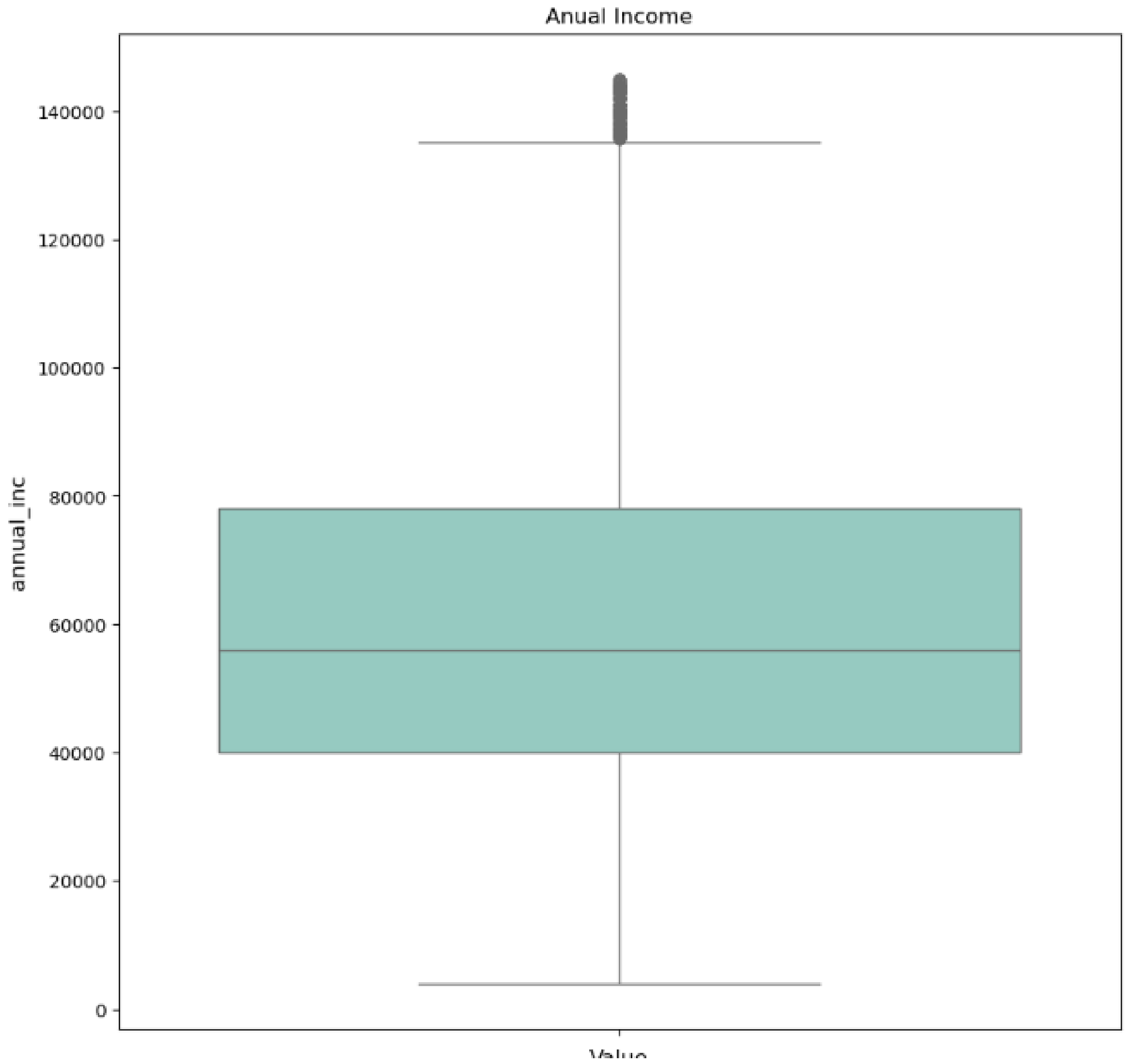


Outliers – Contd.

Annual income after removing outliers

```
# Verifying annual income after removing outliers
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['annual_inc'],orient='v' , palette ='Set3')
plt.title('Anual Income')
plt.xlabel('Value' ,fontsize = 12)
plt.ylabel('annual_inc', fontsize = 12)
plt.show()
```

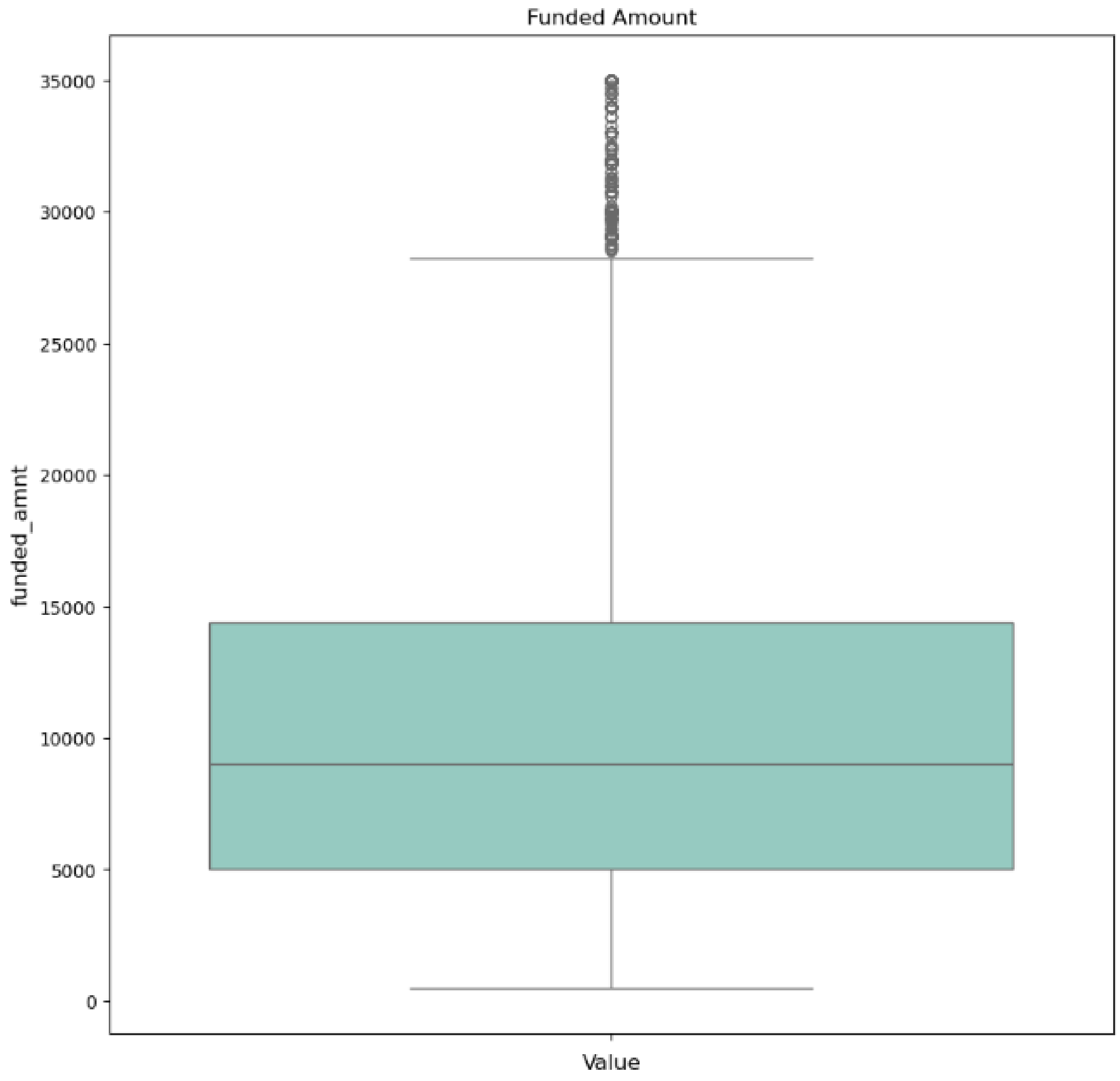
- The Annual income for most of the loan applicants are between 40K and 78K



Outliers – Contd.

Verifying Outliers for Funded amount

```
# verifying outliers for Funded amount
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['funded_amnt'],orient='v' , palette ='Set3')
plt.title('Funded Amount')
plt.xlabel('Value' ,fontsize = 12)
plt.ylabel('funded_amnt', fontsize = 12)
plt.show()
```



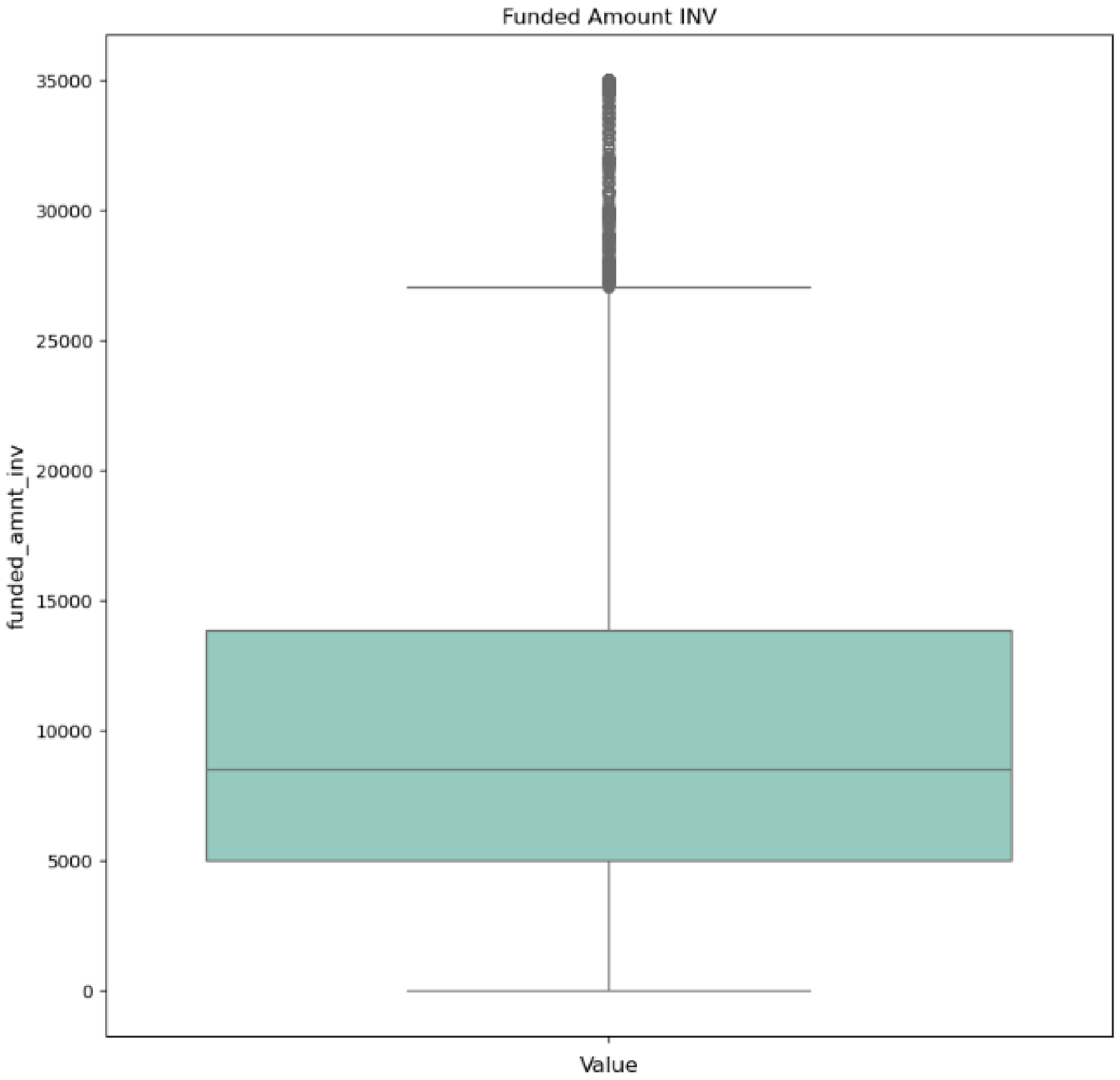
- The funded amount for majority of the applicants is between 5K and 14K

Outliers – Contd.

Verifying Outliers for Funded amount INV

```
# Verifying outliers for Funded amount by invester
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['funded_amnt_inv'],orient='v' , palette ='Set3')
plt.title('Funded Amount INV')
plt.xlabel('Value' ,fontsize = 12)
plt.ylabel('funded_amnt_inv', fontsize = 12)
plt.show()
```

- The funded amount by investor for majority of the applicants is between 5K and 14K

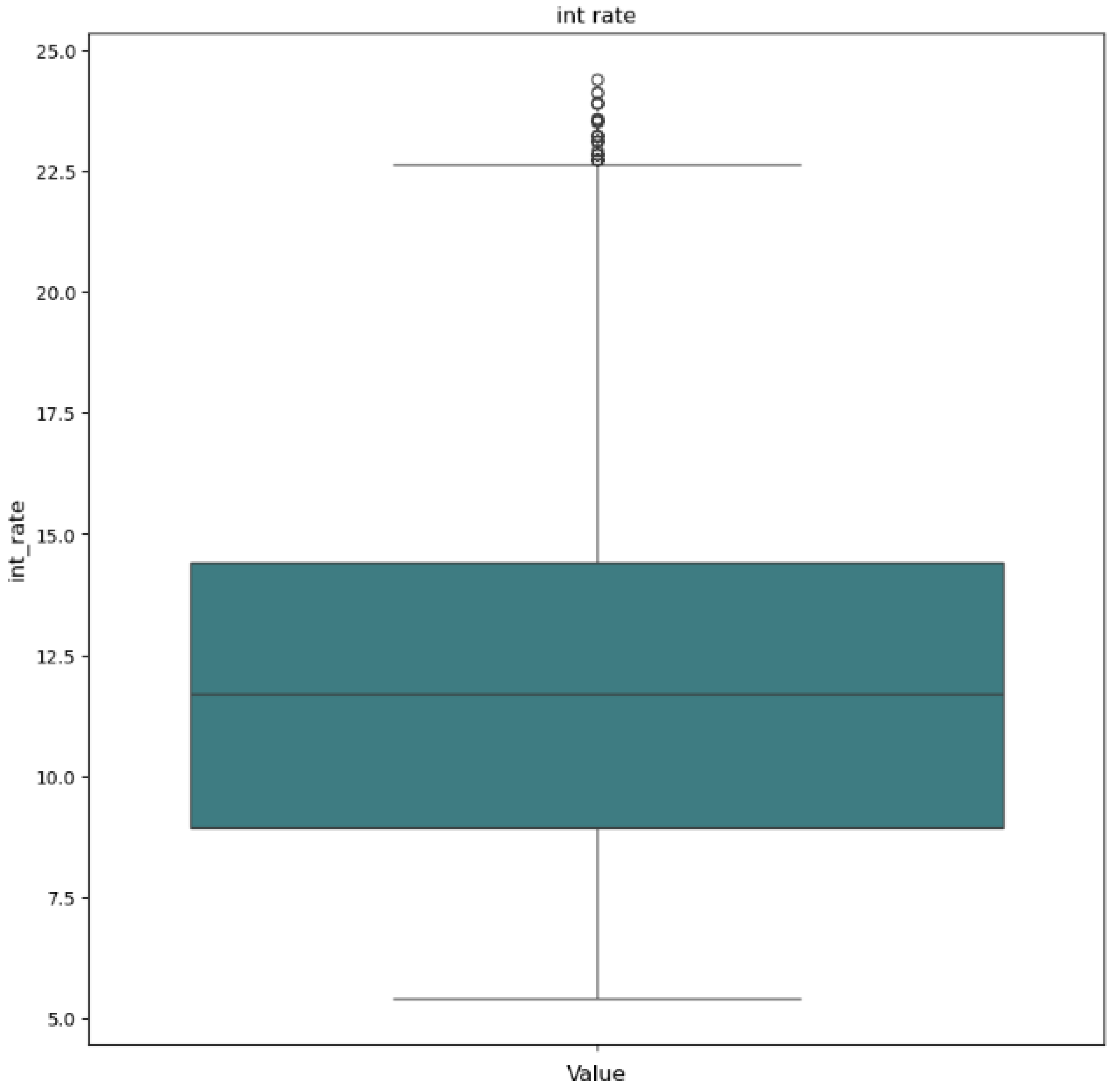


Outliers – Contd.

Verifying Outliers for Rate of interest

```
# Verifying outliers for interest rate
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['int_rate'],orient='v' , palette ='crest')
plt.title('int rate')
plt.xlabel('Value' ,fontsize = 12)
plt.ylabel('int_rate', fontsize = 12)
plt.show()
```

The interest rate for majority of the applicants is 9% to 14%

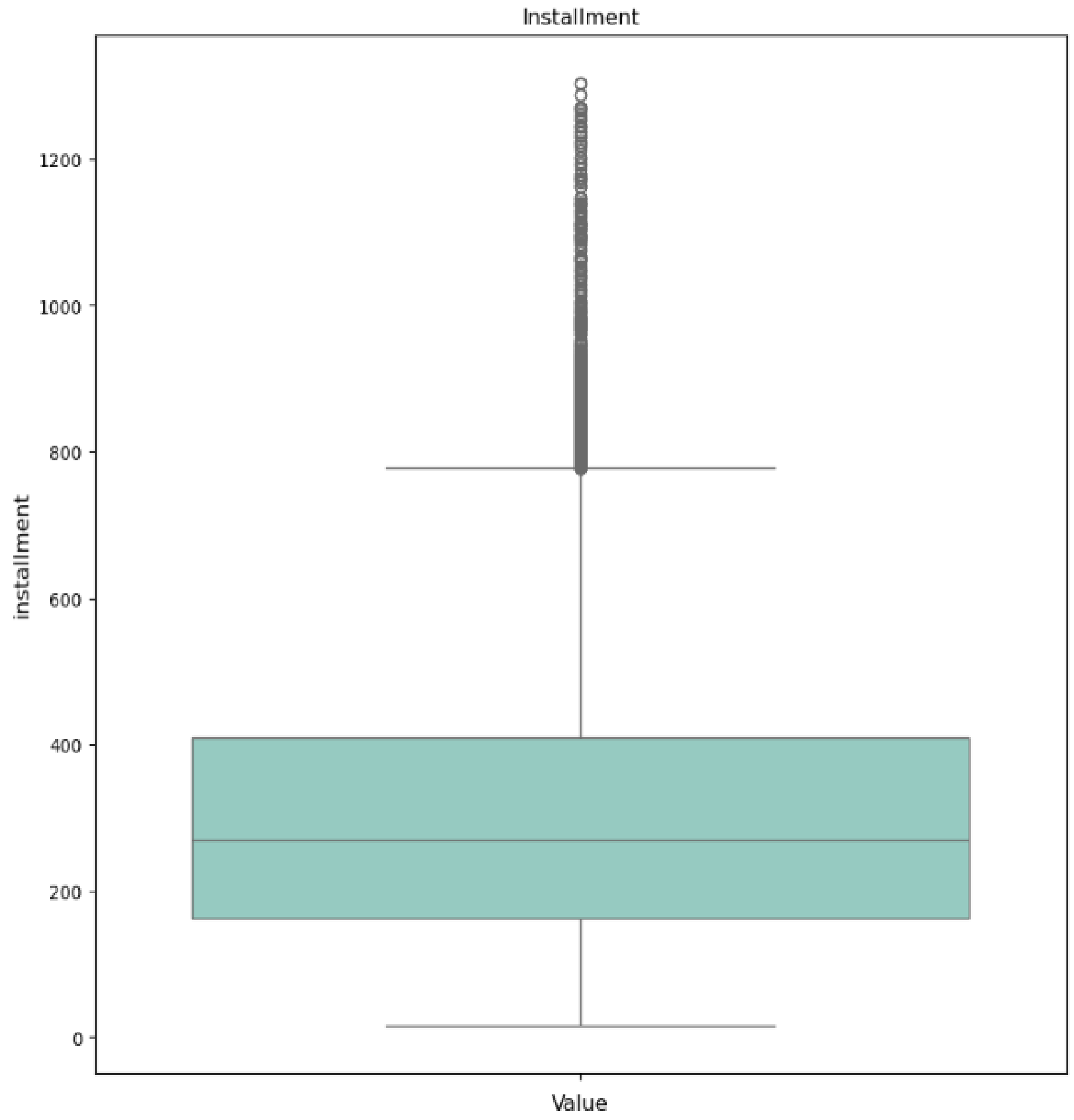


Outliers – Contd.

Verifying Outliers for installment

```
# Verifying outliers for installment
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['installment'],orient='v' , palette ='Set3')
plt.title('Installment')
plt.xlabel('Value' ,fontsize = 12)
plt.ylabel('installment', fontsize = 12)
plt.show()
```

- The installment for majority of the applicants is in between 163 and 400

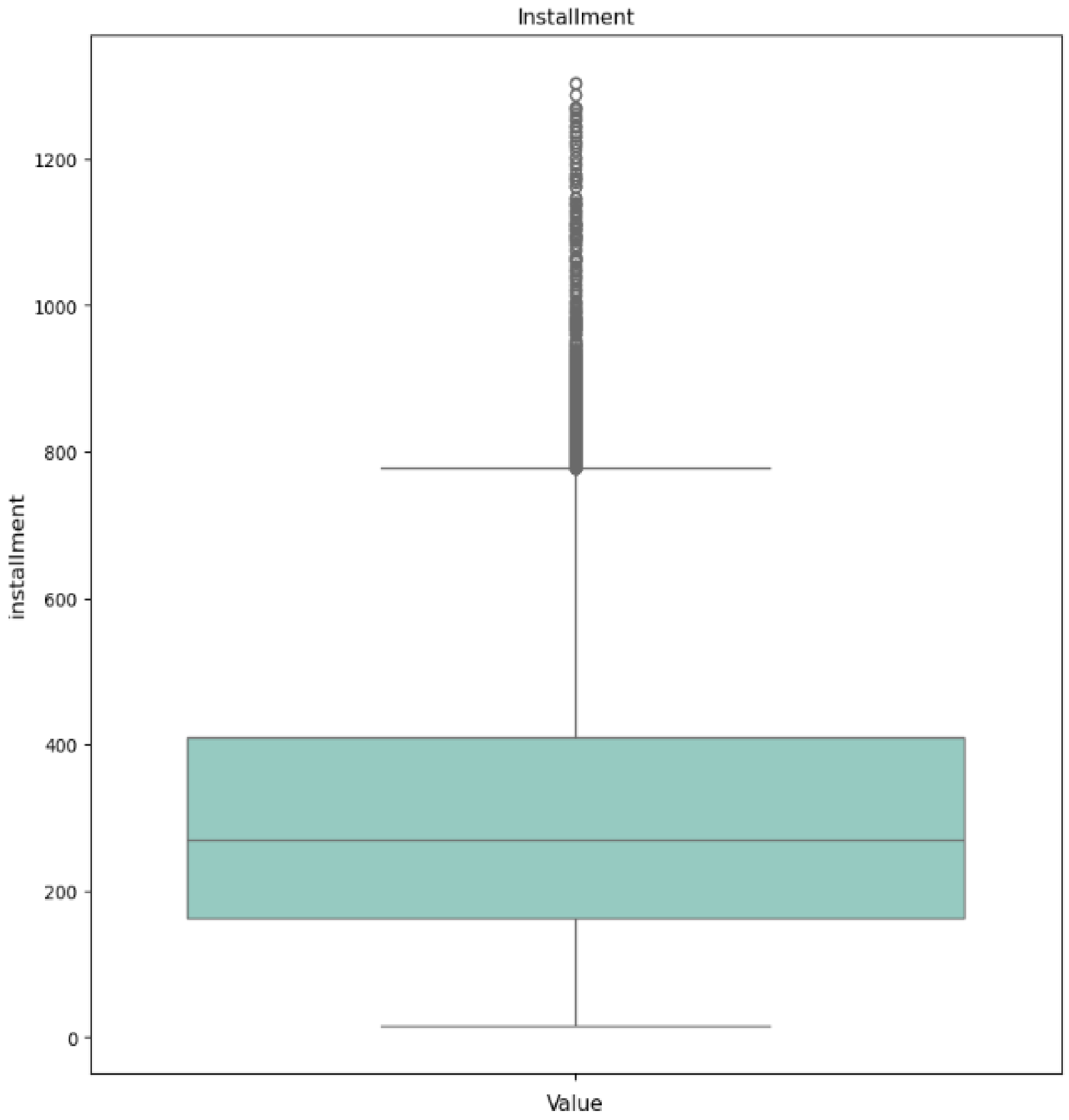


Outliers – Contd.

Verifying Outliers for installment

```
# Verifying outliers for installment
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['installment'],orient='v' , palette ='Set3')
plt.title('Installment')
plt.xlabel('Value' ,fontsize = 12)
plt.ylabel('installment', fontsize = 12)
plt.show()
```

- The installment for majority of the applicants is in between 163 and 400

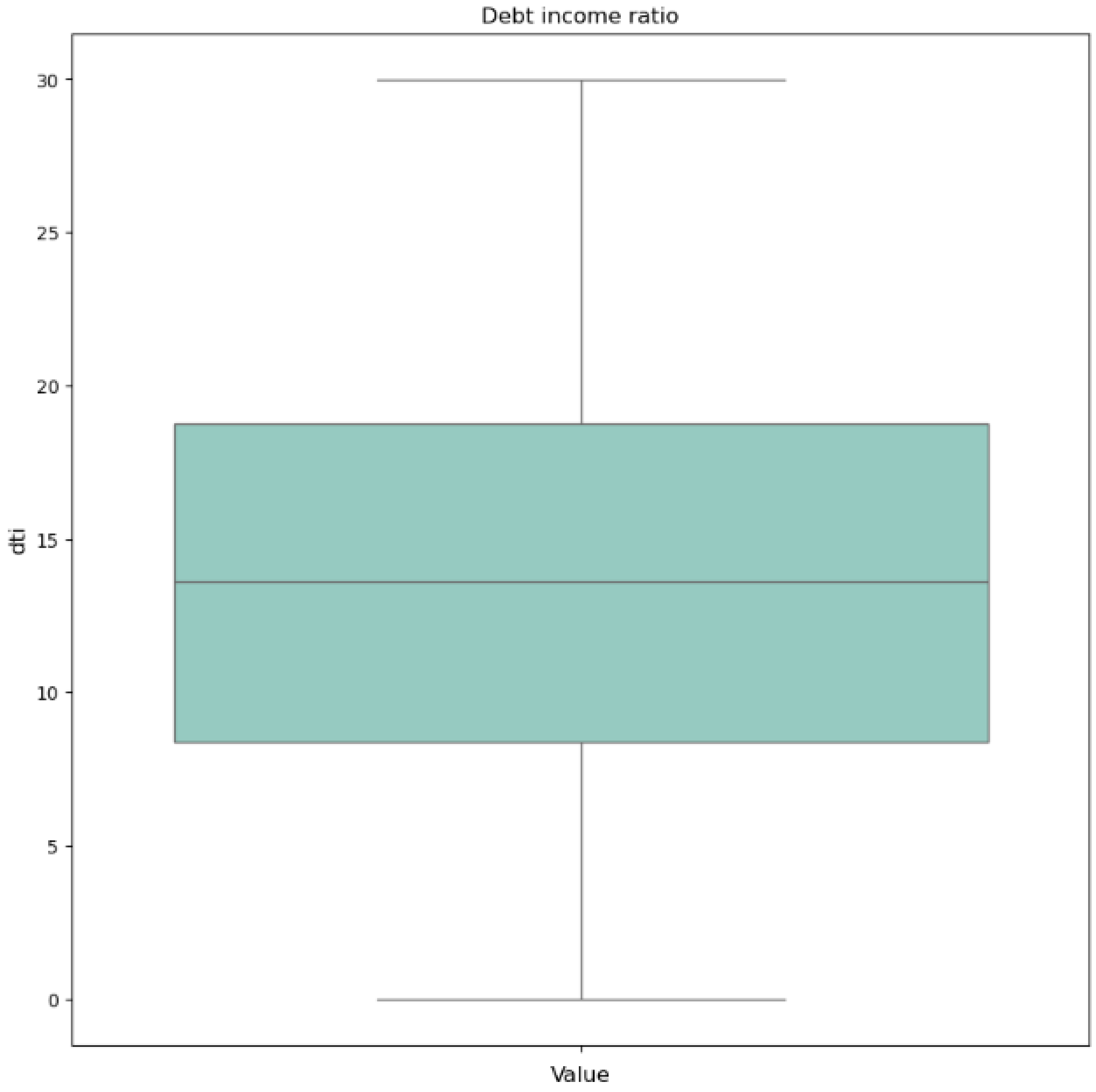


Outliers – Contd.

Verifying Outliers for DTI

```
# Verifying outliers for dti
plt.figure(figsize =(10,10))
sns.boxplot(y=df_loan['dti'],orient='v' , palette ='Set3')
plt.title('Debt income ratio')
plt.xlabel('Value' ,fontsize = 12)
plt.ylabel('dti', fontsize = 12)
plt.show()
```

- The Debt-to-income ratio for majority of applicants are between 8 to 19



Outliers – Summary

- Majority of Loan applicants loan amount is between 5000 to 15000.
- Annual income has outliers. Filtered out the outliers outside of lower and upper bound.
- After removing outliers, the annual income for most of the loan applicants are between 40K and 78K.
- The funded amount for majority of the applicants is between 5K and 14K.
- The funded amount by investor for majority of the applicants is between 5K and 14K.
- The interest rate for majority of the applicants is 9% to 14%.
- The installment for majority of the applicants is between 163 and 400.
- The Debt-to-income ratio for majority of applicants is between 8 and 19.

Derived columns

1. New columns are month and year of loan issued date and then another column to identify quarter wise.
2. Bucketing Loan amount, funded amount, annual income and DTI.

Univariate analysis

- For this analysis we will be focusing only charged off loans
- Classified Ordered categorical variable, Un-Ordered categorical variable and Quantitative variables

Ordered categorical Variables
1.term -- The number of payments on the loan.
2.grade -- LC assigned loan grade
3.sub_grade -- LC assigned loan subgrade
4.emp_length -- Employment length in years
5.mnth_issued -- Month in which loan issued
6.yr_issued -- Year in which loan isssued
7.quarter -- in which quarter loan was issued

Un-Ordered categorical variable
1.Home Ownership - Whether the home is Rented / own house / under mortgage
2.Loan status - whether loan is fully paid / Charged off
3.Verification status - whether loan is Verified / Not verified
4.purpose - Reason for taking loan
5.State - The state to which loan applicant belongs.

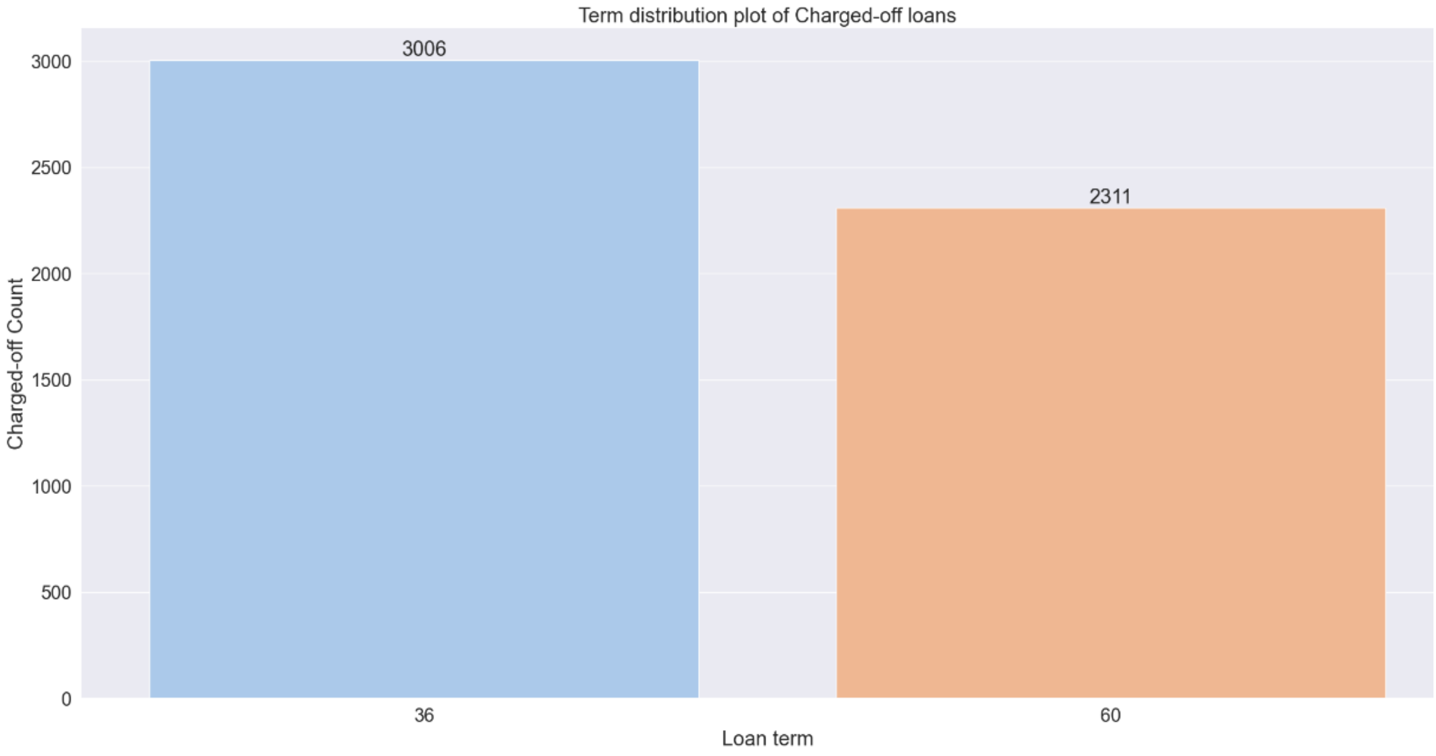
Quantitative variables
1.installment - Loan installment
2.Loanamt_bucket - Loan amount in multiple bins
3.funded_amnt_bucket - Funded amount in multiple bins
4.annual_inc_bucket - Annual income in multiple bins
5.int_rate_bucket - Rate of interest in multiple bins
6.dti_bucket - Debt to income in multiple bins

Ordered categorical Variables -Term

```
# Creating Seaborn categorical plot to check the Term distribution plot of Charged-off loans
catplot=sns.catplot(x='term',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0)
catplot.set(title='Term distribution plot of Charged-off loans', xlabel='Loan term', ylabel='Charged-off Count')

# for annonate
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

Loans with a 36-month term are most likely to default. 3006 Loans under 36 months got defaulted. This indicates that individuals default on short-term loans.

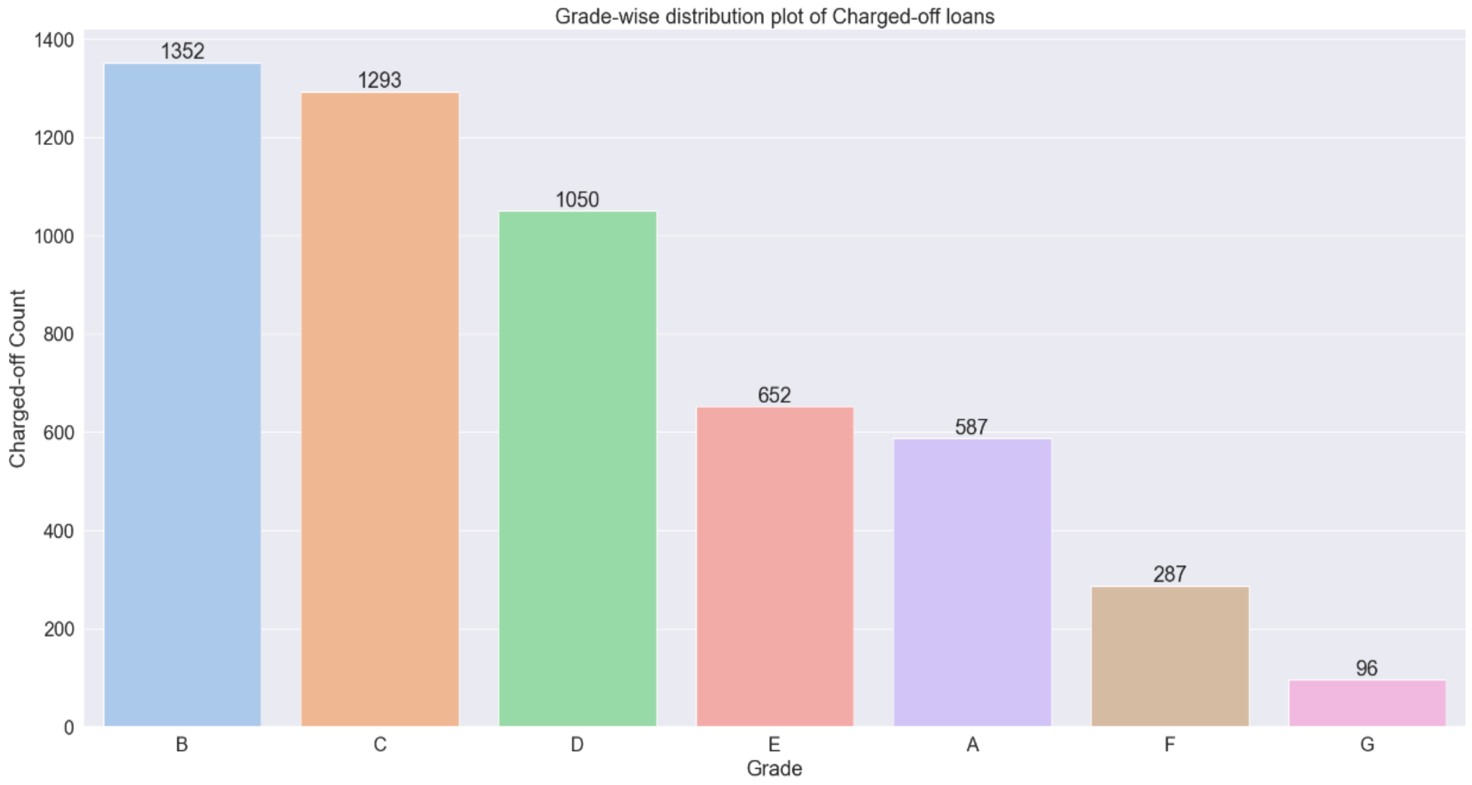


Ordered categorical Variables -Grade

```
# Creating Seaborn categorical plot for Grade wise distribution plot of Charged-off loans
catplot=sns.catplot(x='grade',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0,
                   order=df_loan_chargedoff['grade'].value_counts().index)
catplot.set(title='Grade-wise distribution plot of Charged-off loans', xlabel='Grade', ylabel='Charged-off Count')

# to Annonate
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

Grade B had the highest number of Charged off loan applicants, with a total of 1,352 applicants. The next is C with 1293 loan applicants. This indicates the lending club should pay attention for providing loan to B & C Grade applicants.

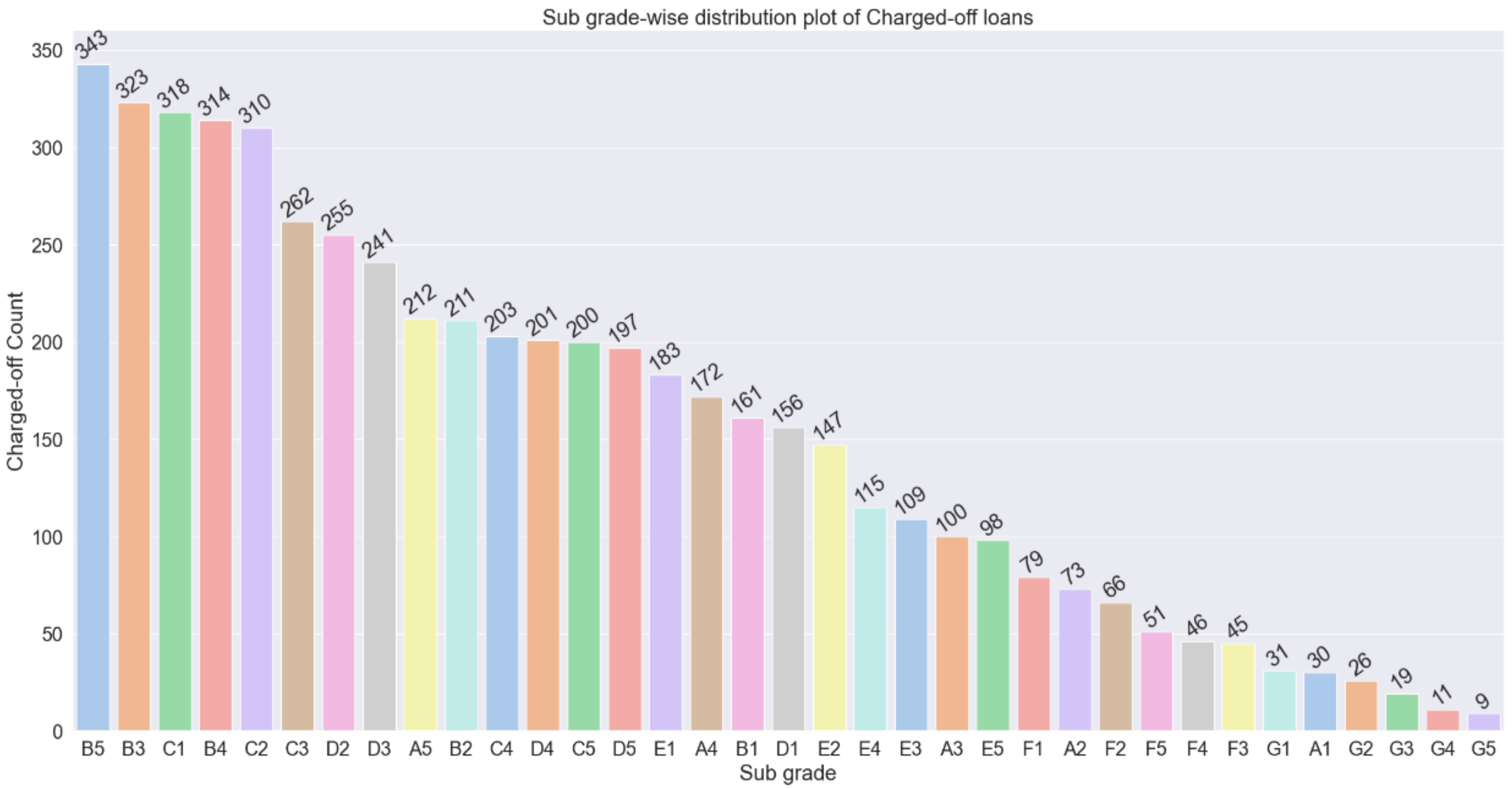


Ordered categorical Variables -Sub grade-wise

```
# Creating Seaborn categorical plot for Sub grade-wise distribution plot of Charged-off loans.

catplot=sns.catplot(x='sub_grade',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2,
                   order=df_loan_chargedoff['sub_grade'].value_counts().index)
catplot.set(title='Sub grade-wise distribution plot of Charged-off loans', xlabel='Sub grade', ylabel='Charged-off Count')
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=35)
plt.show()
```

Sub Grade-wise B3,B4 and B5 are the most defaulters under B. Also in C, we have majority of defaulters in C2 & C3. This means with Grade B & C, the lending club needs to focus on the above sub grades carefully before approving loans.

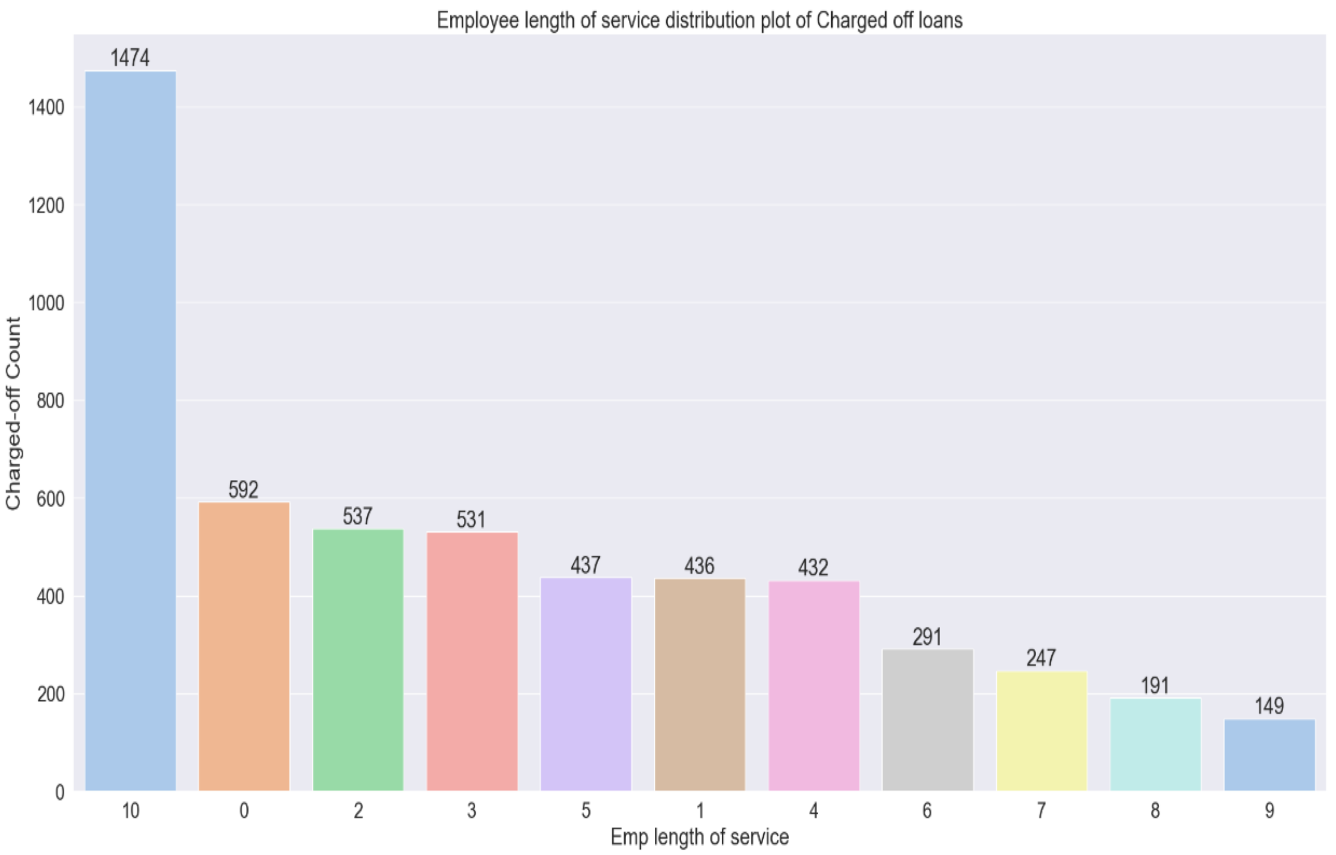


Ordered categorical Variables -Employee length of service

```
# Creating Seaborn categorical plot for Employee length of service distribution plot of Charged off loans.

catplot=sns.catplot(x='emp_length',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2,
                    order=df_loan_chargedoff['emp_length'].value_counts().index)
catplot.set(title='Employee length of service distribution plot of Charged off loans', xlabel='Emp length of service', ylabel='Charged-off Count')
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

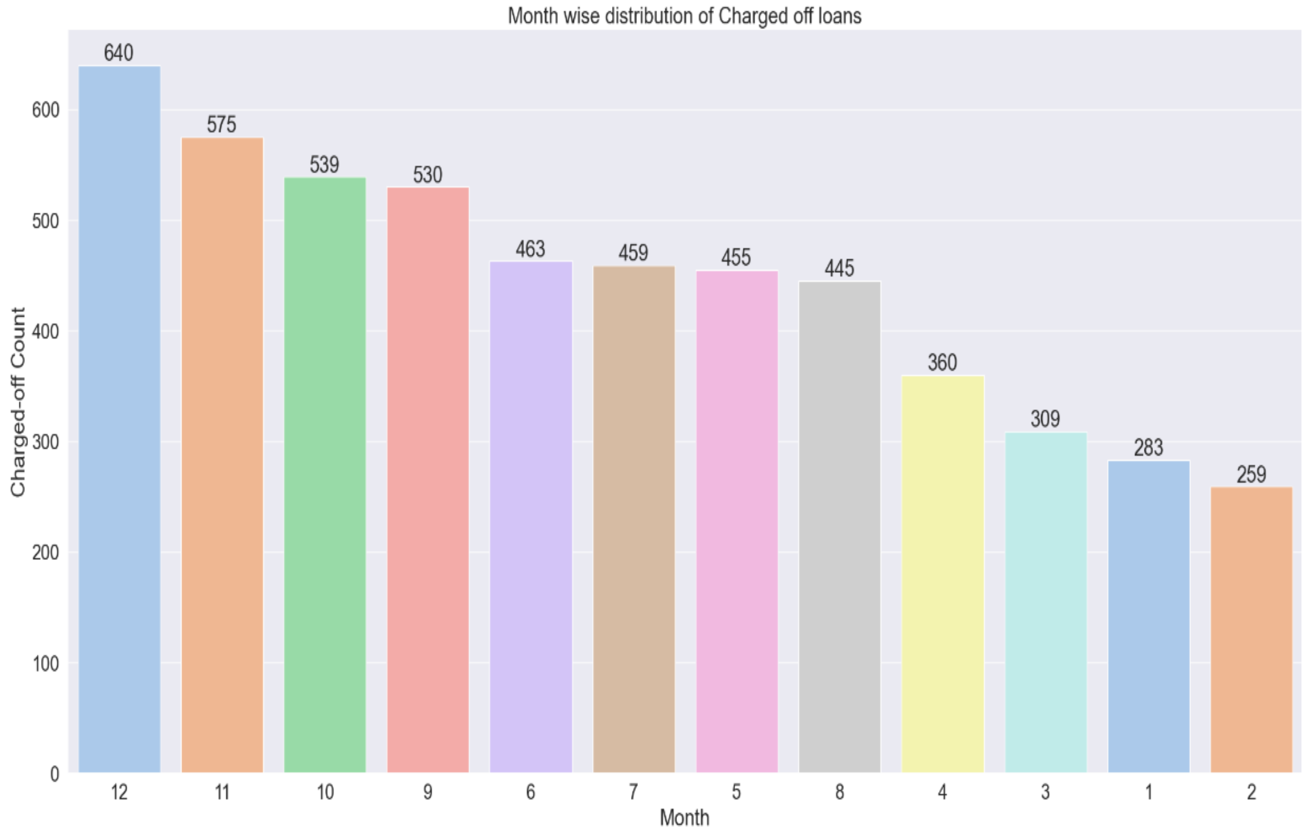
Employees with 10+ years of experiences are the majority defaulters, this figure is 1474 loans, which means the employee experience will not guarantee full payment of loans.



Ordered categorical Variables - Month wise

```
# Creating Seaborn categorical plot for Month wise distribution of Charged off loans
catplot=sns.catplot(x='mnth_issued',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2,
                   order=df_loan_chargedoff['mnth_issued'].value_counts().index)
catplot.set(title='Month wise distribution of Charged off loans', xlabel='Month', ylabel='Charged-off Count')
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

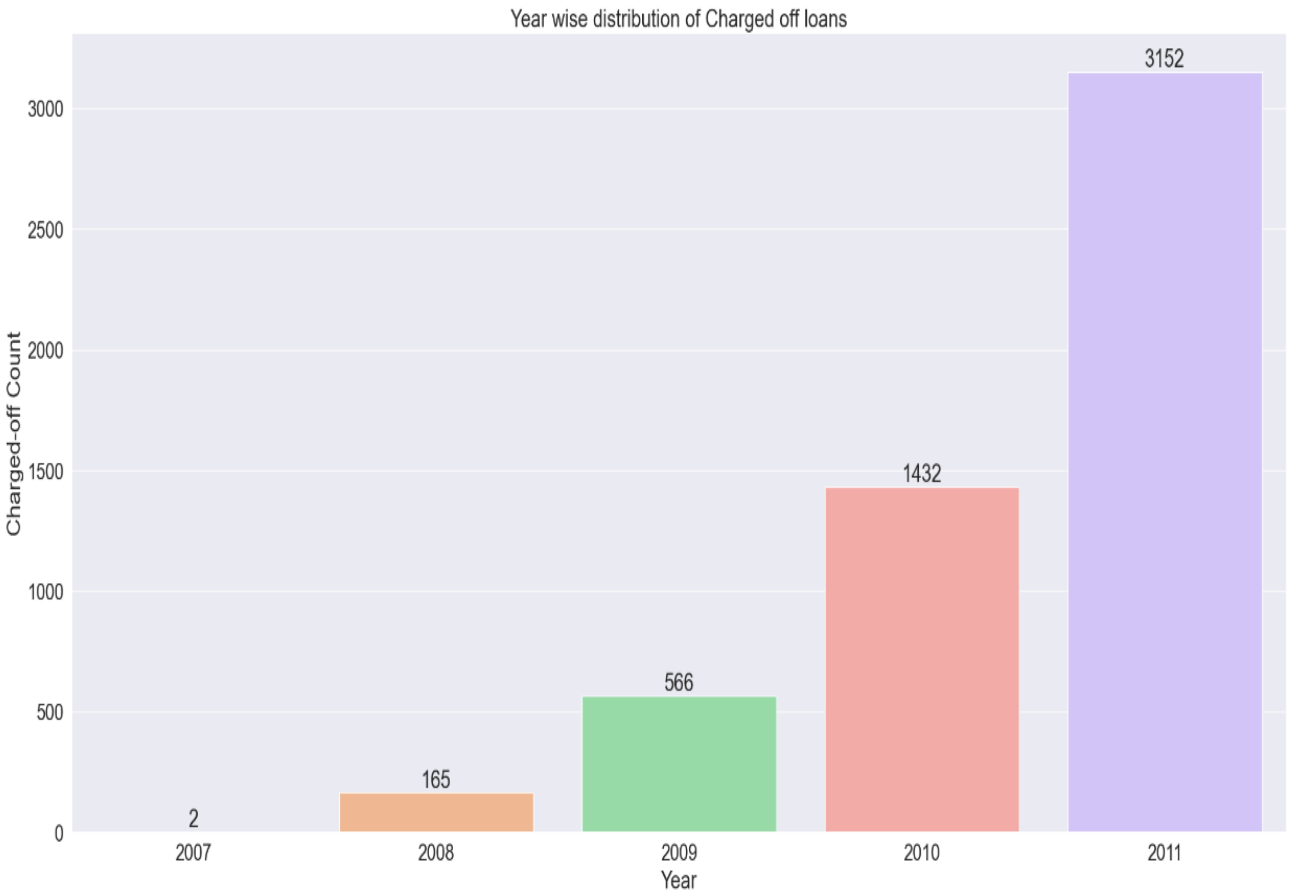
December is most of the month wherein maximum loans gets defaulted. Total of 640 loans got defaulted. This may be due to holiday season. Lending club should do proactive follow-ups during the previous month.



Ordered categorical Variables - Year wise

```
# Creating Seaborn categorical plot for Year wise distribution of Charged off Loans
catplot=sns.catplot(x='yr_issued',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0)
catplot.set(title='Year wise distribution of Charged off loans', xlabel='Year', ylabel='Charged-off Count')
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

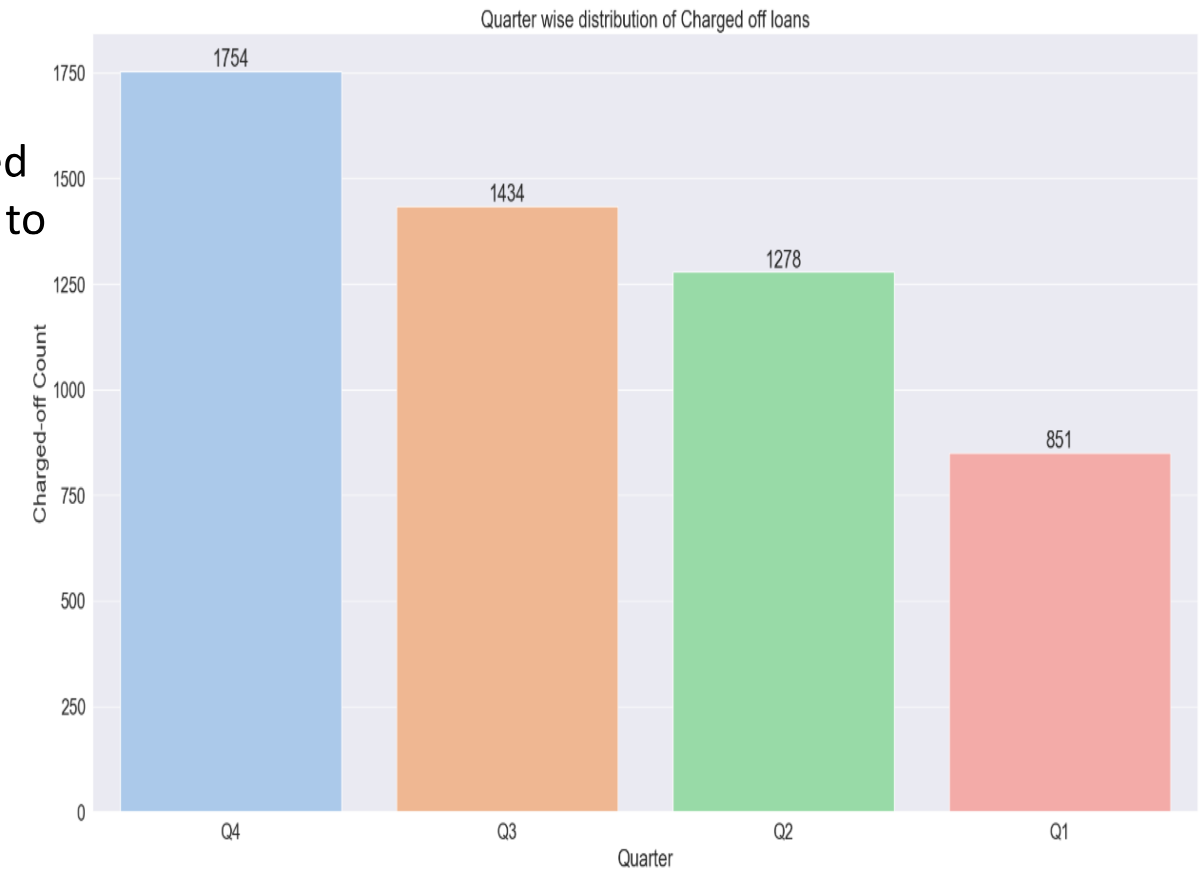
Year 2011 reported 3152 loans got defaulted. This is more than 50% of 2010.The same trend is continued from 2007 to 2010. The trend shows there is likely to be more defaulters in upcoming years.



Ordered categorical Variables - Quarter wise

```
# Creating Seaborn categorical plot for Quarter wise distribution of Charged off loans
catplot=sns.catplot(x='quarter',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0)
catplot.set(title='Quarter wise distribution of Charged off loans', xlabel='Quarter', ylabel='Charged-off Count')
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

Loans are getting defaulted in the last quarter (Q4). Reported that 1754 loans gets defaulted. This may have occurred due to holiday season.



Ordered categorical Variables

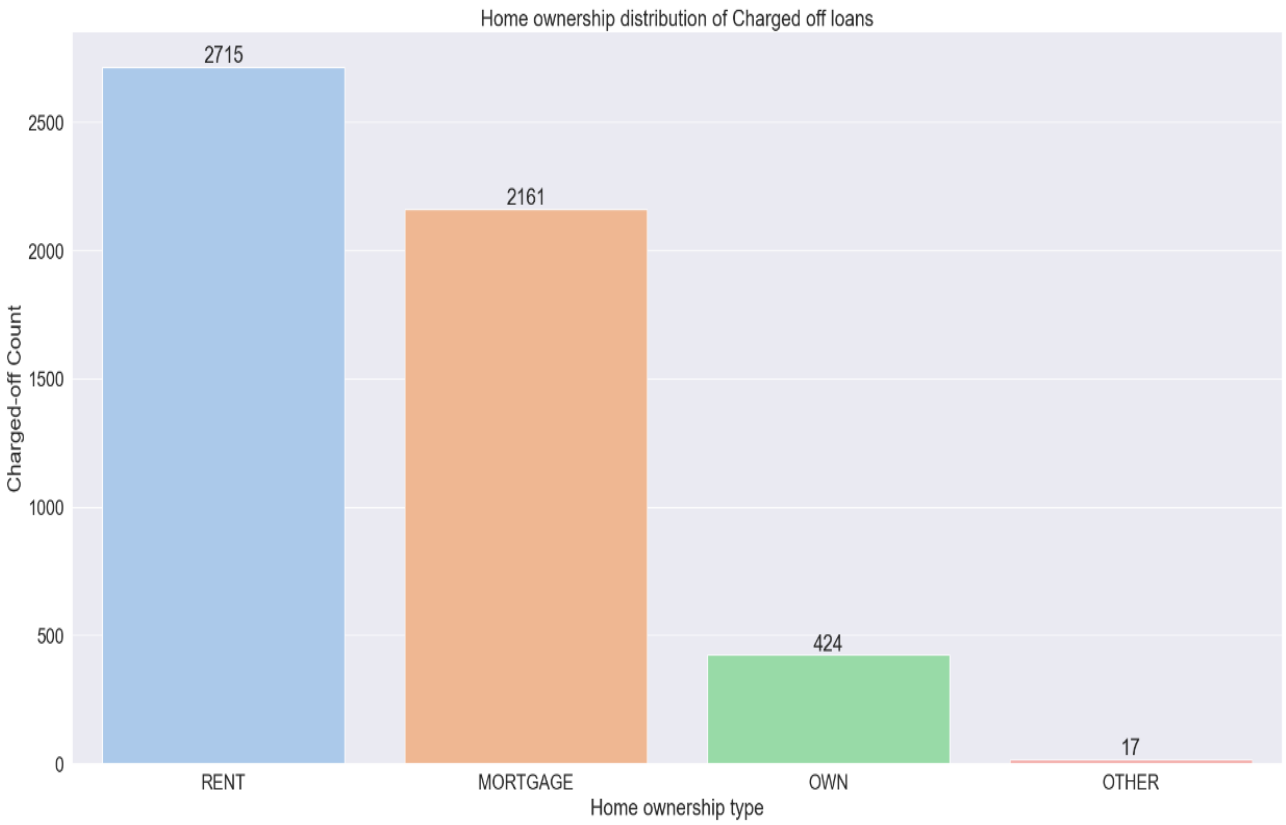
Observations and Inferences from Ordered categorical Variables analysis

- Loans with a 36-month term are most likely to default. 3006 Loans under 36 months got defaulted. This indicates that individuals default on short-term loans.
- Grade B had the highest number of Charged off loan applicants, with a total of 1,352 applicants. The next is C with 1293 loan applicants. This indicates the lending club should pay attention for providing loan to B & C Grade applicants.
- Sub Grade-wise B3,B4 and B5 are the most defaulters under B. Also in C, we have majority of defaulters in C2 & C3. This means with Grade B & C, the lending club needs to focus on the above sub grades carefully before approving loans.
- Employees with 10+ years of experiences are the majority defaulters, this figure is 1474 loans, which means the employee experience will not guarantee full payment of loans.
- December is most of the month wherein maximum loans gets defaulted. Total of 640 loans got defaulted. This may be due to holiday season. Lending club should do proactive follow-ups during the previous month.
- Year 2011 reported 3152 loans got defaulted. This is more than 50% of 2010.The same trend is continued from 2007 to 2010. The trend shows there is likely to be more defaulters in upcoming years.
- Loans are getting defaulted in the last quarter (Q4). Reported that 1754 loans gets defaulted. This may have occurred due to holiday season.

Un-Ordered categorical Variables - Home ownership

```
# Creating Seaborn categorical plot for Home ownership distribution of Charged off loans
catplot=sns.catplot(x='home_ownership',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0,
                   order =df_loan_chargedoff['home_ownership'].value_counts().index)
catplot.set(title='Home ownership distribution of Charged off loans', xlabel='Home ownership type', ylabel='Charged-off Count')
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

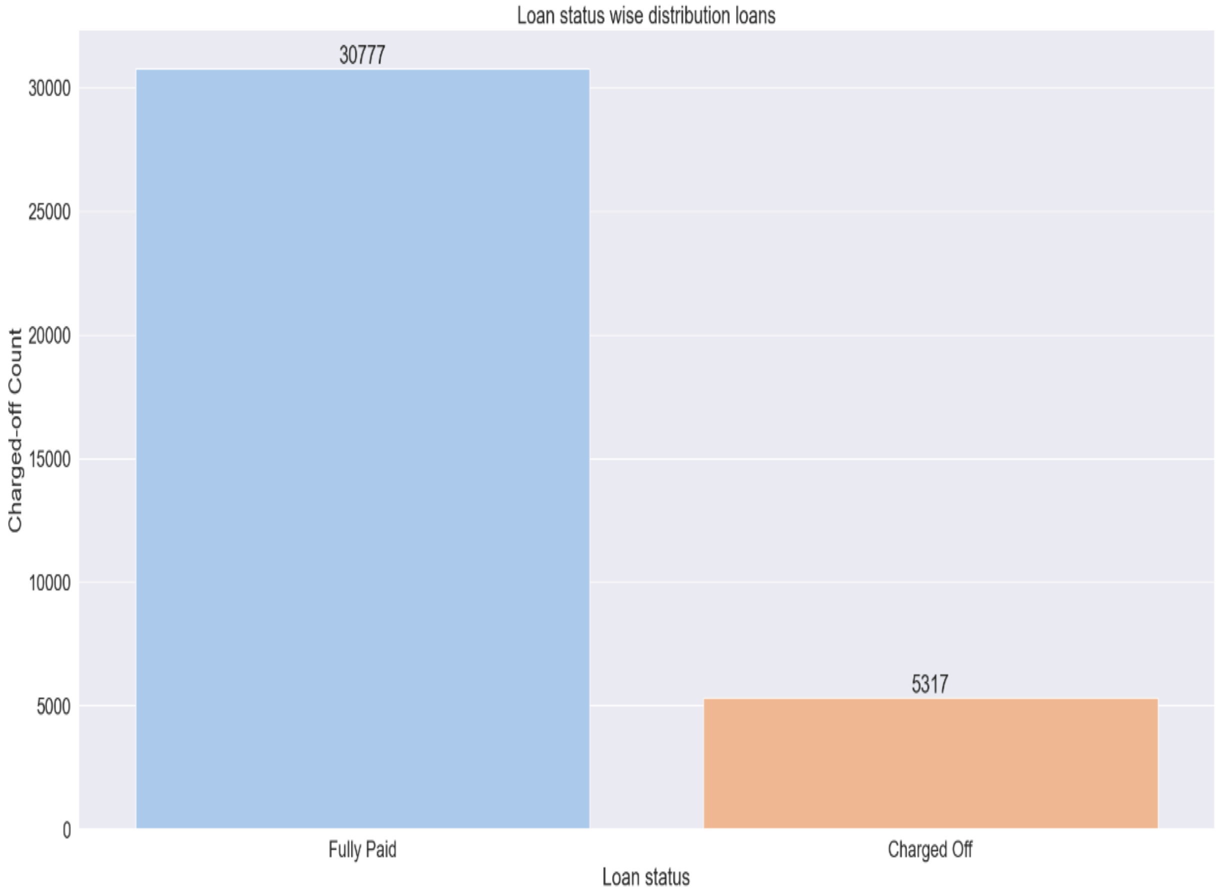
Rented accommodation has the highest defaulters, which is 2715 loans. Lending club should take necessary caution on approving loans for those who are living in rented house as this can likely get defaulted.



Un-Ordered categorical Variables - Loan status

```
# Creating Seaborn categorical plot for Loan status wise distribution of loans
catplot=sns.catplot(x='loan_status',data=df_loan ,kind='count',palette='pastel',height=11, aspect=2.0)
catplot.set(title='Loan status wise distribution loans', xlabel='Loan status', ylabel='Charged-off Count')
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

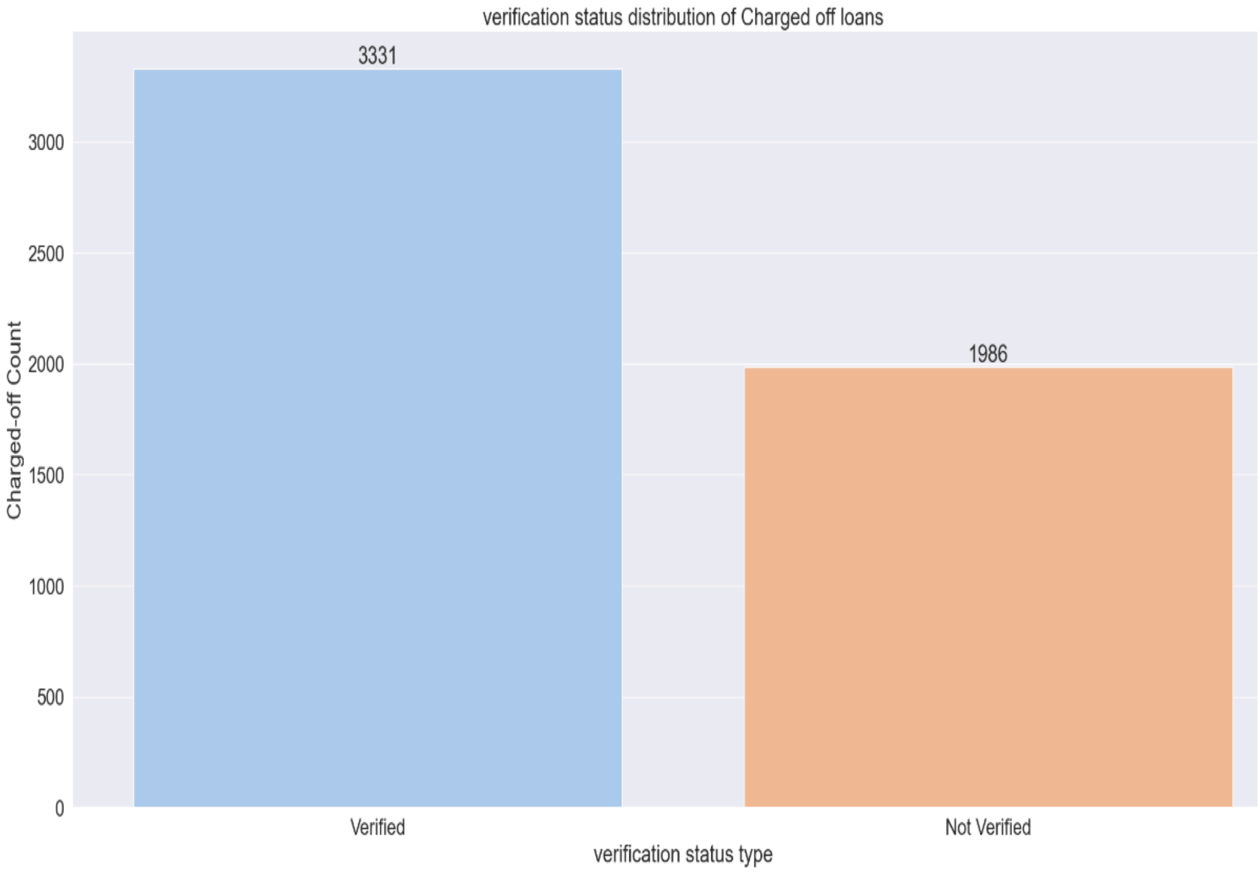
Comparatively majority of the loans are fully paid, which gives a positive indication about the process in place for lending club.



Un-Ordered categorical Variables - verification status

```
# Creating Seaborn categorical plot for verification status distribution of Charged off loans'
catplot=sns.catplot(x='verification_status',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0)
catplot.set(title='verification status distribution of Charged off loans', xlabel='verification status type', ylabel='Charged-off Count')
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

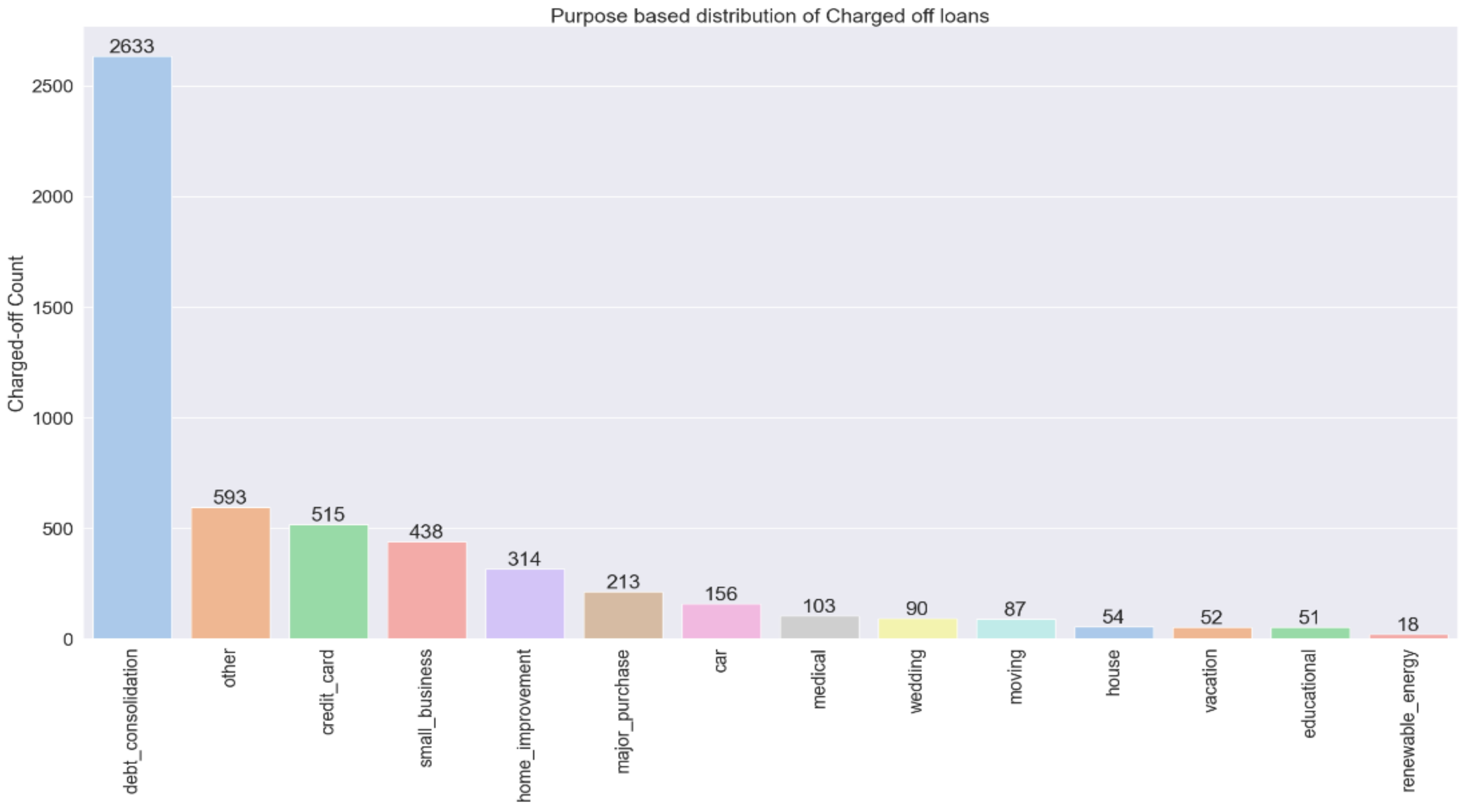
3331 Loans are defaulted which is also verified. This indicates lending club needs to revisit verification process and identify and fix gaps wherever required.



Un-Ordered categorical Variables - Purpose

```
# Creating Seaborn categorical plot for Purpose based distribution of Charged off loans
catplot=sns.catplot(x='purpose',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0 ,
                    order =df_loan_chargedoff['purpose'].value_counts().index)
catplot.set(title='Purpose based distribution of Charged off loans', xlabel='purpose', ylabel='Charged-off Count')
catplot.set_xticklabels(rotation=90)
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

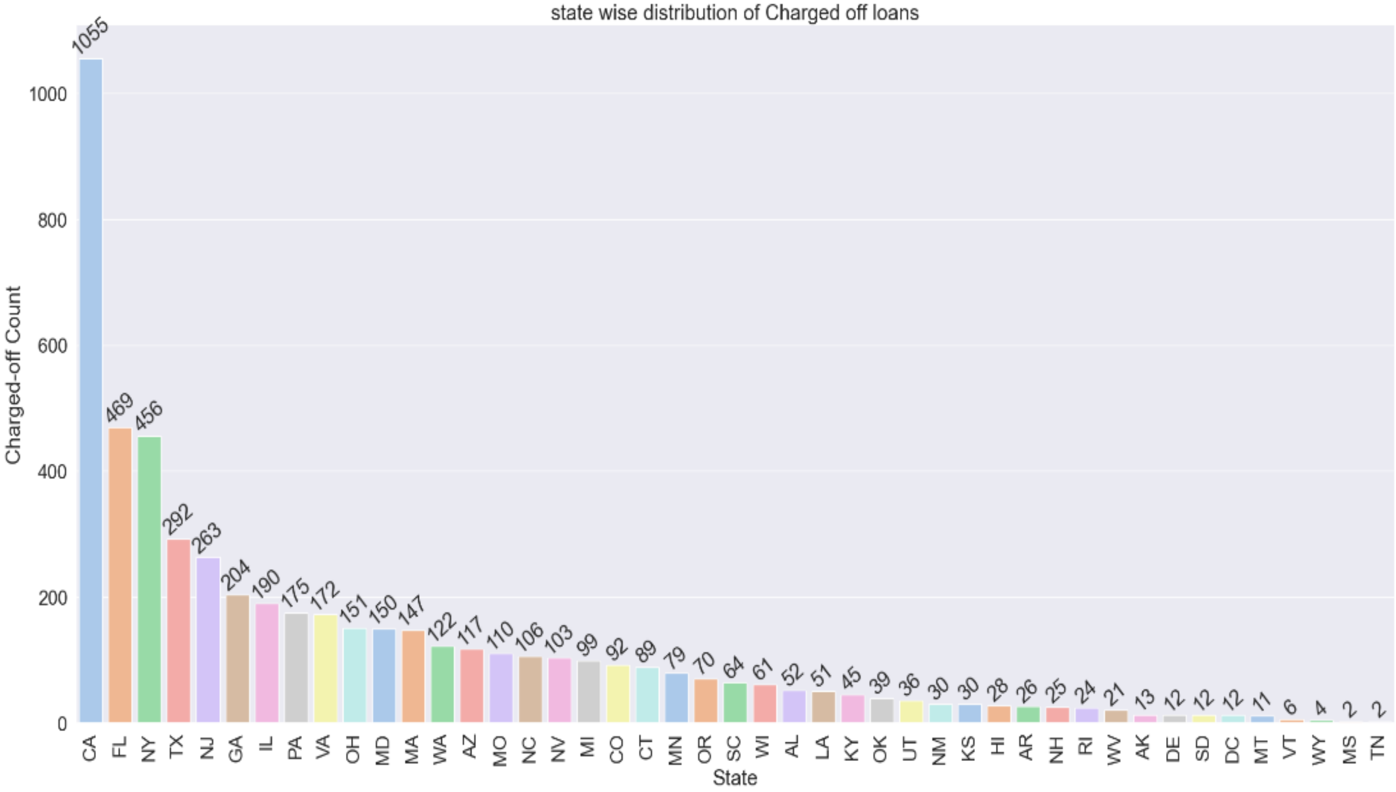
•Majority of loan got defaulted due to Debt consolidation. 2633 loans got defaulted. This indicates that when the loan is granted for this purpose the lending club has to ensure that the applicant has the capability to pay the loan as well.



Un-Ordered categorical Variables - State

```
catplot=sns.catplot(x='addr_state',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0,
                    order =df_loan_chargedoff['addr_state'].value_counts().index)
catplot.set(title='state wise distribution of Charged off loans', xlabel='State', ylabel='Charged-off Count')
catplot.set_xticklabels(rotation=90)
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=40)
plt.show()
```

•The state of CA contributes to the majority of defaulters. Lending club should be more cautious when we approve loan in this State.



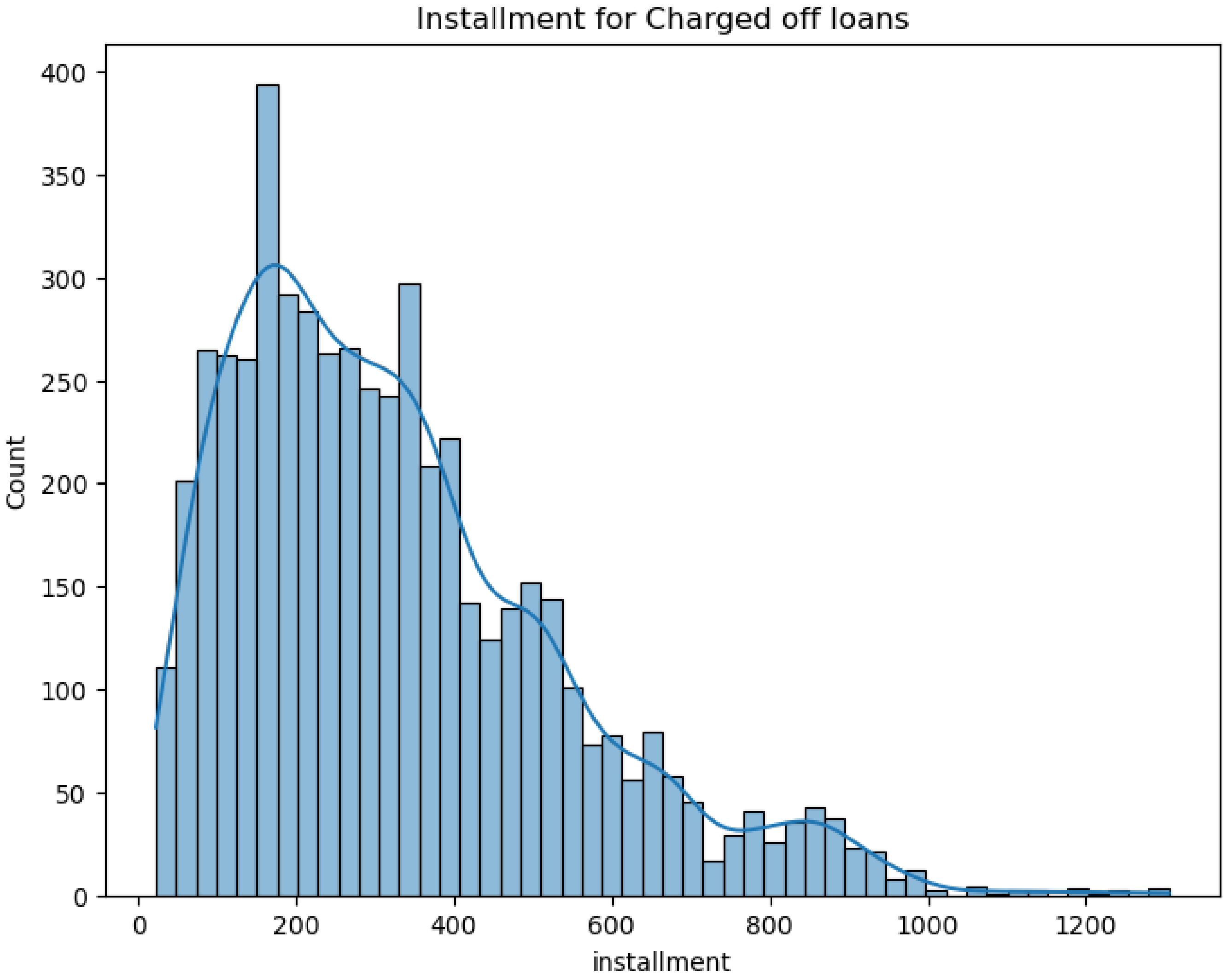
Un-Ordered categorical Variables

- **Observations and Inferences from Un-Ordered categorical Variables analysis**
  - Rented accommodation has the highest defaulters, which is 2715 loans. Lending club should take necessary caution on approving loans for those who are living in rented house as this can likely get defaulted.
  - Comparatively majority of the loans are fully paid, which gives a positive indication about the process in place for lending club.
  - 3331 Loans are defaulted which is also verified. This indicates lending club needs to revisit verification process and identify and fix gaps wherever required.
  - Majority of loan got defaulted due to Debt consolidation. 2633 loans got defaulted. This indicates that when the loan is granted for this purpose the lending club has to ensure that the applicant has the capability to pay the loan as well.
  - The state of CA contributes to the majority of defaulters. Lending club should be more cautious when we approve loan in this State.

Quantitative variables -Installment

```
# Plotting installments
plt.figure(figsize=(8,6))
sns.histplot(df_loan_chargedoff,x='installment',bins=50,kde=True)
plt.title('Installment for Charged off loans')
plt.show()
```

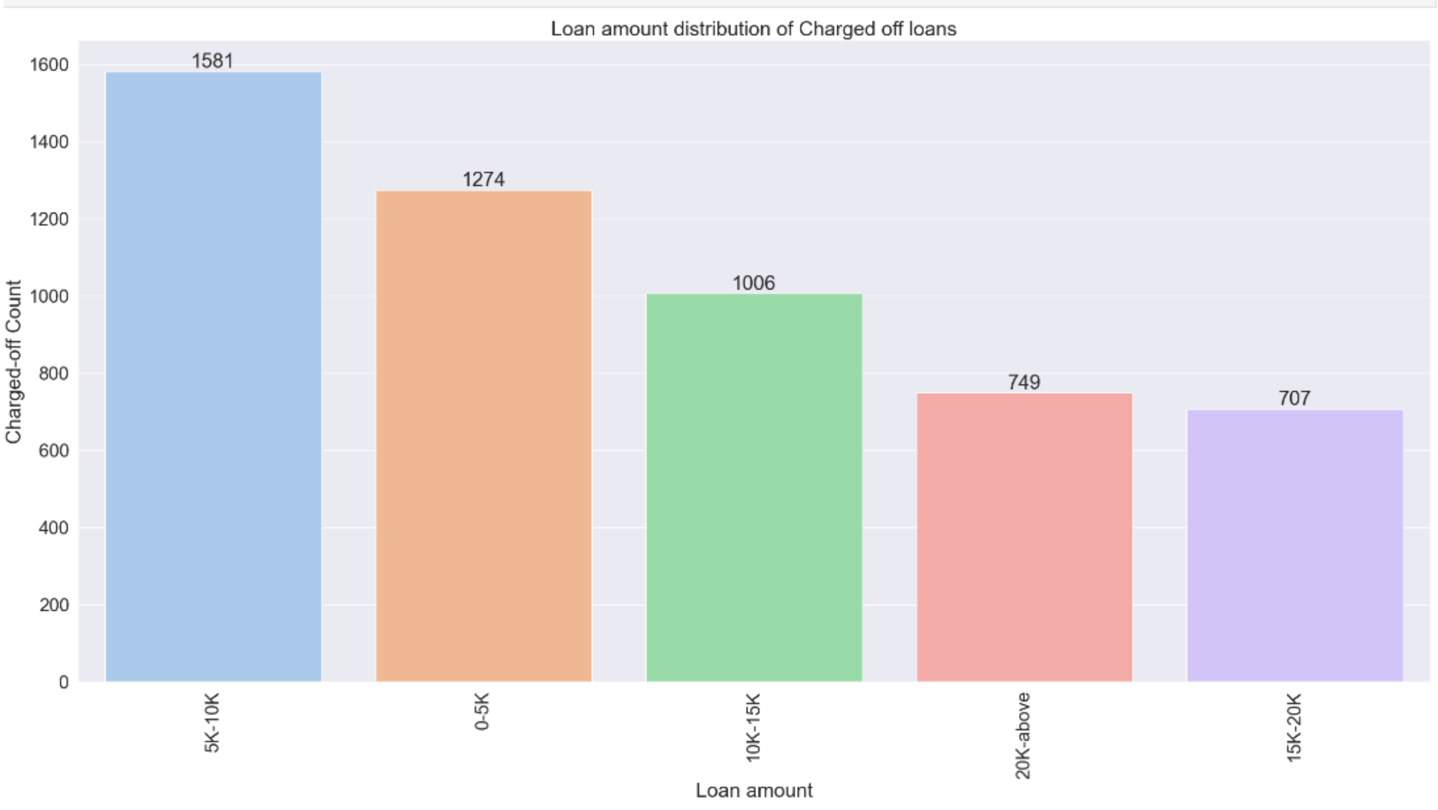
Majority of loan installment for defaulters lies between 160-400. This gives an indication for Lending company to monitor closely on these applicants with similar installments



Quantitative variables –Loan amount

```
# Plotting Loan amount from the derived column Loanamt_bucket of defaulters
catplot=sns.catplot(x='Loanamt_bucket',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0,
                    order =df_loan_chargedoff['Loanamt_bucket'].value_counts().index)
catplot.set(title='Loan amount distribution of Charged off loans', xlabel='Loan amount', ylabel='Charged-off Count')
catplot.set_xticklabels(rotation=90)
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

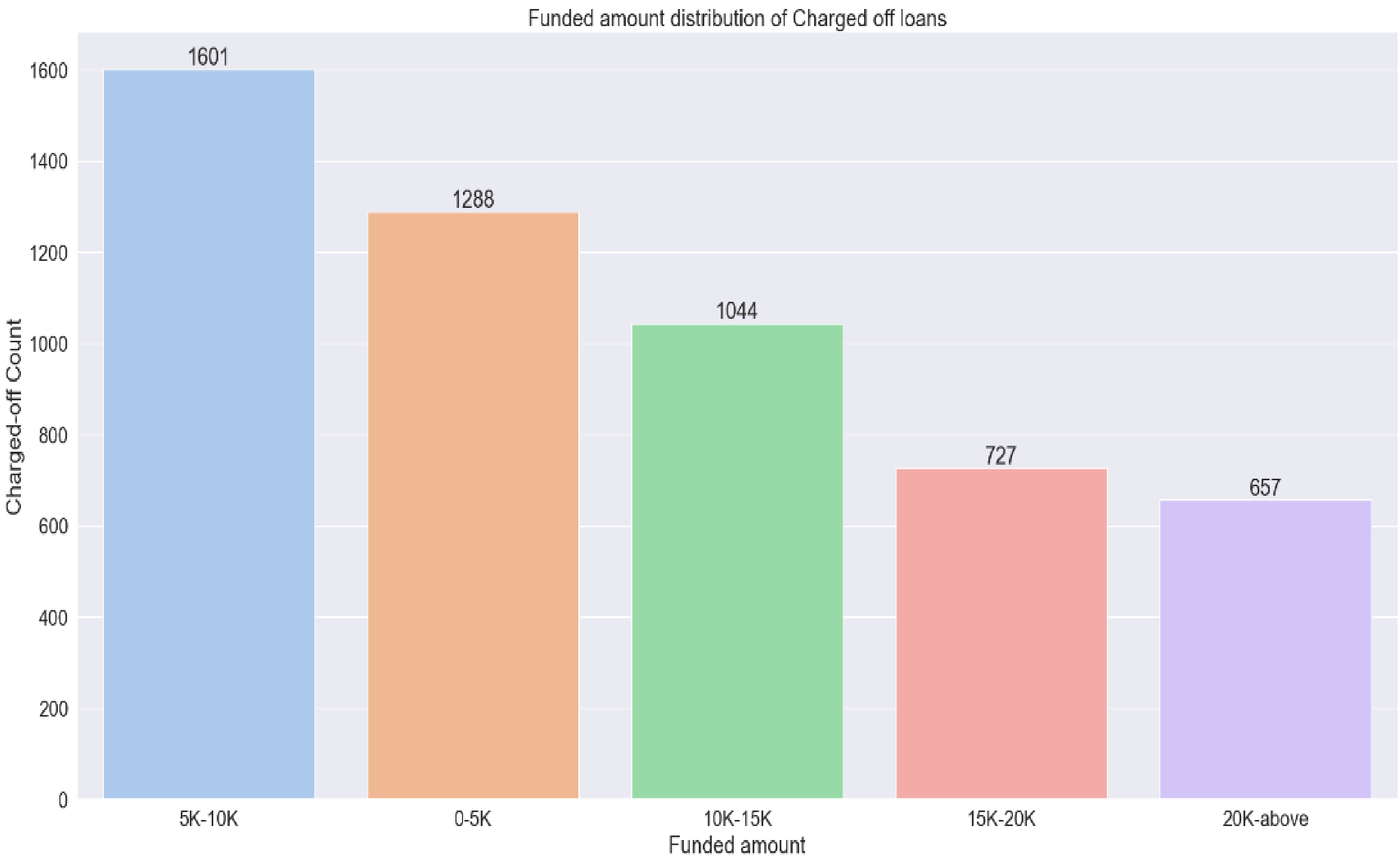
Most defaulters are those with loan amount between 5K to 10K . The lending company should evaluate applicants who apply for bigger amount. Should verify repayment capability as well.



Quantitative variables – Funded amount

```
catplot=sns.catplot(x='funded_amnt_bucket',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0,
                    order =df_loan_chargedoff['funded_amnt_bucket'].value_counts().index)
catplot.set(title='Funded amount distribution of Charged off loans', xlabel='Funded amount', ylabel='Charged-off Count')
catplot.set_xticklabels(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

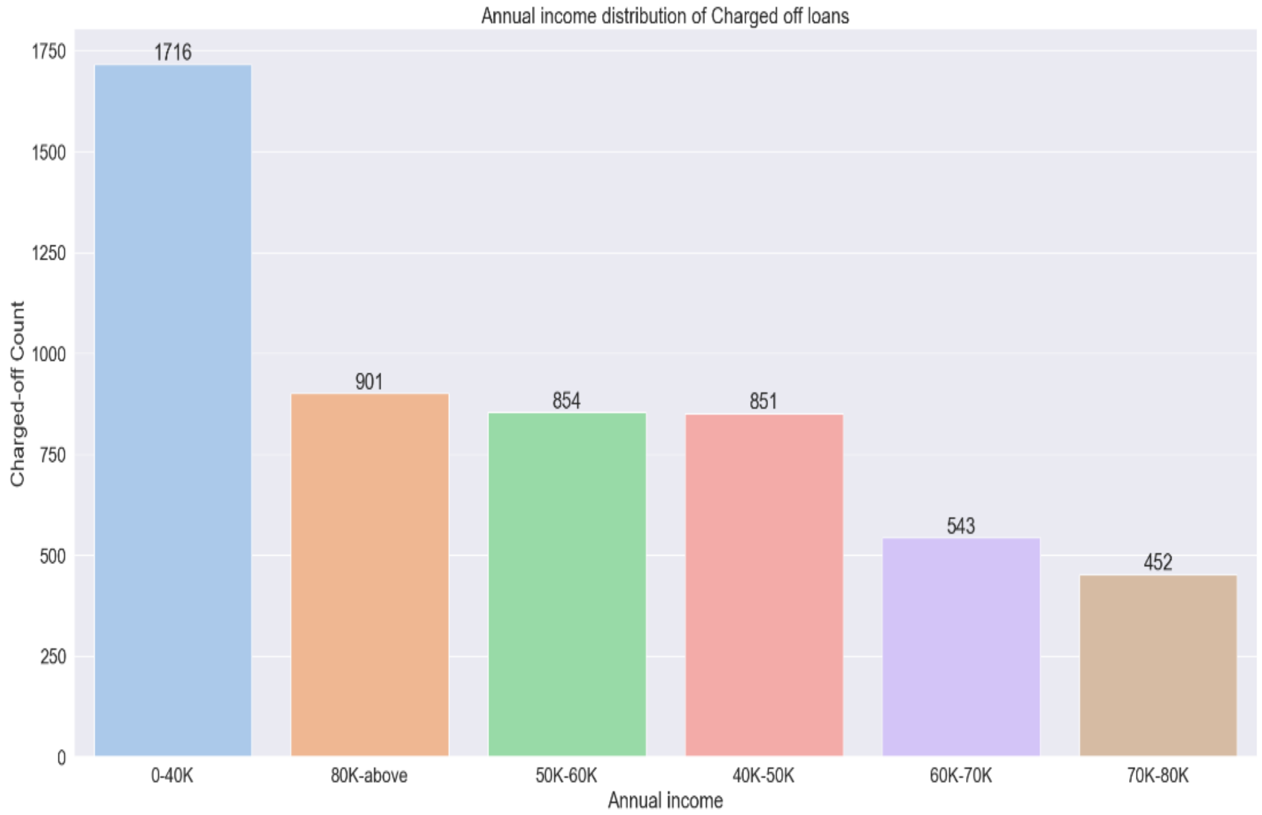
1601 Loan applicants who received funds between 5K to 10K are most of the defaulters. The lending company should be cautious on providing large funds to applicants. Lending company should verify the repayment capability before approval.



Quantitative variables – Annual income

```
catplot=sns.catplot(x='annual_inc_bucket',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0,
                    order =df_loan_chargedoff['annual_inc_bucket'].value_counts().index)
catplot.set(title='Annual income distribution of Charged off loans', xlabel='Annual income', ylabel='Charged-off Count')
catplot.set_xticklabels(rotation=0)
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

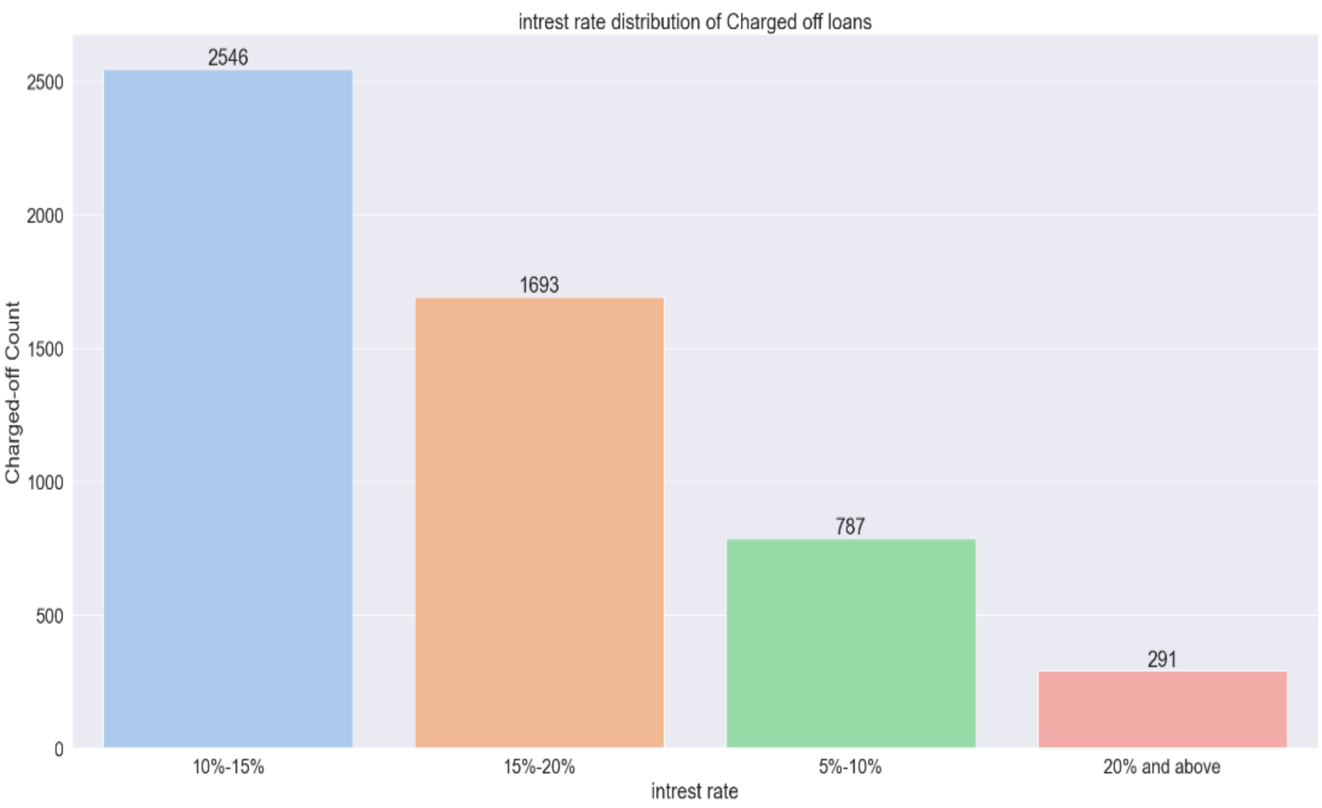
•1716 Loan applicants with salary range from 0-40K are most of the defaulters. Lending company should check the repayment capability of employee who is in the lower salary band.



Quantitative variables – Interest rate

```
catplot=sns.catplot(x='int_rate_bucket',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0,
                    order =df_loan_chargedoff['int_rate_bucket'].value_counts().index)
catplot.set(title='interest rate distribution of Charged off loans', xlabel='interest rate', ylabel='Charged-off Count')
catplot.set_xticklabels(rotation=0)
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

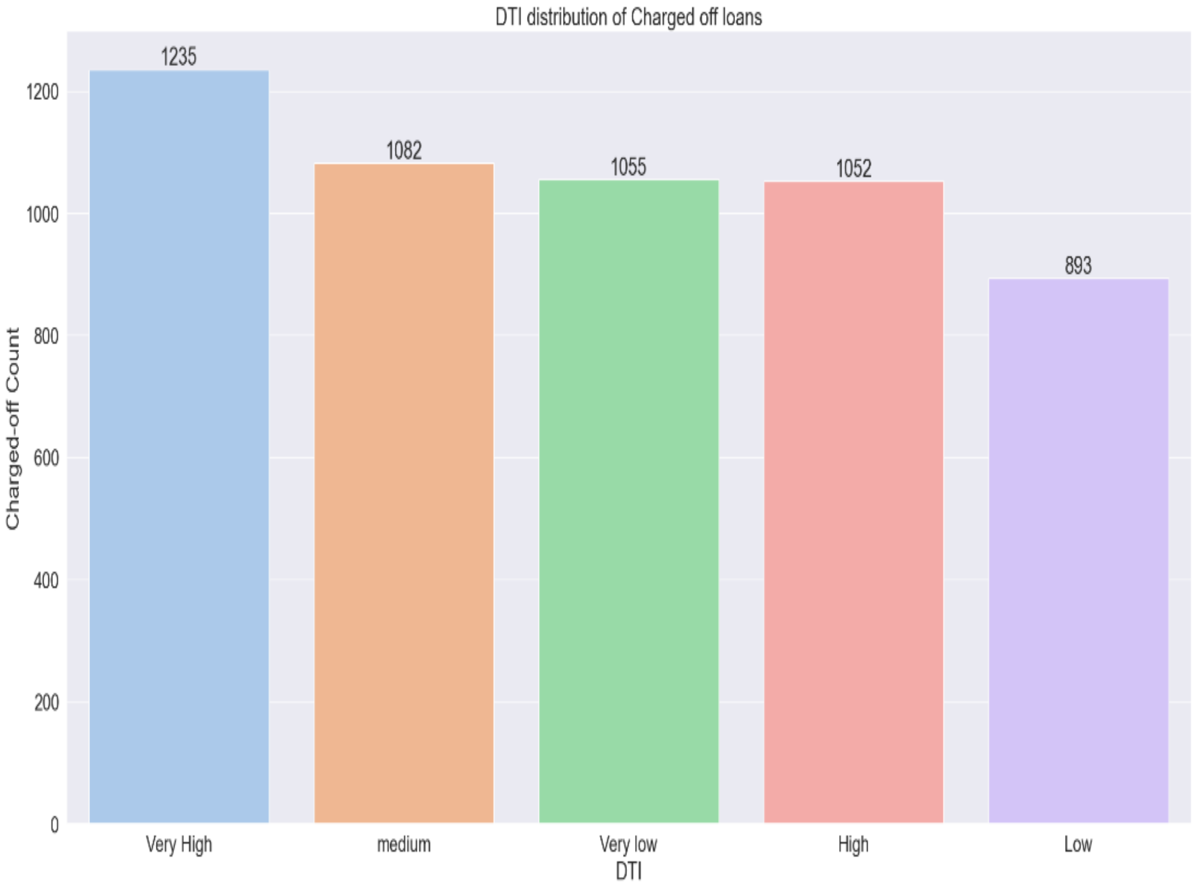
2546 Loan applicants having 10-15% interest rate contribute to the majority of charged off loans. A lower interest rate could possibly mitigate the risk of credit loss.



Quantitative variables – DTI

```
catplot=sns.catplot(x='dti_bucket',data=df_loan_chargedoff ,kind='count',palette='pastel',height=11, aspect=2.0,
                    order =df_loan_chargedoff['dti_bucket'].value_counts().index)
catplot.set(title='DTI distribution of Charged off loans', xlabel='DTI', ylabel='Charged-off Count')
catplot.set_xticklabels(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

•1235 Applicants have very high Debt-to-income ratio. The lending company should put strict debt to income ratio requirements in order to have sustainable levels for the debt against income.



Quantitative variables

- **Observations and Inferences from Quantitative analysis**
  - •Majority of loan installment for defaulters lies between 160-400. This gives an indication for Lending company to monitor closely on these applicants with similar installments.

  - •Most defaulters are those with loan amount beween 5K to 10K . The lending company should evaluate applicants who apply for bigger amount. Should verify repayment capability as well.

  - •1601 Loan applicants who received funds between 5K to 10K are most of the defaulters. The lending company should be cautious on providing large funds to applicants. Lending company should verify the repayment capability before approval.

  - •1716 Loan applicants with salary range from 0-40K are most of the defaulters. Lending company should check the repayment capability of employee who is in the lower salary band.

  - •2546 Loan applicants having 10-15% interest rate contribute to the majority of charged off loans. A lower interest rate could possibly mitigate the risk of credit loss.

  - •1235 Applicants have very high Debt-to-income ratio. The lending company should put strict debt to income ratio requirements in order to have sustainable levels for the debt against income.

Bivariate analysis

Ordered , Unordered & Quantitative variables will be analysed against Loan status

Ordered categorical variable
- term -- The number of payments on the loan.
- grade -- LC assigned loan grade.
- sub_grade -- LC assigned loan subgrade.
- emp_length -- Employment length in years.
- mnth_issued -- Month in which loan issued.
- Yr_issued -- Year in which loan issued.
- quarter -- in which quarter loan is issued

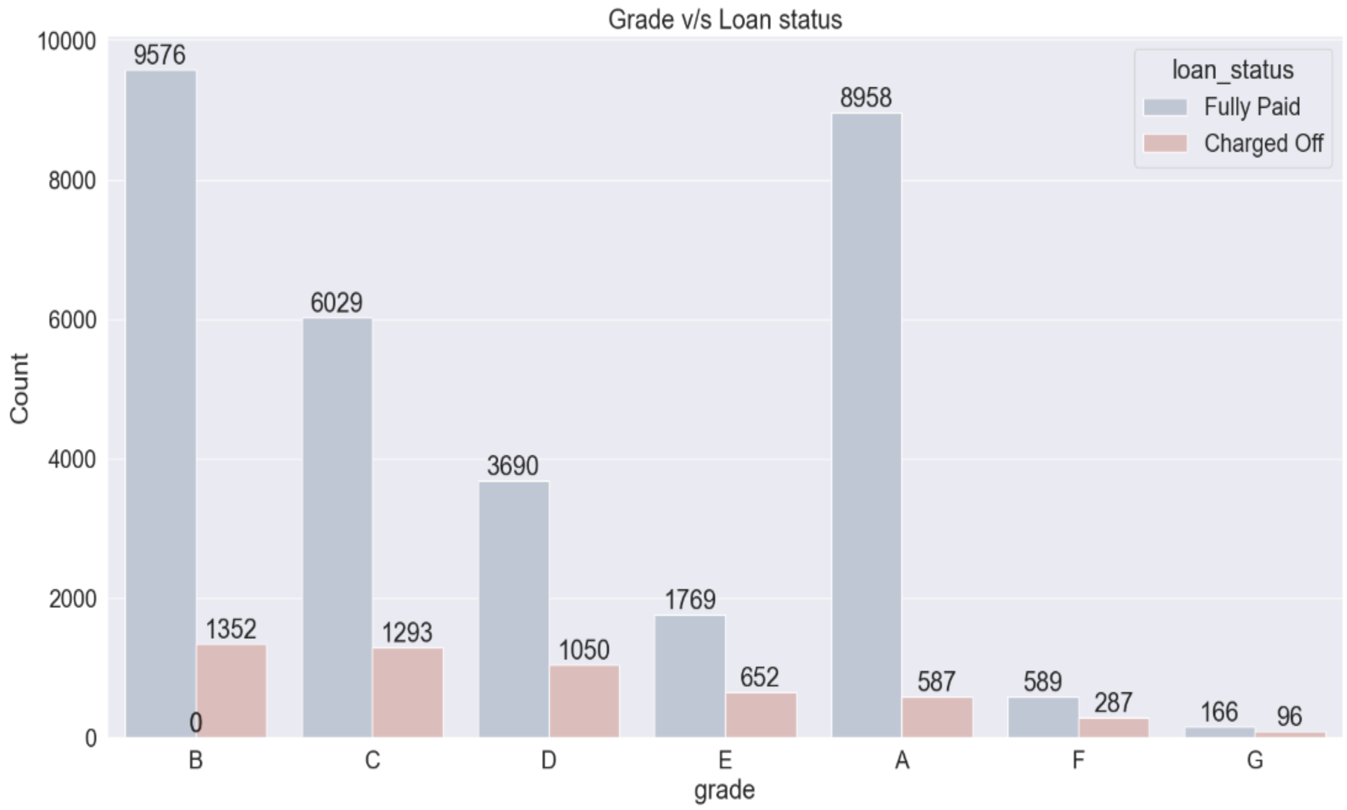
Unordered categorical variables
- Home Ownership - Whether the home is Rented / own house / under mortgage.
- Verification status - whether loan is Verified / Not verified.
- purpose - Reason for taking loan.
- State - The state to which loan applicant belongs.

Quantitative variables
- Loanamt_bucket - Loan amount in multiple bins.
- funded_amnt_bucket - Funded amount in multiple bins.
- annual_inc_bucket - Annual income in multiple bins.
- int_rate_bucket - Rate of interest in multiple bins.
- dti_bucket - Debt to income in multiple bins.

Bivariate analysis - Grade v/s Loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.countplot(x='grade',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['grade'].value_counts().index,palette='vlag')
ax.set_xlabel('grade')
ax.set_ylabel('Count')
ax.set_title('Grade v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

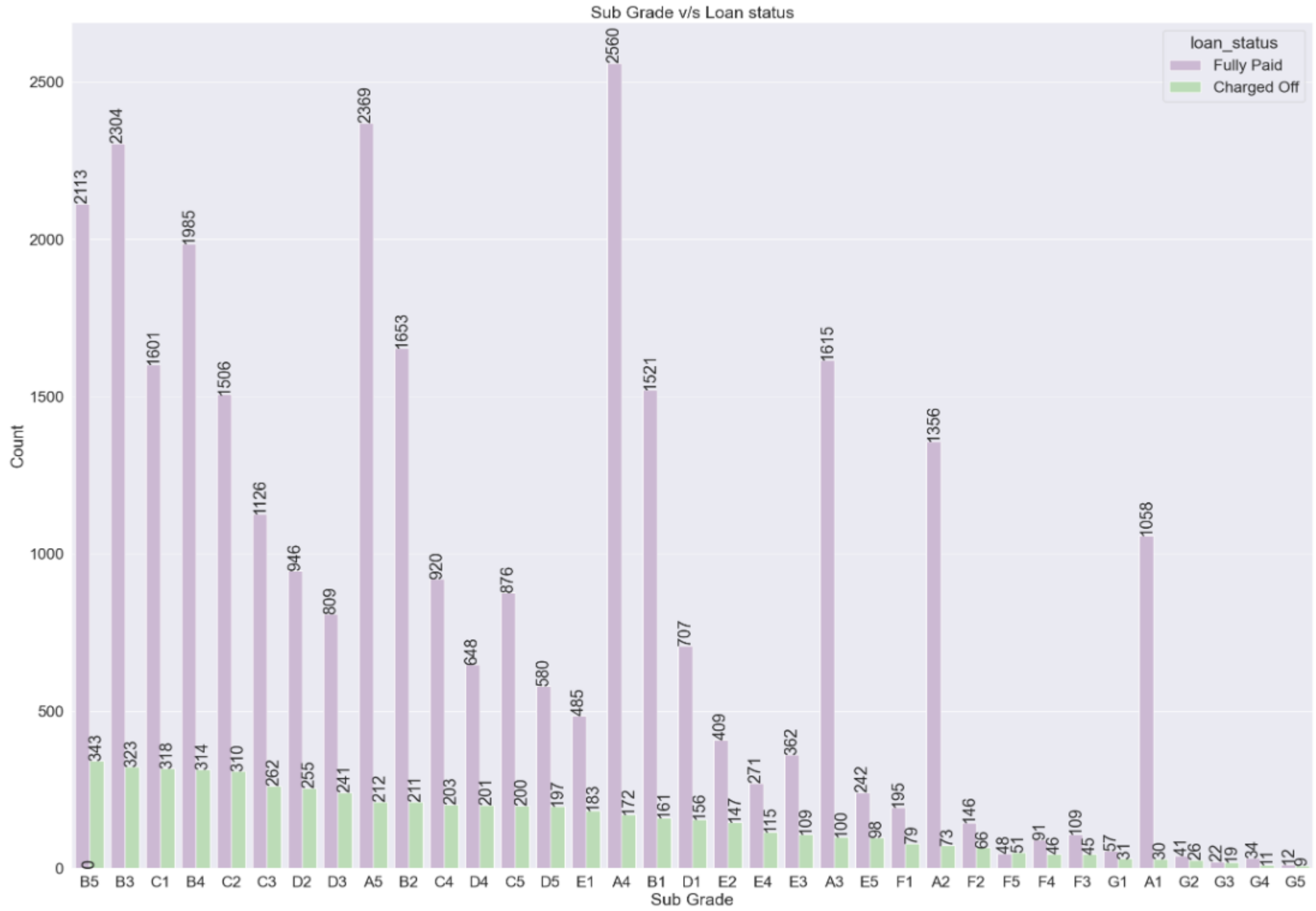
Loan applicants in Grades B, C, and D account for the majority of ‘Charged Off’ loans. However, Grades A and B make up the largest portion of total loan applicants. This indicates that Grade A applicants positively impact the business, while caution is needed when approving loans for Grade B, C and D applicants.



Bivariate analysis - subgrade v/s loan status

```
fig, ax = plt.subplots(figsize=(28,20), dpi=100)
sns.countplot(x='sub_grade',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['sub_grade'].value_counts().index,palette='PRGn')
ax.set_xlabel('Sub Grade')
ax.set_ylabel('Count')
ax.set_title('Sub Grade v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=90)
plt.show()
```

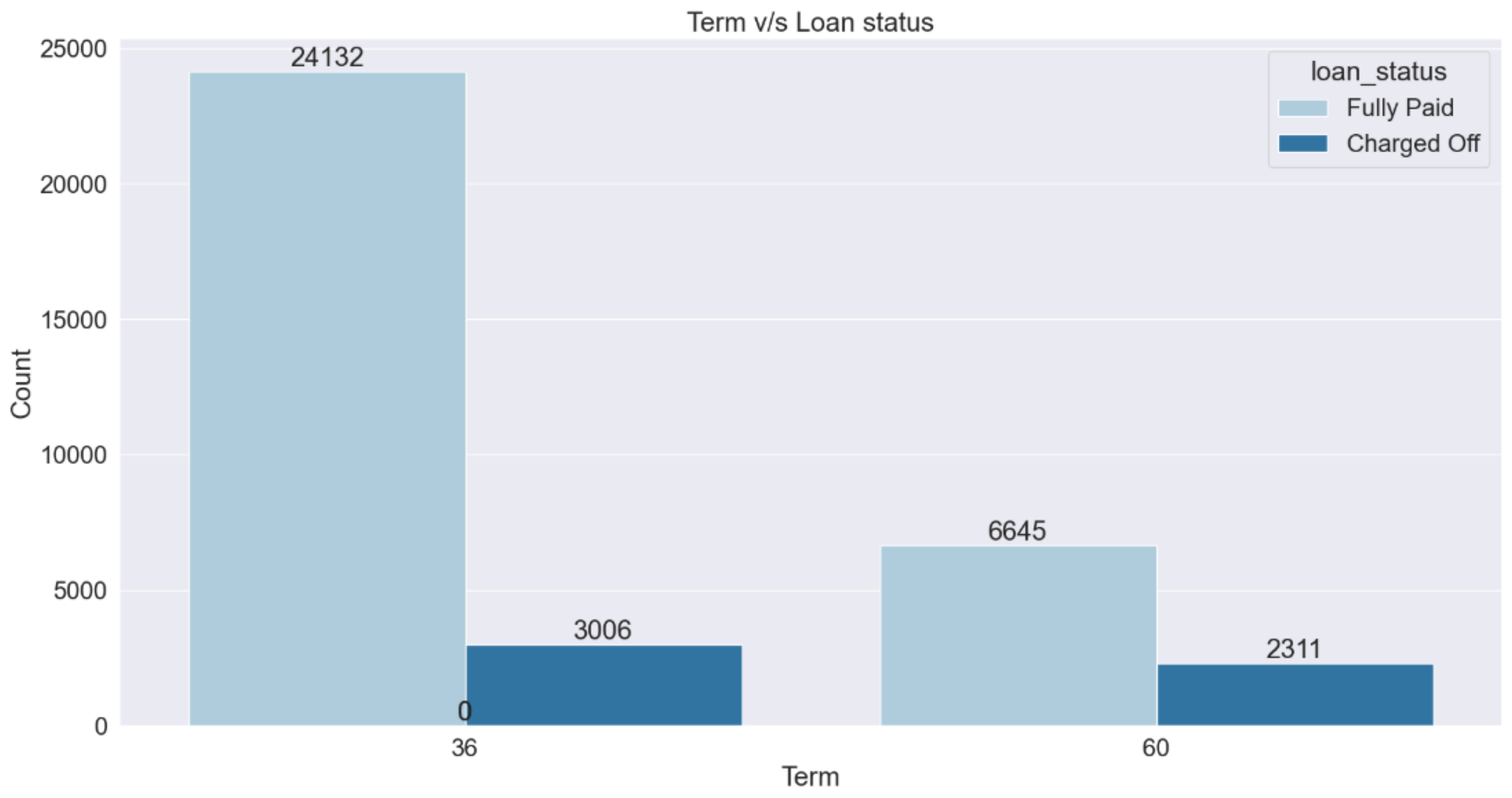
Loan subgrade B3,B4,B5,C1,C2,C3 & D2 contribute to most of Charged off loans.



Bivariate analysis - Term v/s Loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.countplot(x='term',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['term'].value_counts().index,palette='Paired')
ax.set_xlabel('Term')
ax.set_ylabel('Count')
ax.set_title('Term v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

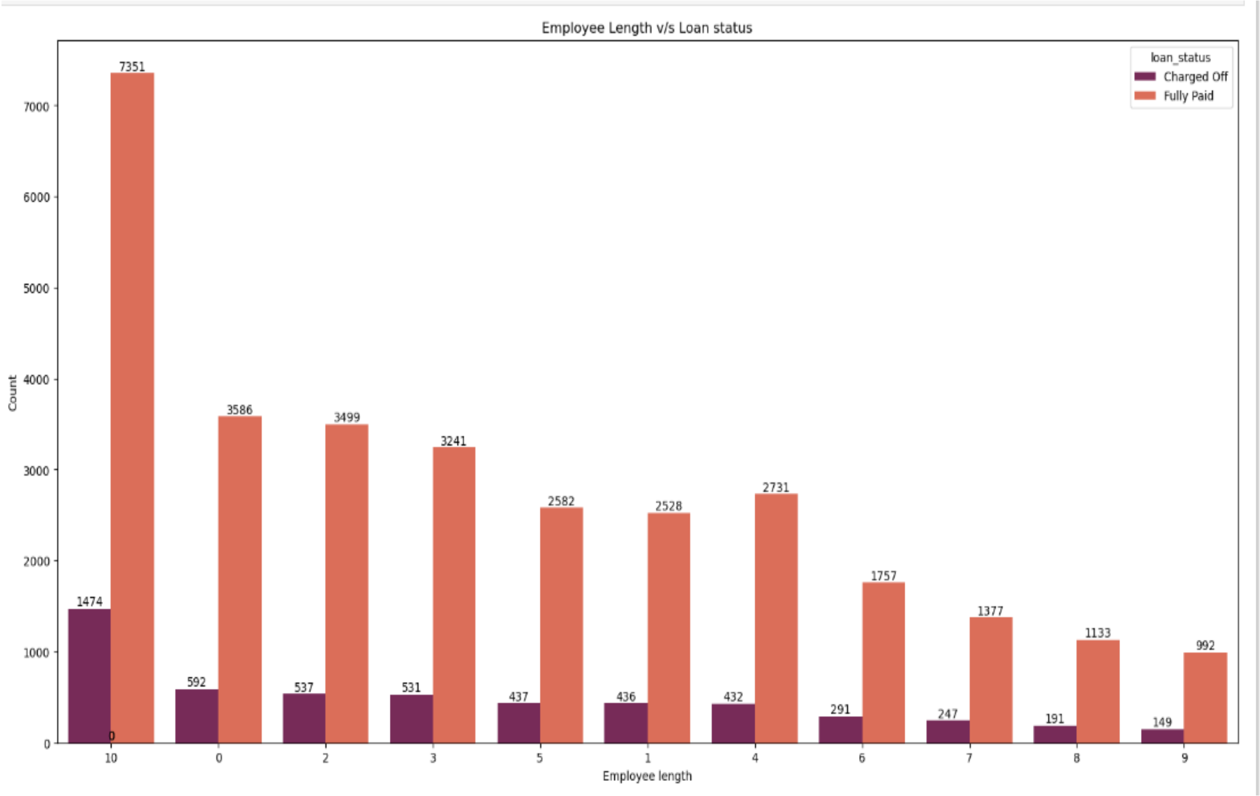
Most charged-off loans are short-term. However, when comparing the total loans for each term, the percentage of long-term loans being charged off is higher than that of short-term loans.



Bivariate analysis - Employee length v/s Loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.countplot(x='emp_length',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['emp_length'].value_counts().index,palette='rocket')
ax.set_xlabel('Employee length')
ax.set_ylabel('Count')
ax.set_title('Employee Length v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

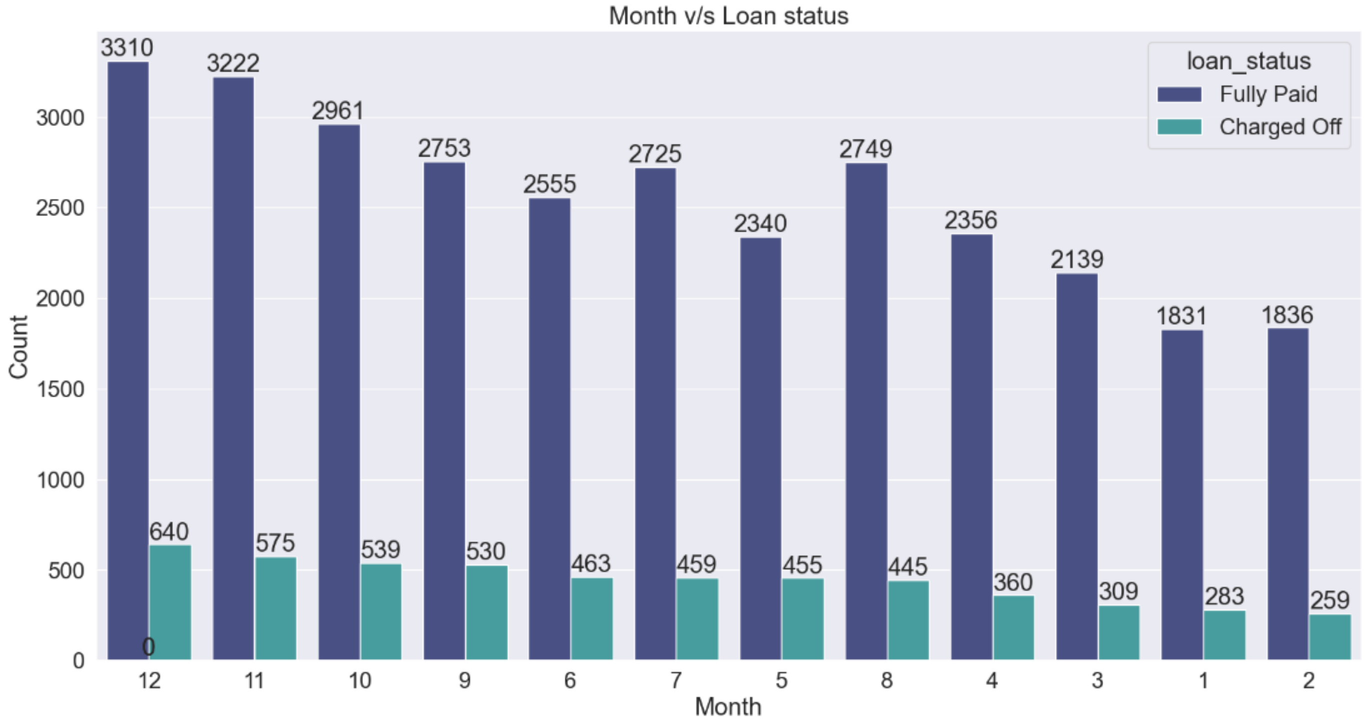
Employees having 10 or more years of experience are majority of loan applicants. They are also most likely to be defaulters.



Bivariate analysis - Month v/s loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.countplot(x='mnth_issued',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['mnth_issued'].value_counts().index,palette='mako')
ax.set_xlabel('Month')
ax.set_ylabel('Count')
ax.set_title('Month v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

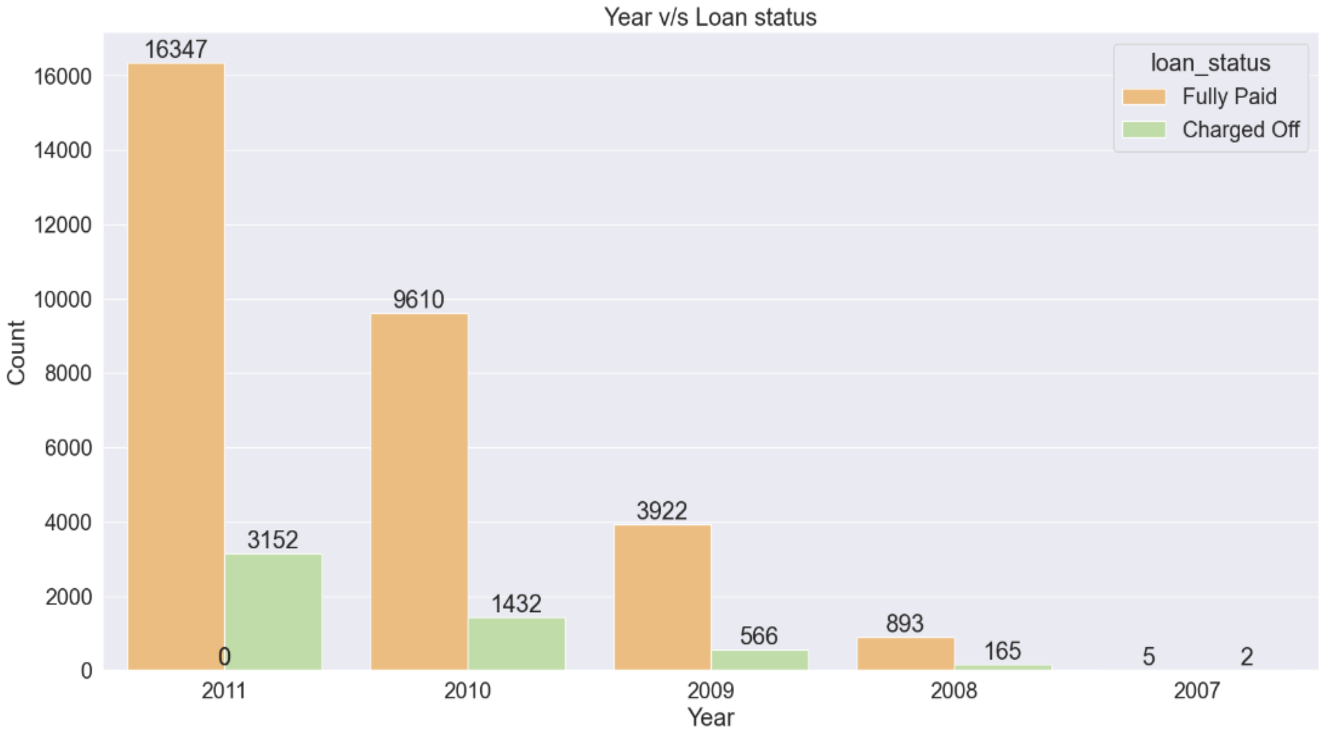
December month is the month with the maximum loan applicants.



Bivariate analysis - Year v/s Loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.countplot(x='yr_issued',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['yr_issued'].value_counts().index,palette='Spectral')
ax.set_xlabel('Year')
ax.set_ylabel('Count')
ax.set_title('Year v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

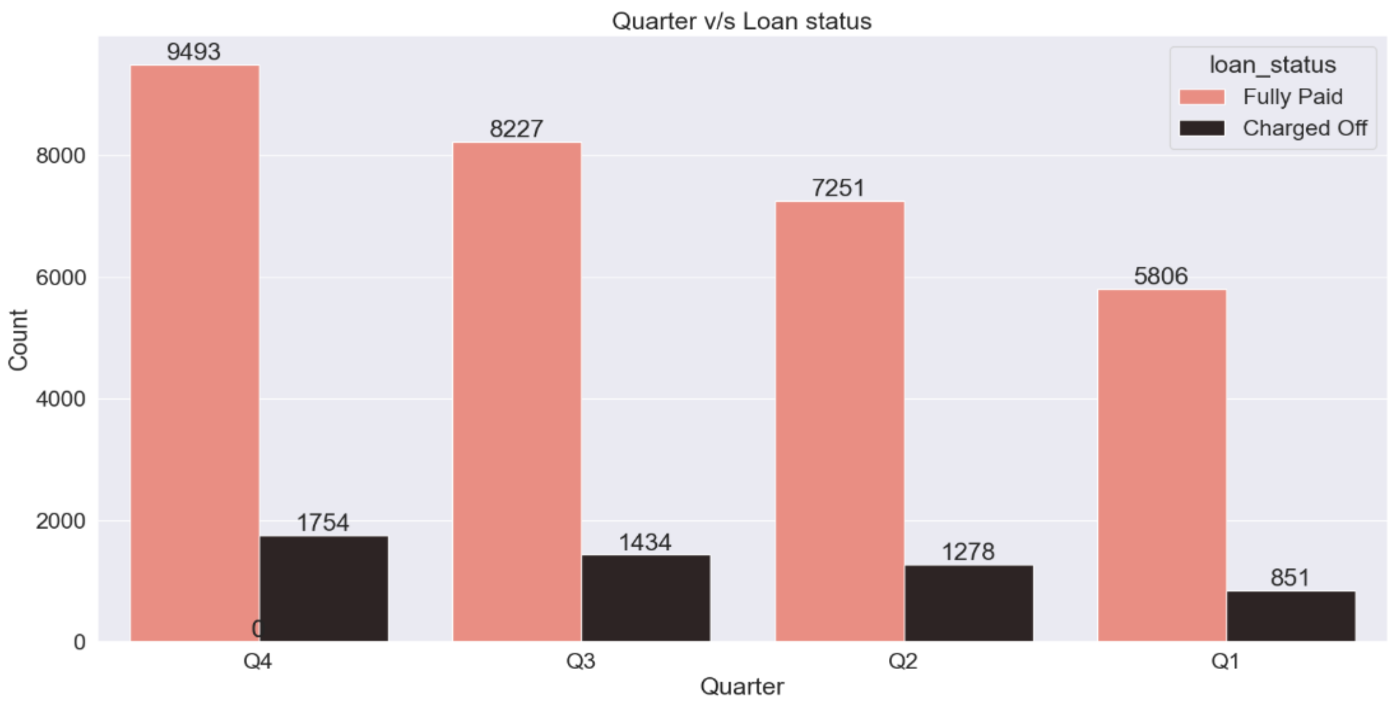
From 2007 to 2011 there is a positive trend of loan applicants. This shows there is a positive trend of more applicants in upcoming years.



Bivariate analysis - Quarter v/s loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.countplot(x='quarter',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['quarter'].value_counts().index,palette='dark:salmon_r')
ax.set_xlabel('Quarter')
ax.set_ylabel('Count')
ax.set_title('Quarter v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
     plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

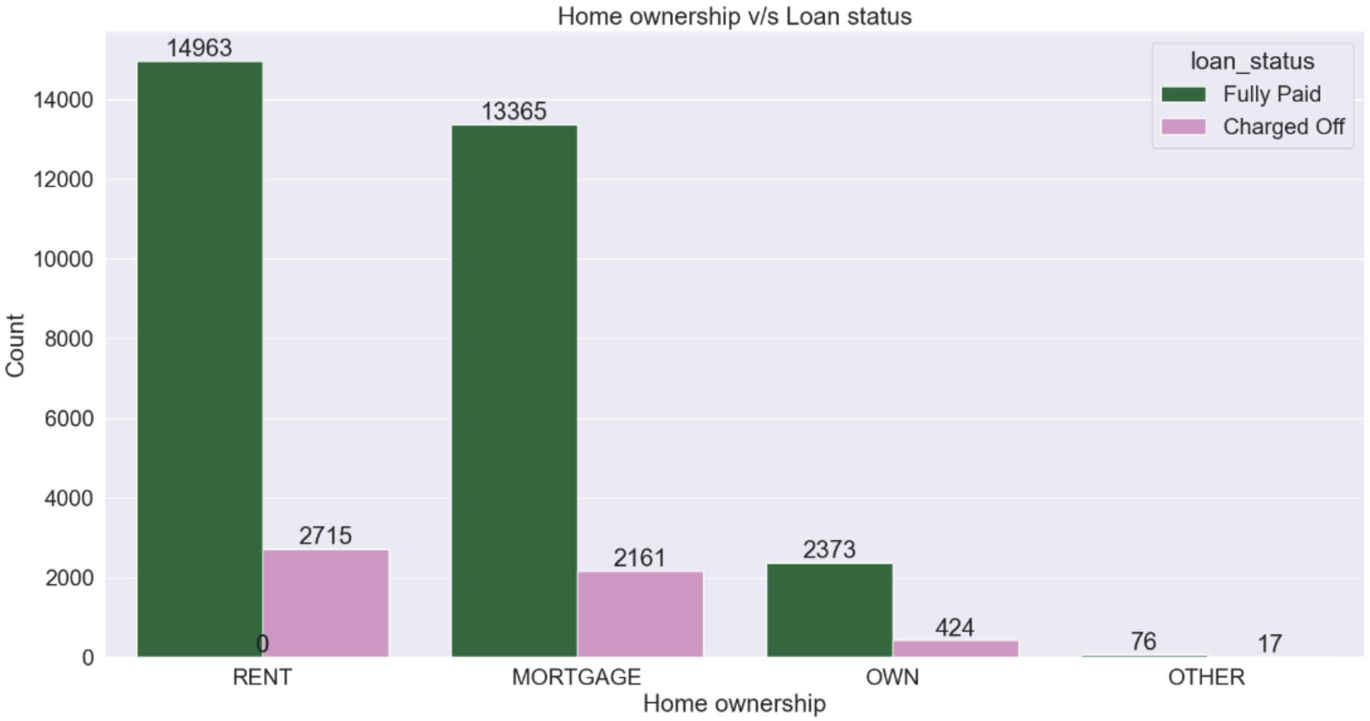
4th Quarter indicates receiving maximum loan applications. This may be because of the holiday season.



Bivariate analysis - Home ownership v/s loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.countplot(x='home_ownership',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['home_ownership'].value_counts().index,palette='cubehelix')
ax.set_xlabel('Home ownership')
ax.set_ylabel('Count')
ax.set_title('Home ownership v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

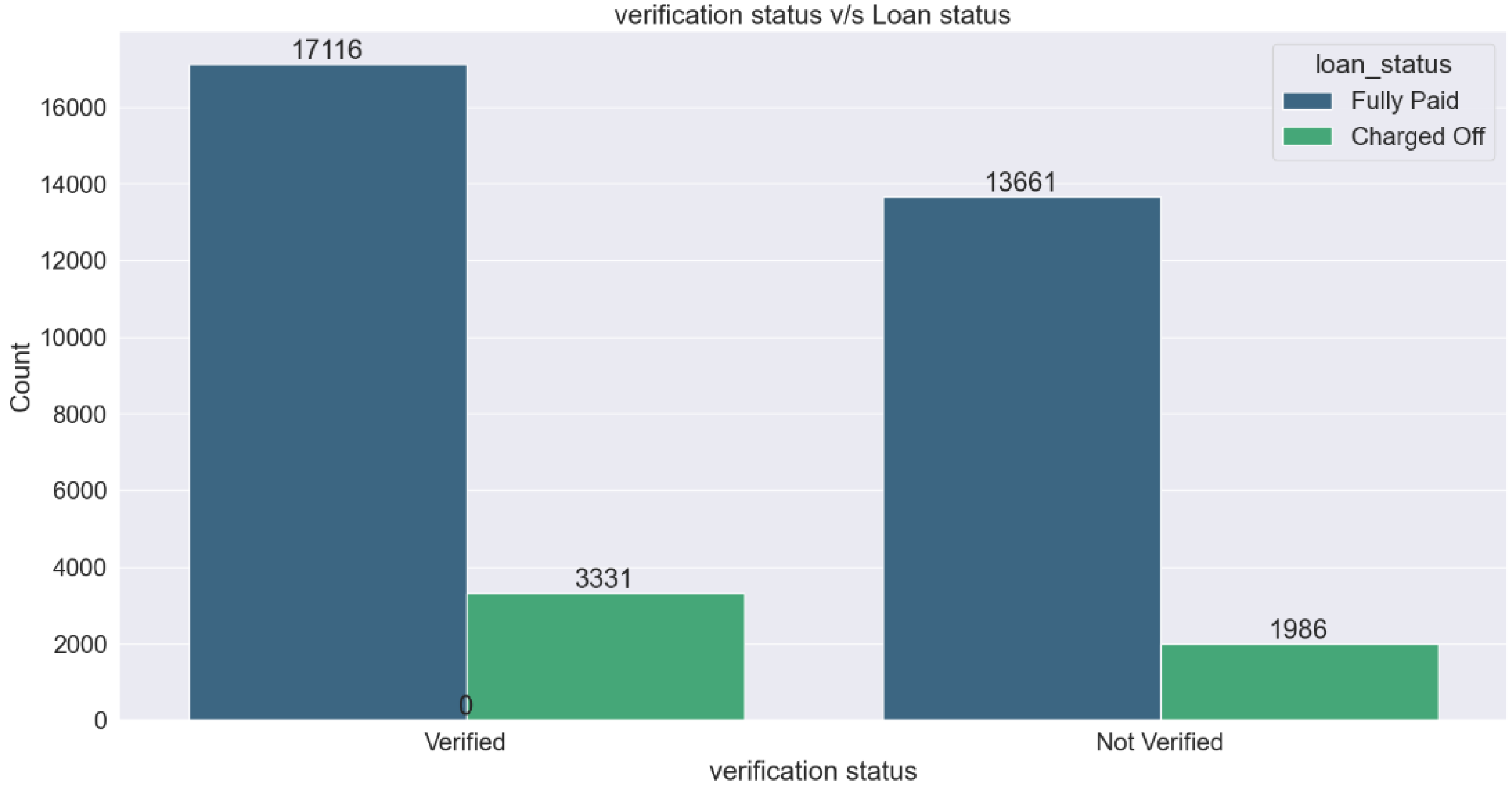
Loan applications received from those living in rented house and mortgage contribute to the majority and they are more likely to be defaulters



Bivariate analysis - Verification status v/s loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.countplot(x='verification_status',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['verification_status'].value_counts().index,palette='viridis')
ax.set_xlabel('verification status')
ax.set_ylabel('Count')
ax.set_title('verification status v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

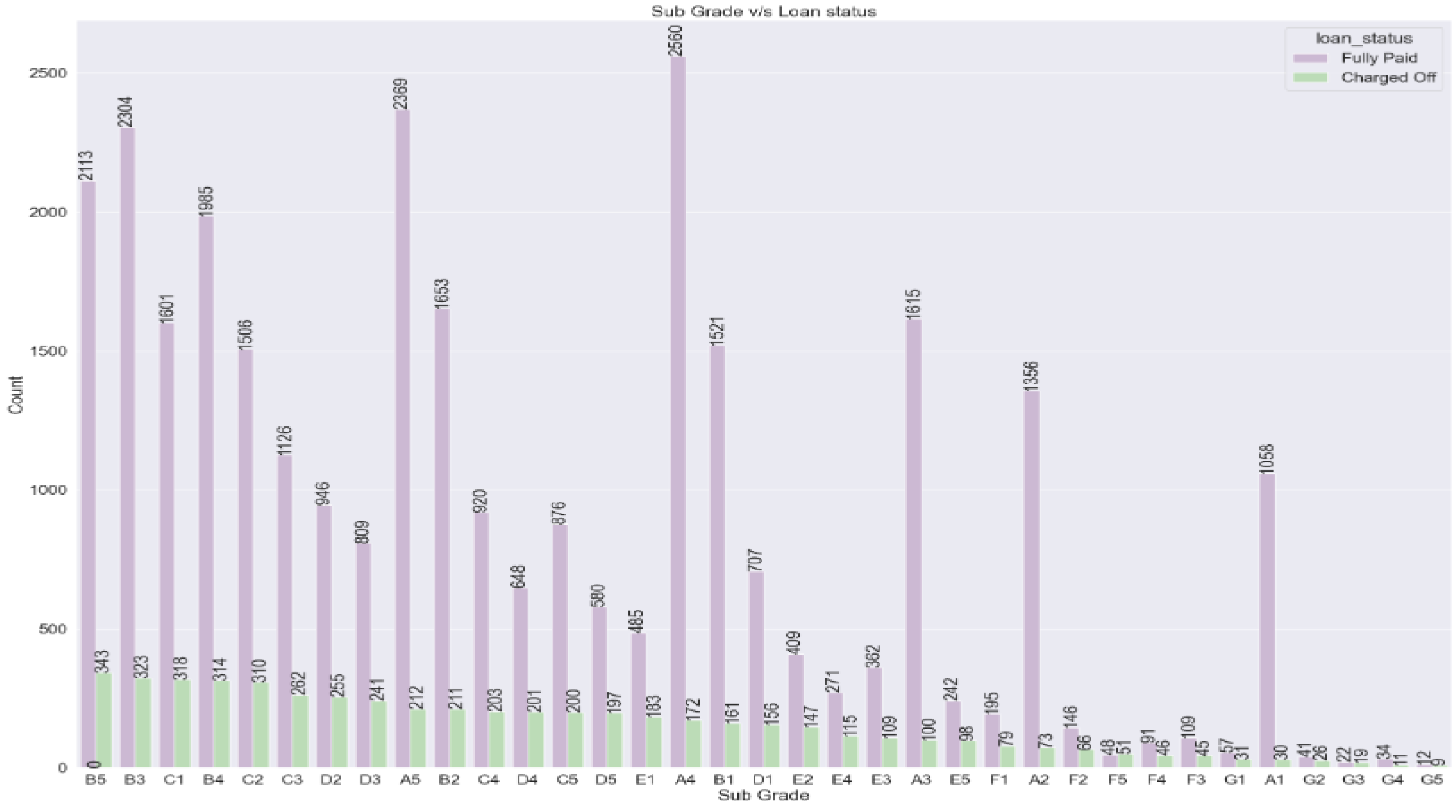
Even the verified loans are getting defaulted. The verification process needs to be strengthened



Bivariate analysis - Purpose v/s loan status

```
fig, ax = plt.subplots(figsize=(20,20), dpi=100)
sns.countplot(x='purpose',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['purpose'].value_counts().index,palette='PiYG')
ax.set_xlabel('purpose')
ax.set_ylabel('Count')
ax.set_title('purpose v/s Loan status')
plt.xticks(rotation=90)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=45)
plt.show()
```

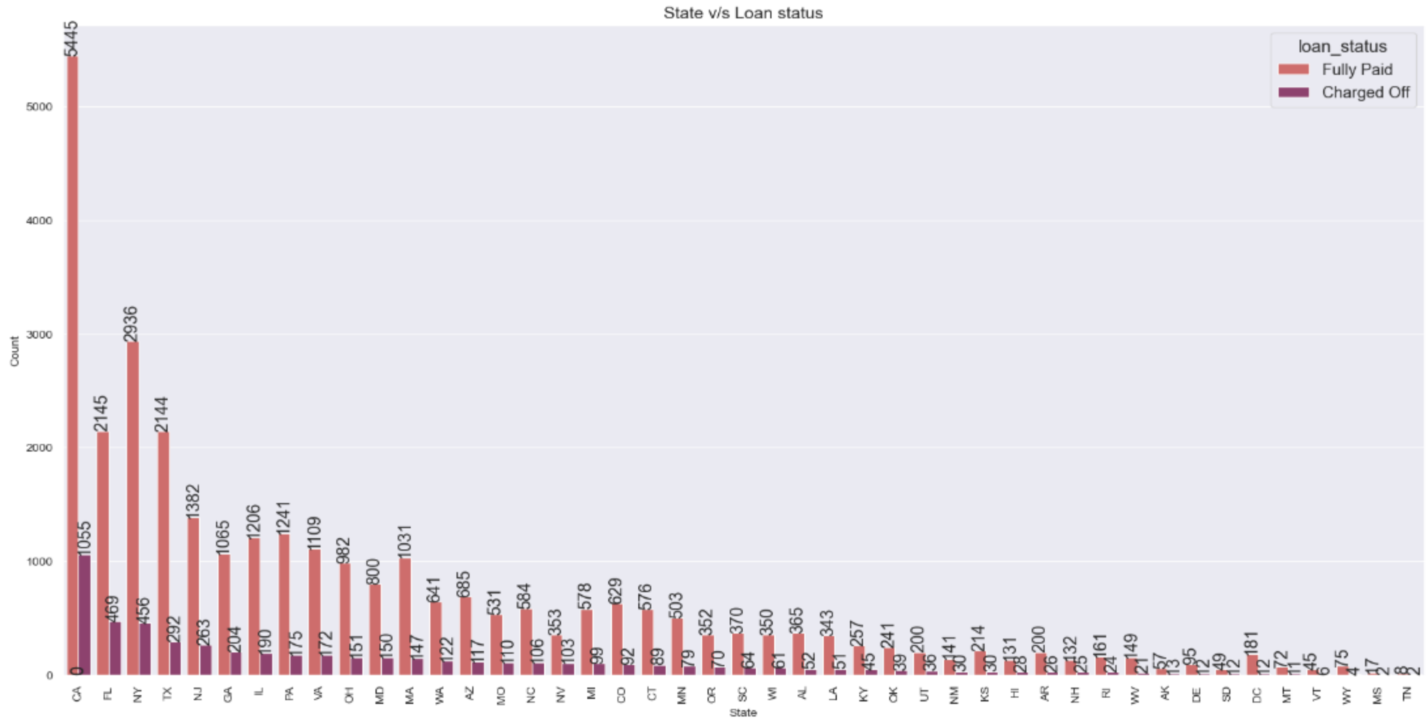
Majority loan has been approved on debt consolidation and this mostly contributes to charged off loans.



Bivariate analysis - Home ownership v/s loan status

```
fig, ax = plt.subplots(figsize=(20,10), dpi=100)
sns.set_context("paper", font_scale=1.5)
sns.countplot(x='addr_state',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['addr_state'].value_counts().index,palette='flare')
ax.set_xlabel('State',fontsize=10)
ax.set_ylabel('Count',fontsize=10)
ax.set_title('State v/s Loan status' ,fontsize=14)
ax.tick_params(axis='both', which='major', labelsize=10)
plt.xticks(rotation=90)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=90)
plt.show()
```

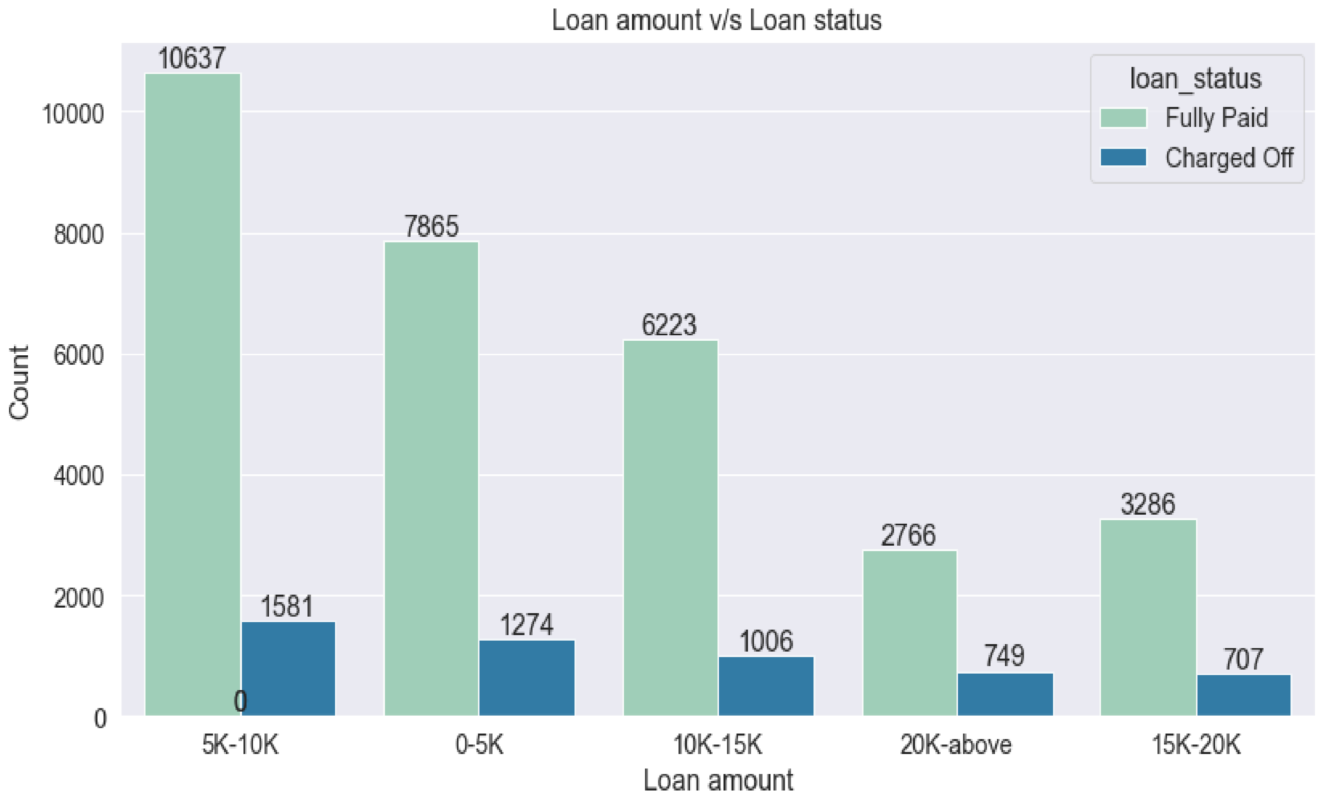
There are more loan applicants from States CA, FL & NY and they make to the majority of the defaulters list. Going forward there will be more defaulters from the above mentioned states.



Bivariate analysis - Loan amount v/s loan status

```
fig, ax = plt.subplots(figsize=(12,6), dpi=80)
sns.countplot(x='Loanamt_bucket',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['Loanamt_bucket'].value_counts().index,palette='YlGnBu')
ax.set_xlabel('Loan amount')
ax.set_ylabel('Count')
ax.set_title('Loan amount v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

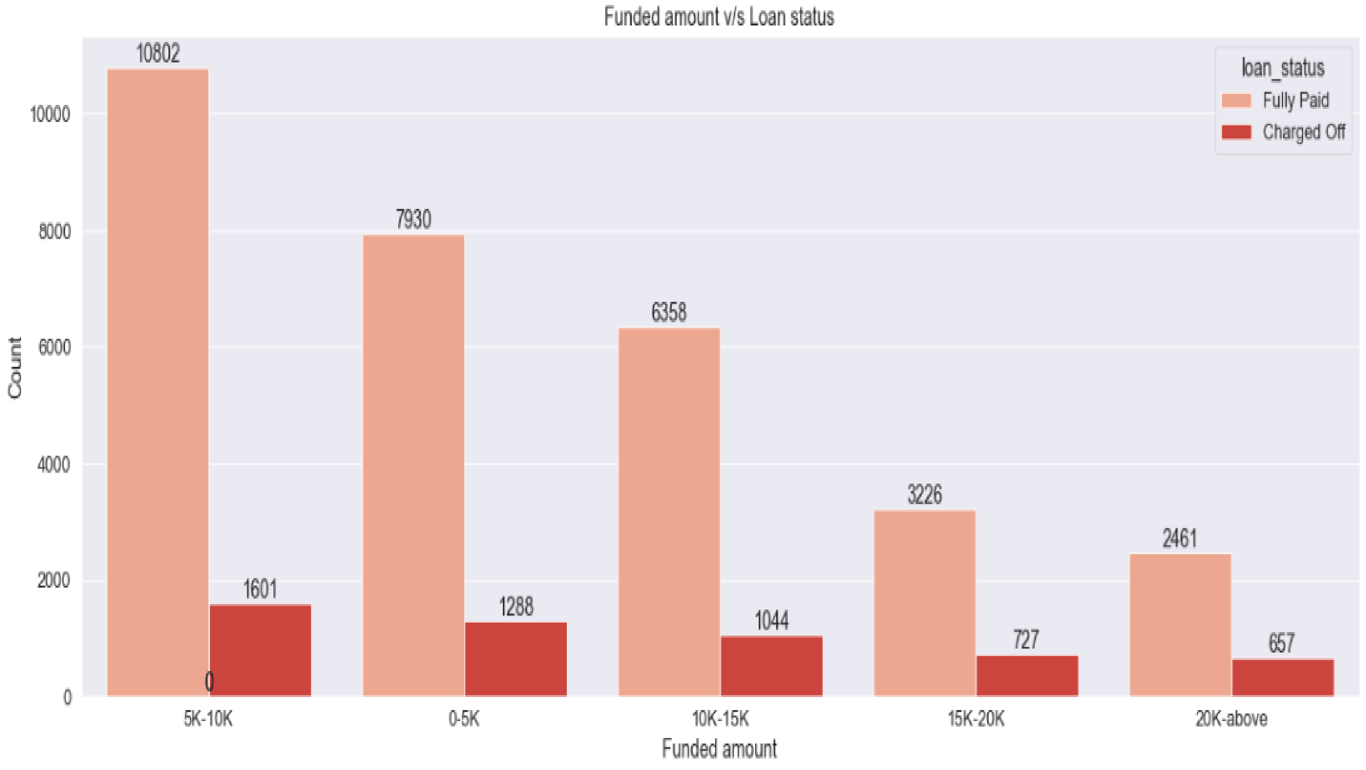
Majority of loan amount applied is between 5K to 10K and the majority defaulters are in this category. This pattern could continue as trend. The lending company should be cautious when approving higher amount loans.



Bivariate analysis - Funded amount v/s Loan status

```
fig, ax = plt.subplots(figsize=(20,7), dpi=70)
sns.countplot(x='funded_amnt_bucket',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['funded_amnt_bucket'].value_counts().index,palette='Reds')
ax.set_xlabel('Funded amount')
ax.set_ylabel('Count')
ax.set_title('Funded amount v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

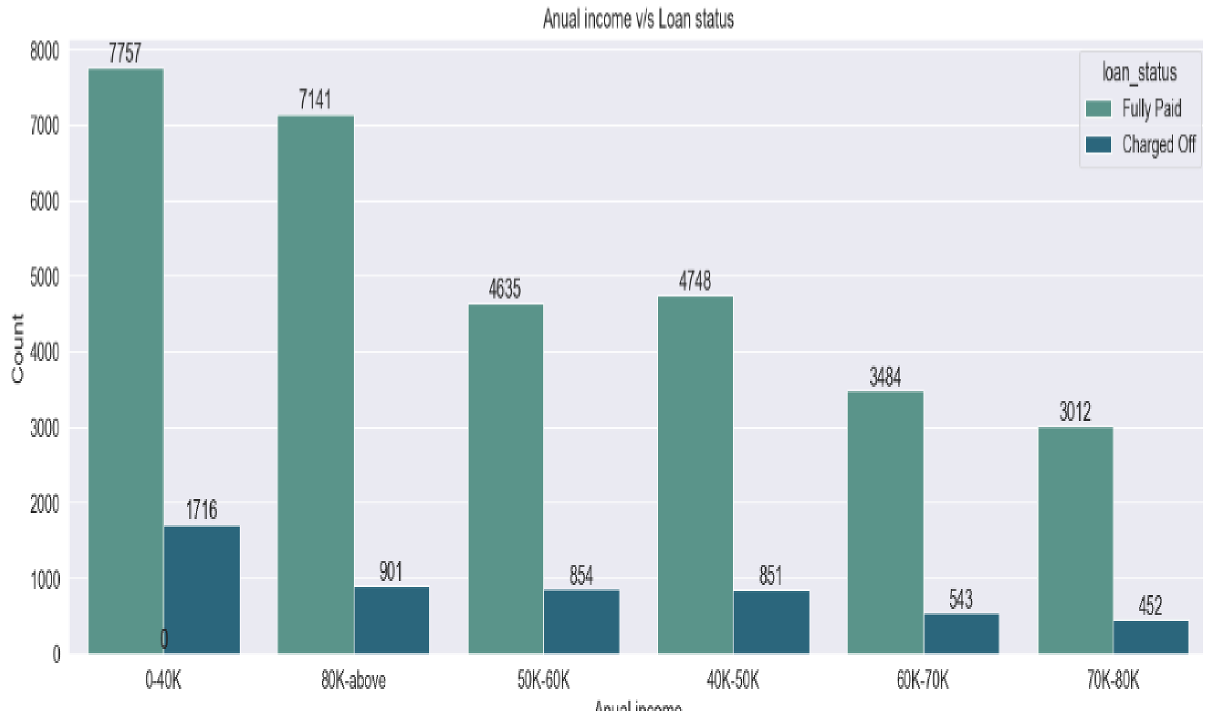
Majority of loan approved are in the range of 5K-10K. Defaulters being majority in this category the lending company should perform proper check before loan approval.



Bivariate analysis - Annual income v/s Loan status

```
fig, ax = plt.subplots(figsize=(20,6), dpi=100)
sns.countplot(x='annual_inc_bucket',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['annual_inc_bucket'].value_counts().index,palette='crest')
ax.set_xlabel('Anual income')
ax.set_ylabel('Count')
ax.set_title('Anual income v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

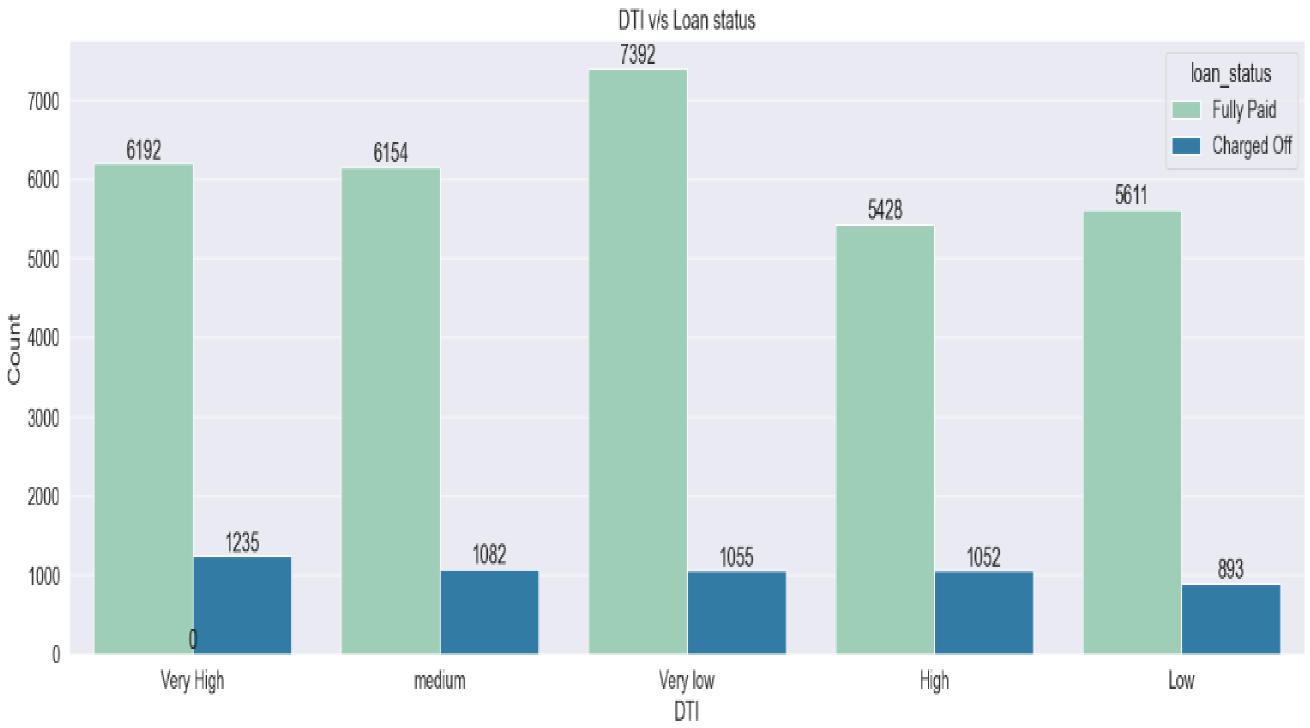
Majority of loan applicants are with salary less than 40000 and they contribute to most of the charged off loans. Less annual income leads to defaulting tendency.



Bivariate analysis - DTI v/s Loan status

```
fig, ax = plt.subplots(figsize=(20,6), dpi=100)
sns.countplot(x='dti_bucket',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['dti_bucket'].value_counts().index,palette='YlGnBu')
ax.set_xlabel('DTI')
ax.set_ylabel('Count')
ax.set_title('DTI v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()), ha='center',va='bottom',rotation=0)
plt.show()
```

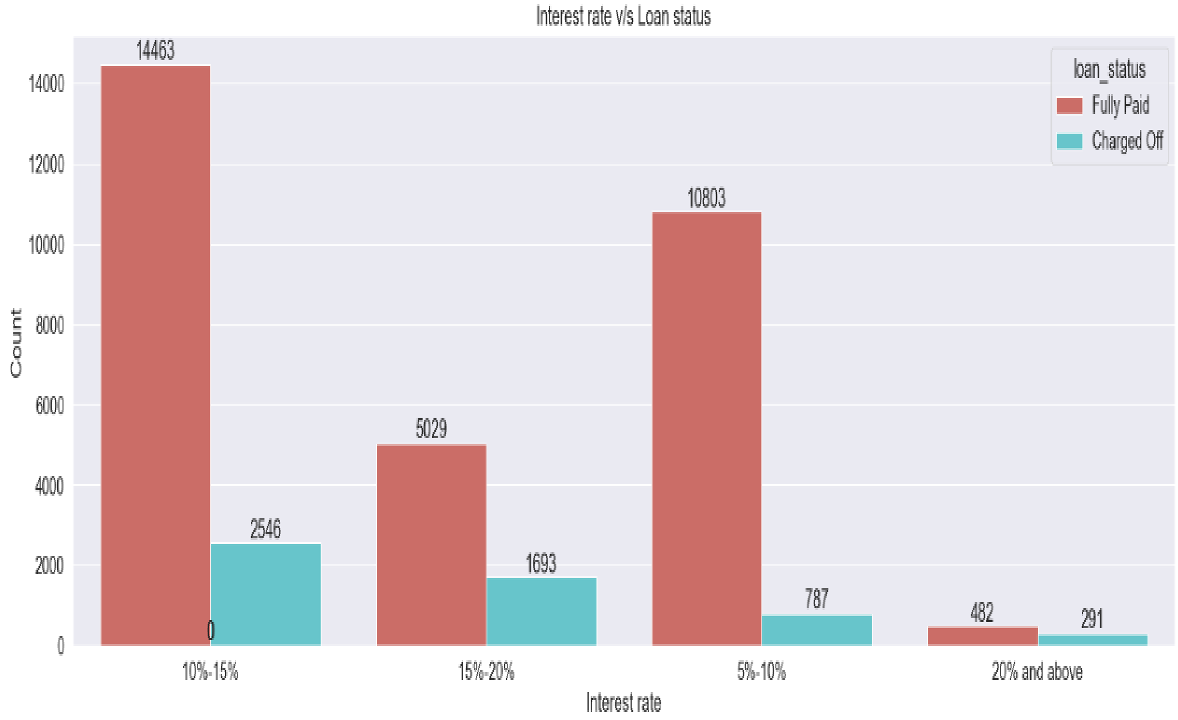
High count of loans get charged off in cases where the Debt-to-income ratio is very high. Lending company should focus on granting loans for applicants with low DTI to reduce credit loss.



Bivariate analysis - Interest rate V/s Loan status

```
fig, ax = plt.subplots(figsize=(20,6), dpi=100)
sns.countplot(x='int_rate_bucket',hue='loan_status',data=df_loan,
              order=df_loan_chargedoff['int_rate_bucket'].value_counts().index,palette='hls')
ax.set_xlabel('Interest rate')
ax.set_ylabel('Count')
ax.set_title('Interest rate v/s Loan status')
plt.xticks(rotation=0)
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}',(p.get_x() + p.get_width() / 2,p.get_height()),
                       ha='center',va='bottom',rotation=0)
plt.show()
```

A significant count of loan defaulters received loans with interest rates falling within the range of 10% to 15%.Lending company should modify the interest rates such that the customer can repay and the risk on credit loss can be avoided.



Observations and inferences from Bivariate analysis

- **Ordered categorical variable**
    - Loan applicants in Grades B, C, ano D account for the majority of ‘Charged Off’ loans. However, Grades A and B make up the largest portion of total loan applicants. This indicates that Grade A applicants positively impact the business, while caution is needed when approving loans for Grade B, C and D applicants.
    - Loan subgrade B3,B4,B5,C1,C2,C3 & D2 contribute to most of Charged off loans.
    - Most charged-off loans are short-term. However, when comparing the total loans for each term, the percentage of long-term loans being charged off is higher than that of short-term loans. In this case company should assess risk with long term loan and adjust term or the interest rate accordingly.
    - Employees having 10 or more years of experience are majority of loan applicants. They are also most likely to be defaulters. Experience should not be the only criteria for approving loans.
    - December month is the month with the maximum number of loan applicants. The company should adopt efficient way to meet customer needs with minimal risk on credit loss.
    - From 2007 to 2011 there is a positive trend of loan applicants. This shows there is a positive trend of more applicants in upcoming years. The company should also note that with increasing trend of loan applicants, robust risk mitigation measures should be adopted.

Observations and inferences from Bivariate analysis

- **Unordered categorical variable**
  - Loan applications received from those living in rented houoe and mortgage contribute to the majority and they are more likely to be defaulters. The lending company should take appropriate measures while lending loans to applicants with rented houses and mortgage.

  - Verified loans are also getting defaulted. The Verification process need to be strengthened.

  - Majority loan has been approved on debt consolidation and this mostly contributes to charged off loans. To manage this situation, company should take appropriate measures to ensure loan repayment.

  - There are more loan applicants from States CA, FL & NY and they make to the majority of the defaulters list. Going forward there will be more defaulters from the above mentioned states.Company should take more stringent rules and efficient methodology while approving loans to applicants from these states.

  - High count of loans get charged off in cases where the Debt-to-income ratio is very high. Lending company should focus on granting loans for applicants with low DTI to reduce credit loss.

  - A significant count of loan defaulters received loans with interest rates falling within the range of 10% to 15%. Lending company should modify the interest rates such that the customer can repay and the risk on credit loss can be avoided.

Observations and inferences from Bivariate analysis

- **Quantitative Variables**
  - Majority of loan amount applied is between 5K to 10K and the majority defaulters are in this category. This pattern could continue as trend. The lending company should be cautious when approving higher amount loans.

  - Majority of loan applicants are with salary less than 40000 and they contribute to most of the charged off loans. Less annual income leads to defaulting tendency. The company should implement a cap on loan amounts based on income levels to ensure affordability for the borrower.

  - Higher DTI and interest rate ranging from 10 to 15% are likely to contribute to loan default. Company should evaluate on DTI and adjust interest rates accordingly so that customer can repay and the risk on credit loss can be avoided.

Correlation analysis

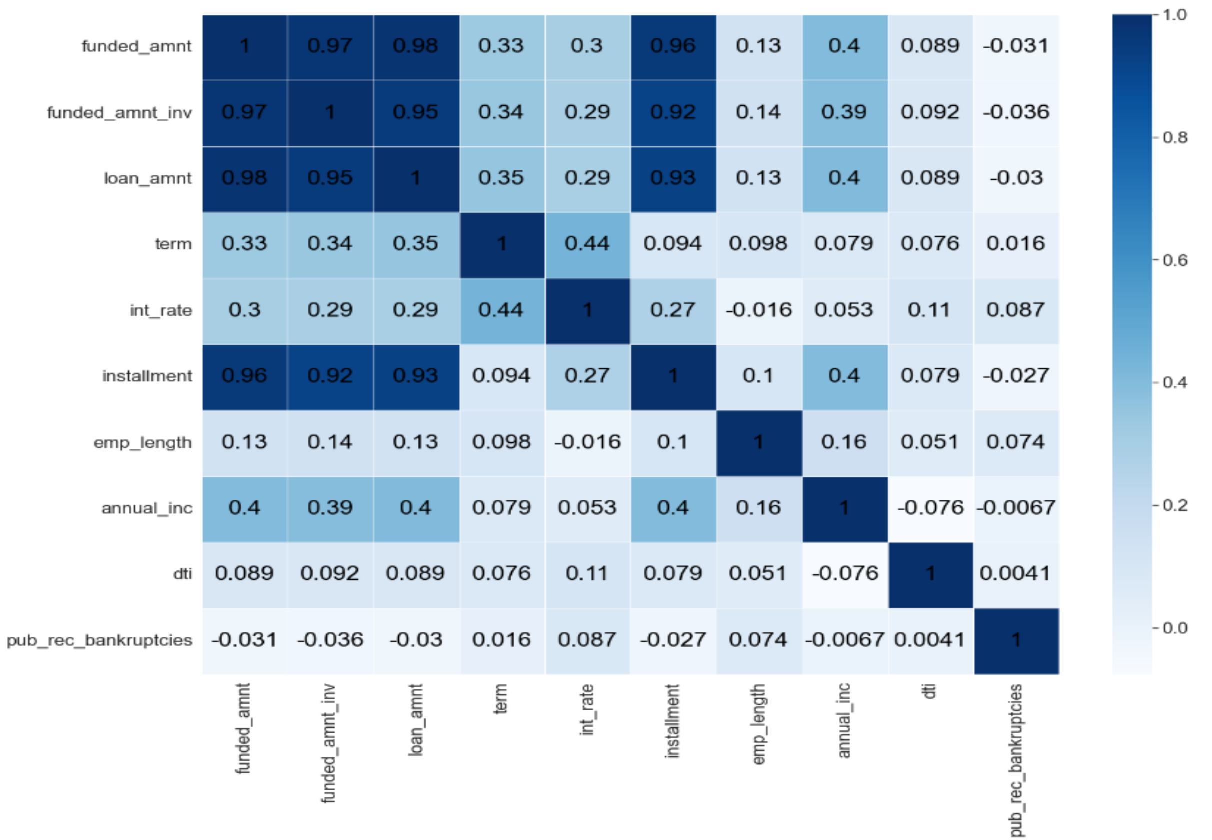
```
df_corr = df_loan[['funded_amnt', 'funded_amnt_inv' ,'loan_amnt','term' , 'int_rate','installment','emp_length',
                   'annual_inc','dti','pub_rec_bankruptcies']]

corr=df_corr.corr()

plt.figure(figsize = (20,15),dpi=80)
sns.set(font_scale=1.8)
sns.heatmap(corr,annot=True,xticklabels=corr.columns.values,yticklabels=corr.columns.values,linewidths=.5,cmap="Blues",
            annot_kws={"size": 25 ,"color": "black"})
plt.show()
```

Inference from Correlation Analysis

- **Strong correlation**
  - installment has strong correlation with loan_amnt, funded_amnt_inv and funded_amnt.
  - loan_amnt has strong correlation with funded_amnt and funded_amnt_inv.
- **Weak correlation**
  - dti has weak correlation with most of the columns.
  - emp_length has weak correlation with most of the columns.
- **Negative correlation**
  - pub_rec_bankruptcies has negative correlation with most of the fields
  - emp_length has negative correlation with int_rate.
  - annual_inc has negative correlation with dti.



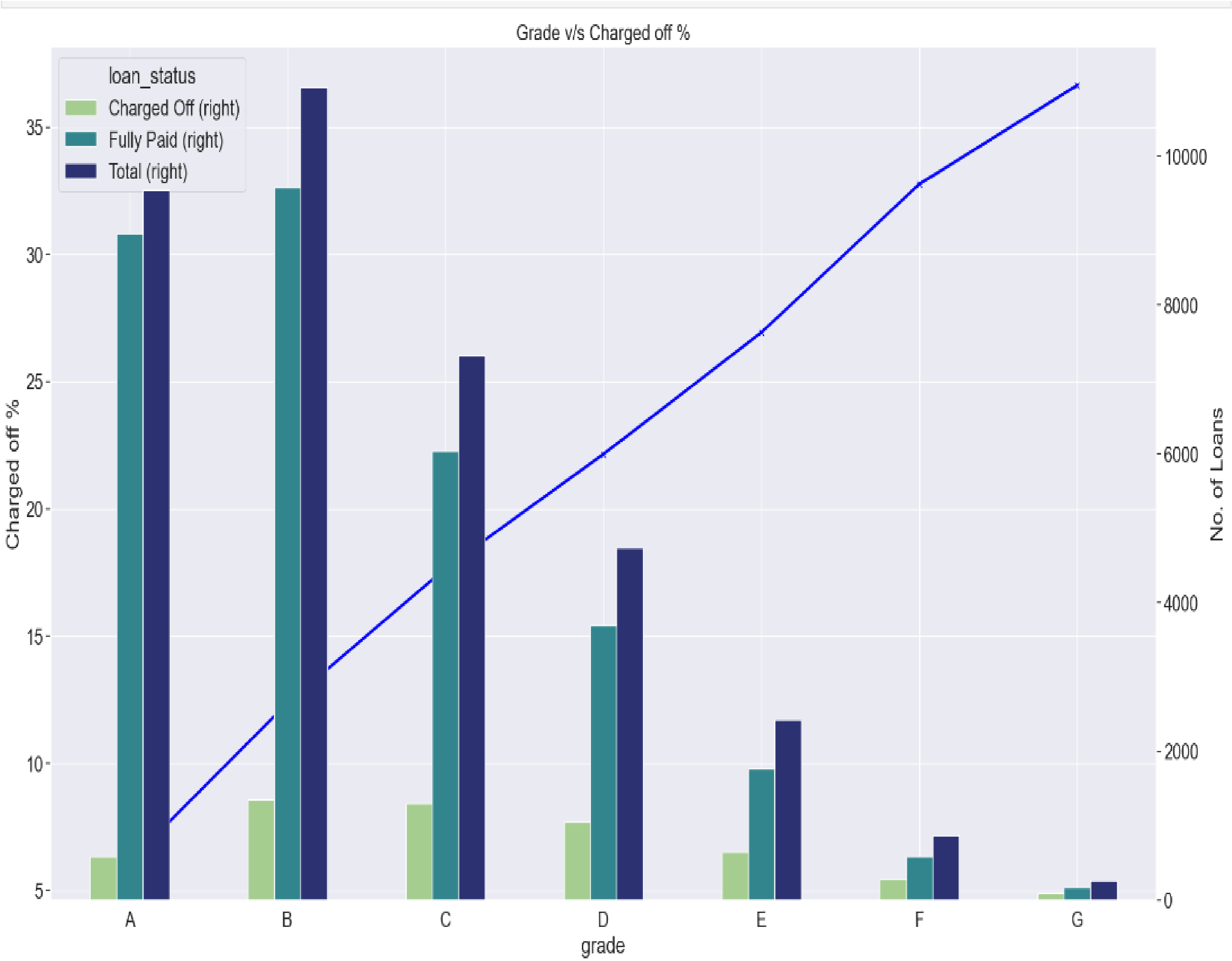
Multivariate analysis

Multivariate analysis will be done for the following columns against the Loan status and Charge off %

1.grade
2.Sub Grade
3.Emp_length
4.State
5.Purpose
6.Home ownership
7.verification status
8.anual income
9.DTI
10.interest rate

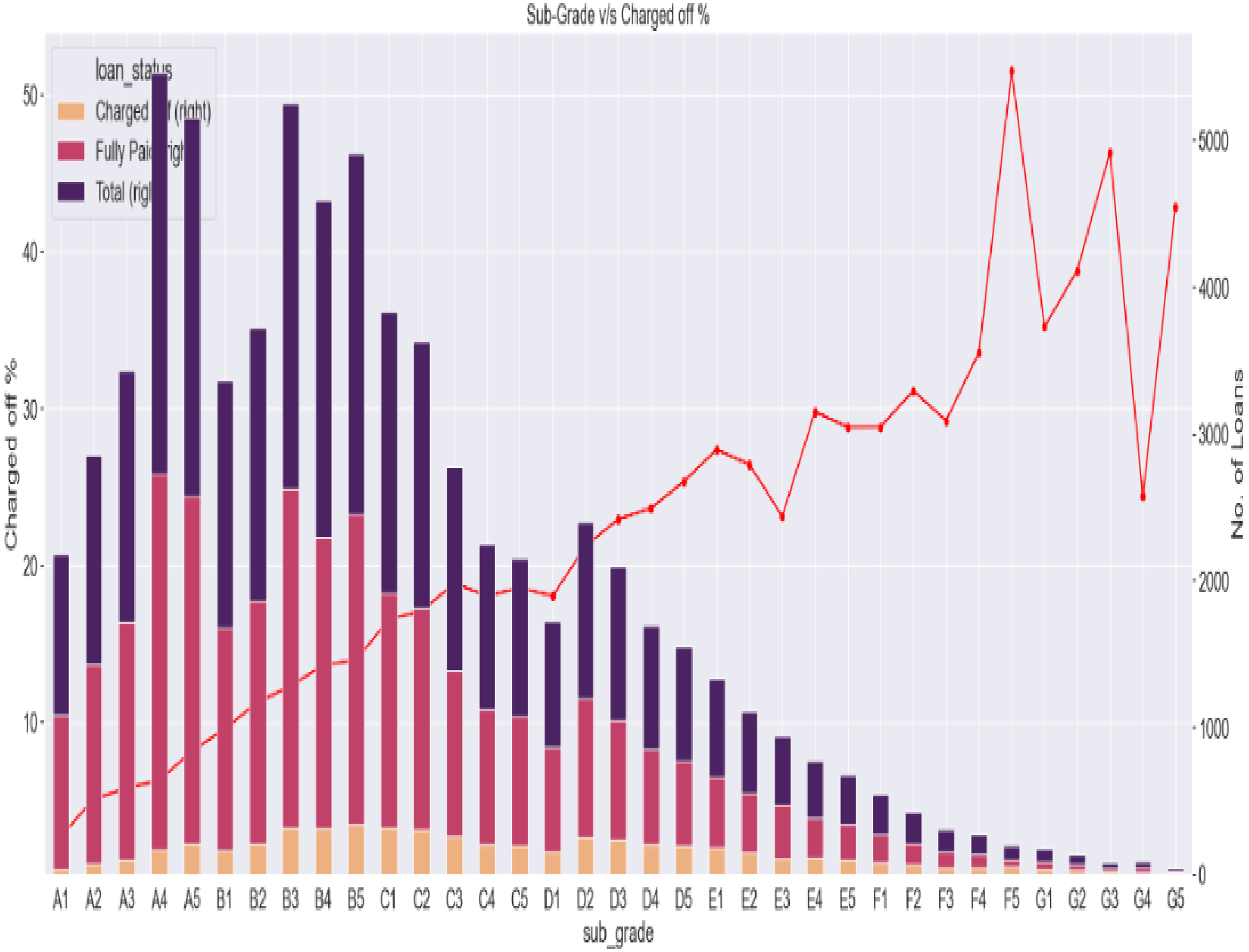
Multivariate analysis - Grade

The tendency to default is higher for Grades B, C, and D. However, the charge-off percentages relative to the total number of loans are highest for Grades G and F, at 36.6% and 32.8%, respectively. Despite having fewer loans in Grades G and F, caution should be applied when approving loans on these grades too as the likelihood of default is higher in these grades as well.



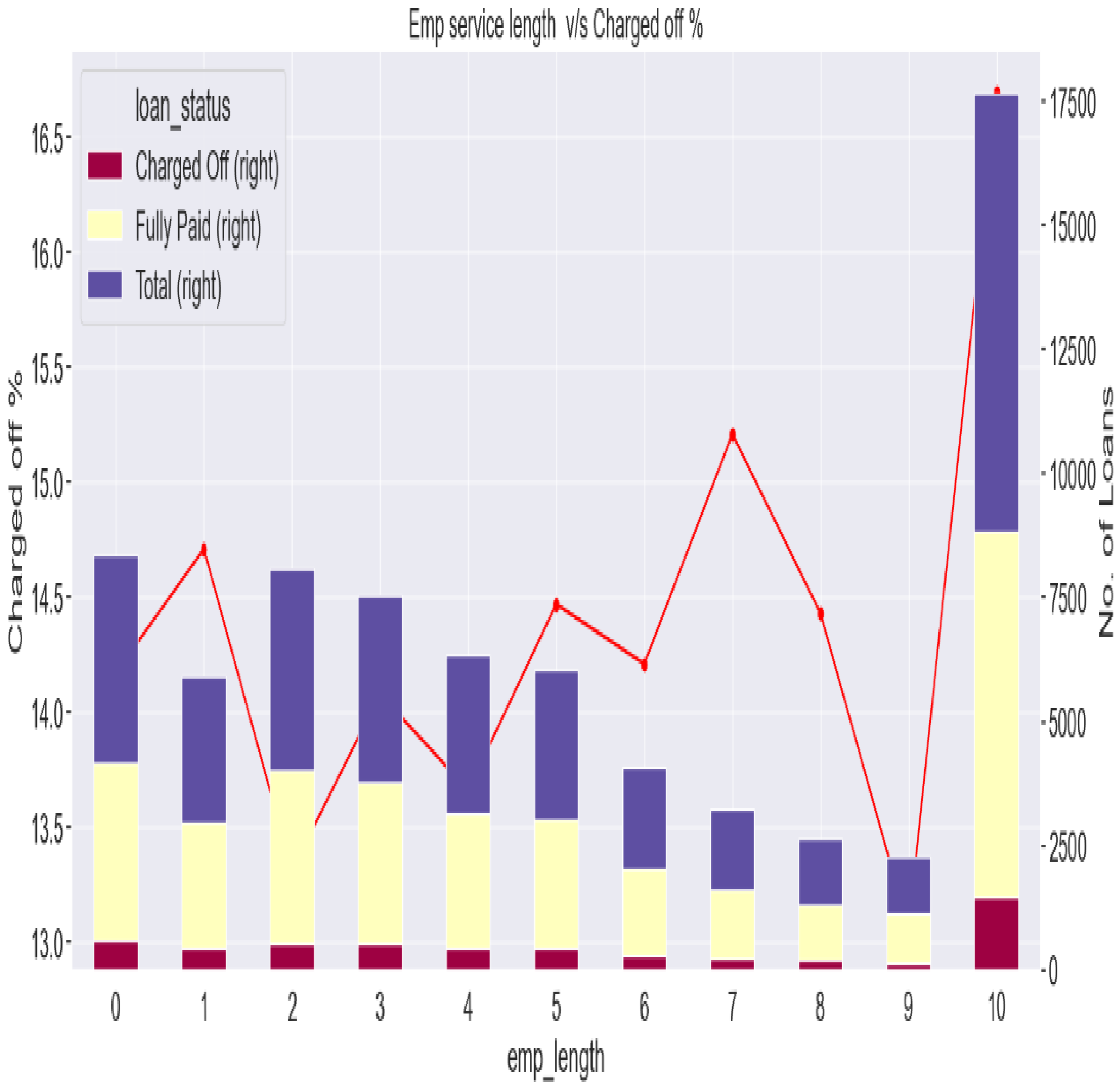
Multivariate analysis – Sub Grade

Borrowers in sub-grades B3, B4, B5, C1, C2, and D2 are likely to default. However, in terms of charge-off percentage relative to the total number of loans, F5 and G3 have the highest rates. The best performance in this category is seen in grades A1 to A5, with less than 10% of borrowers defaulting. Company should also provide special attention to subgrades F5 and G3 while approving loans.



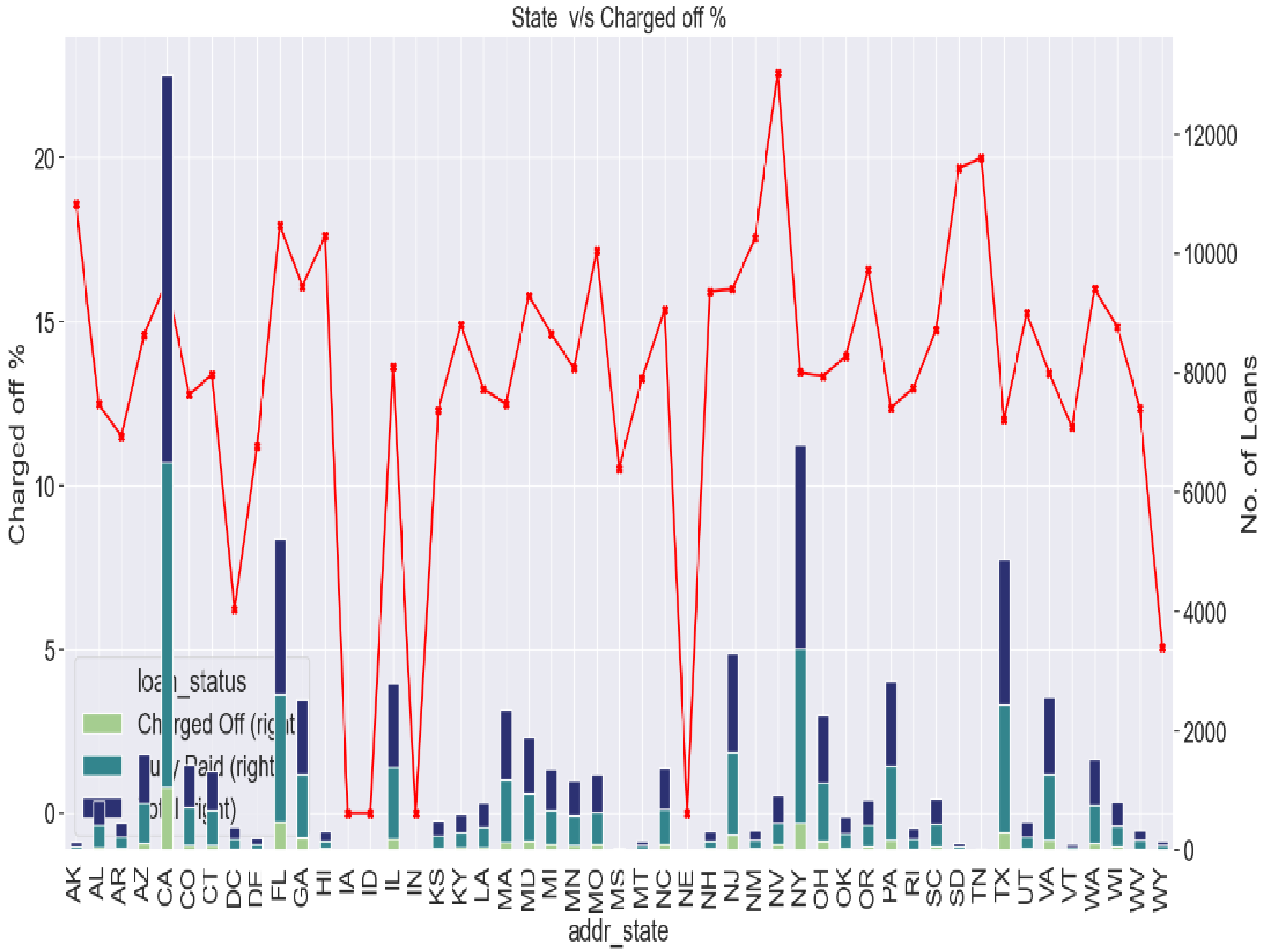
Multivariate analysis – Employee length of service

Employees with over 10 years of service had a charge-off rate of 16.5%.This clearly indicates that the length of service is not a reliable criterion for loan approval.



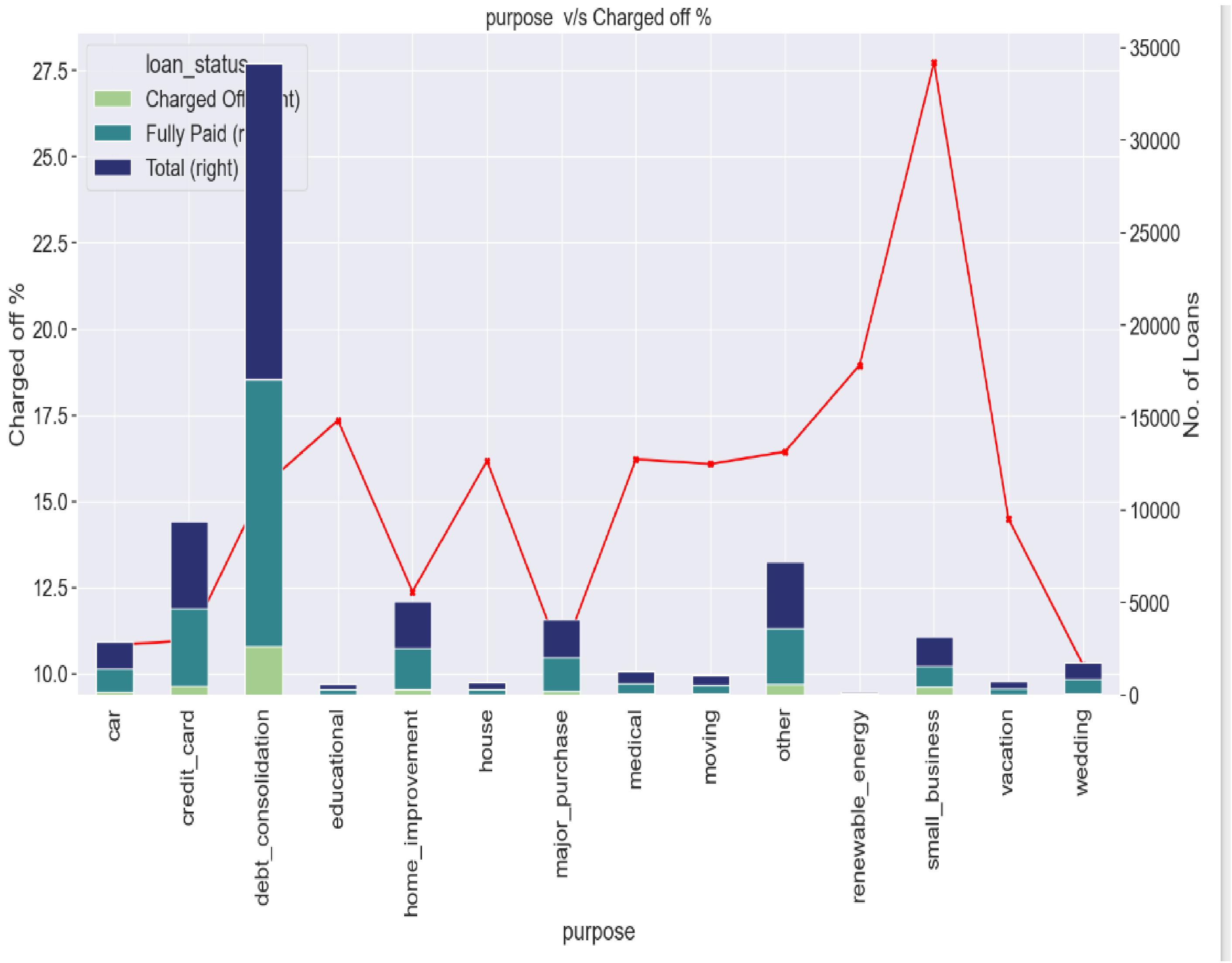
Multivariate analysis - State

Although the majority of loan defaulters are from CA, FL, and NY, there are 17 states with a charge-off rate more than 15%. The company should implement an efficient methodology to thoroughly evaluate loan applicants from these states before approving loans.



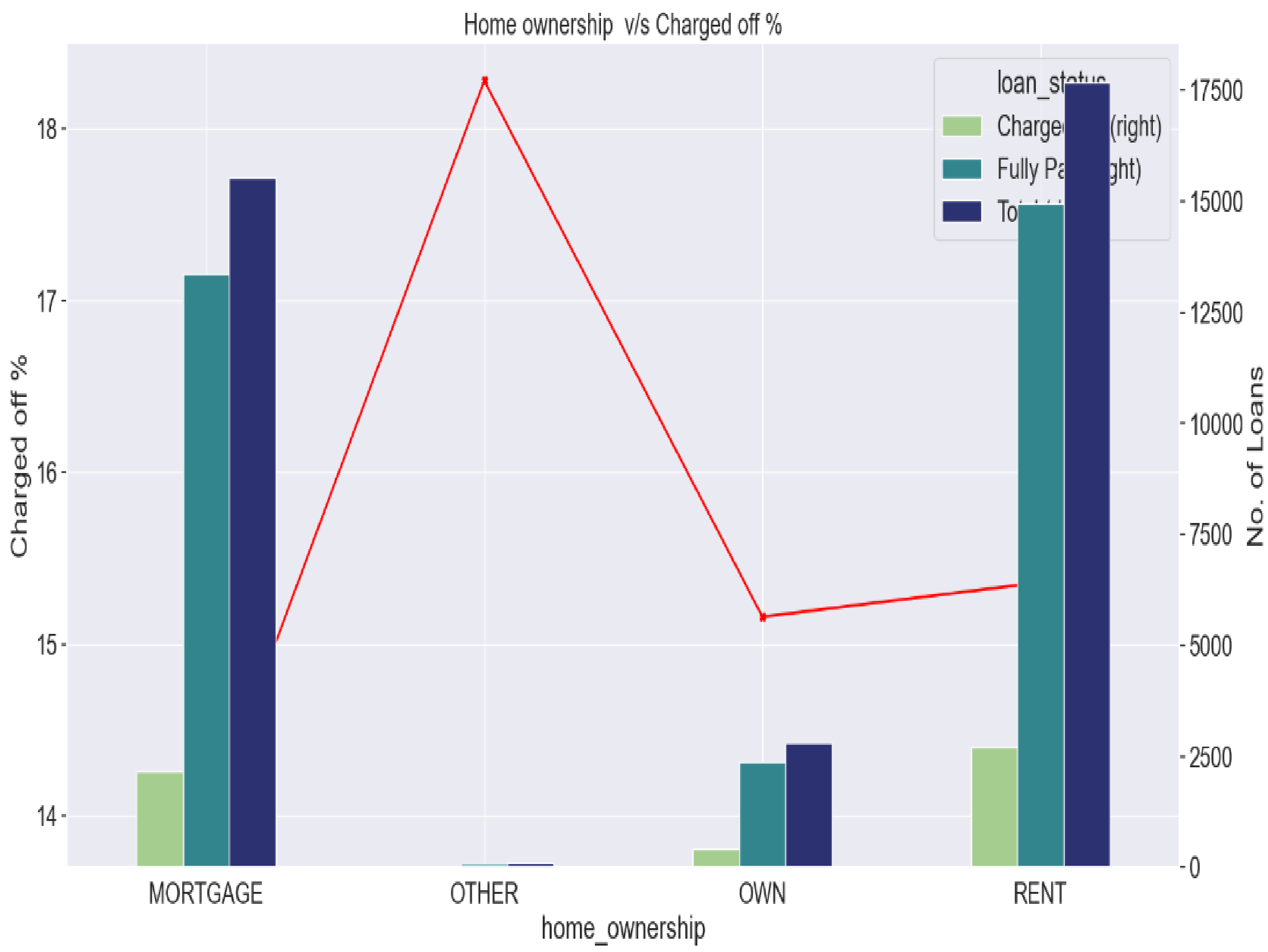
Multivariate analysis - purpose

Even though majority of defaulters are in debt consolidation , considering Charge off percentage, small Business contribute the maximum which is 27.7%. Lending company should have proper measures in place for approving small business loans.



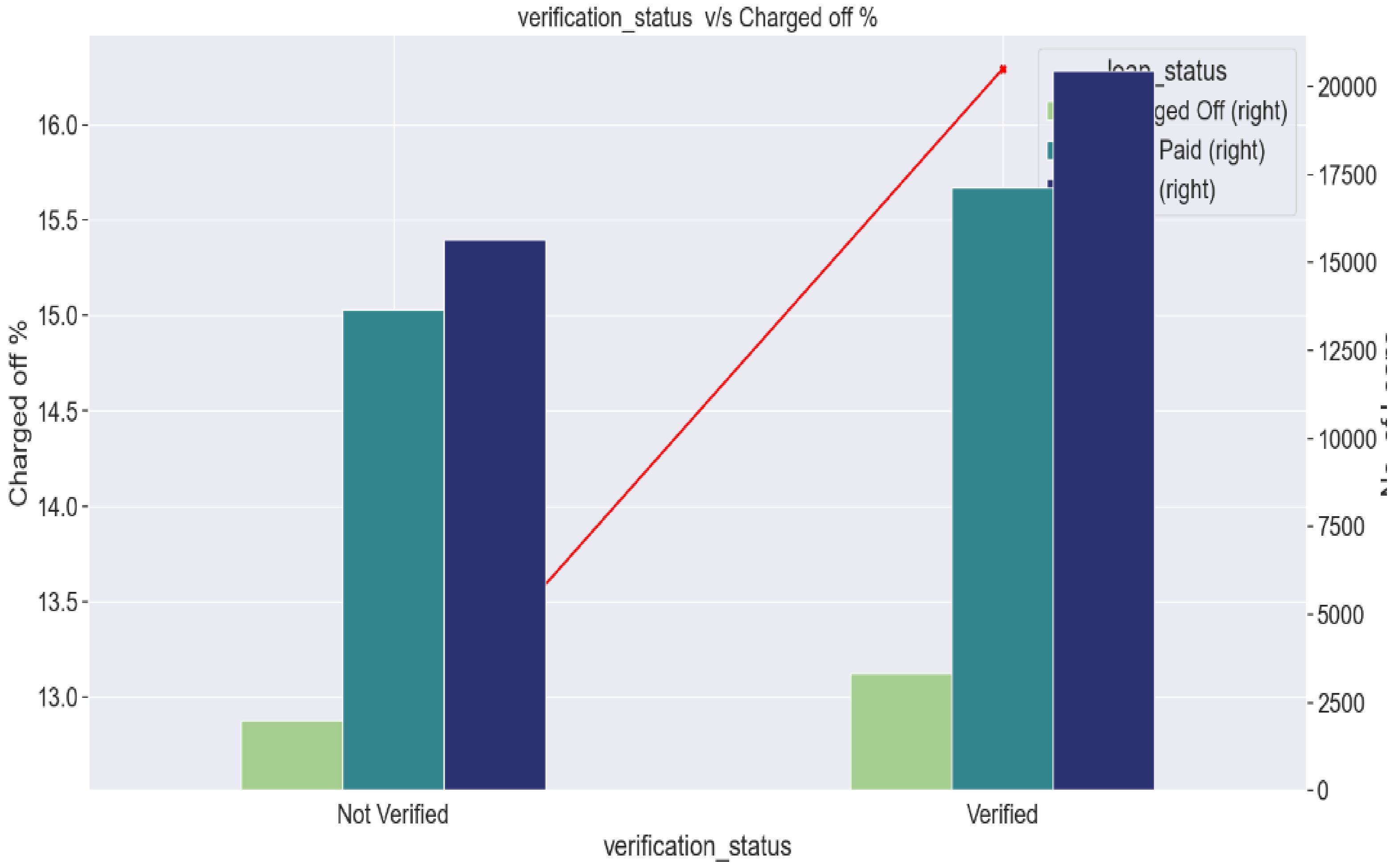
Multivariate analysis – Home ownership

Charge off % is more in Other living type.



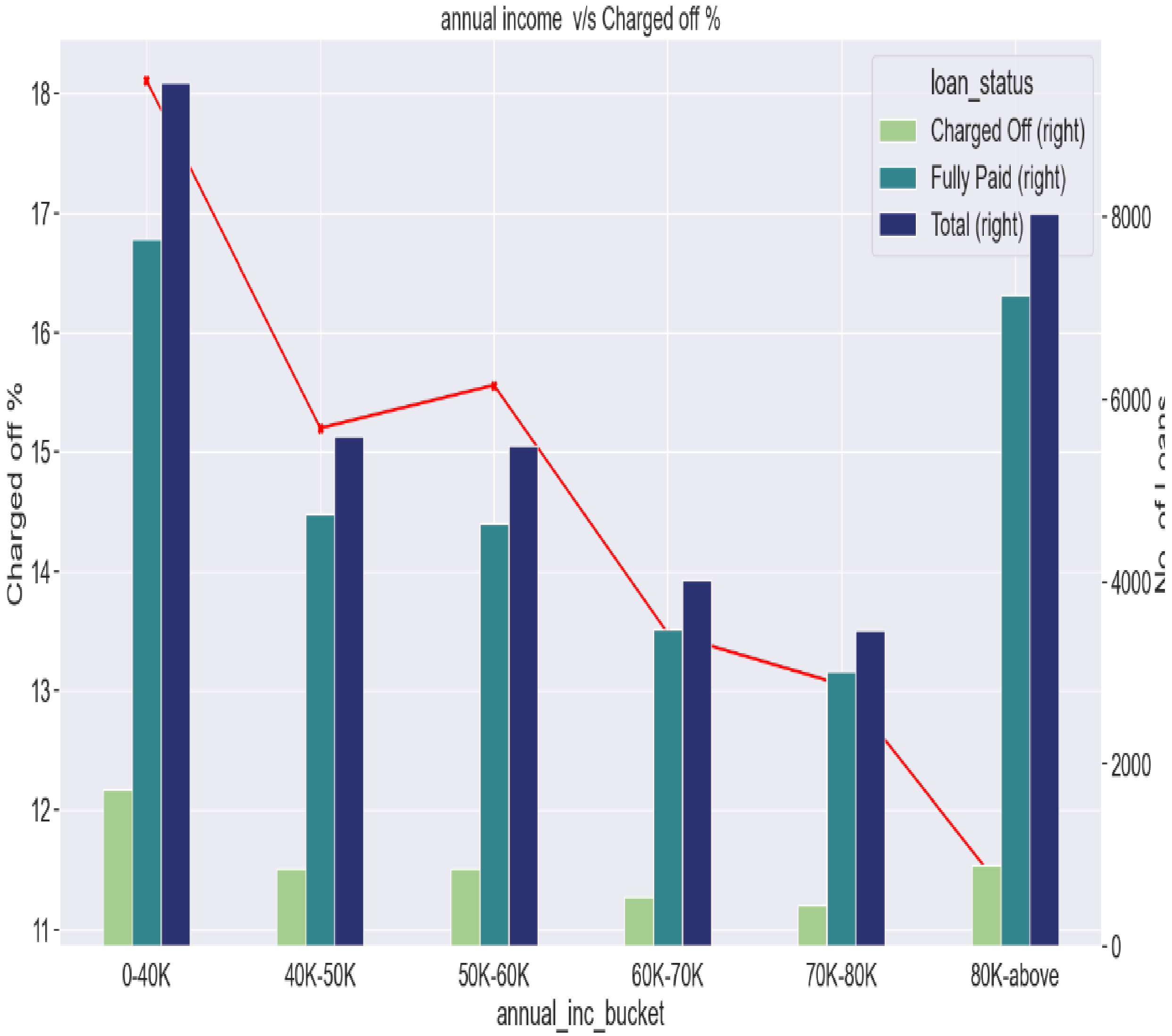
Multivariate analysis – Verification status

Charge of % is more on Verified loans



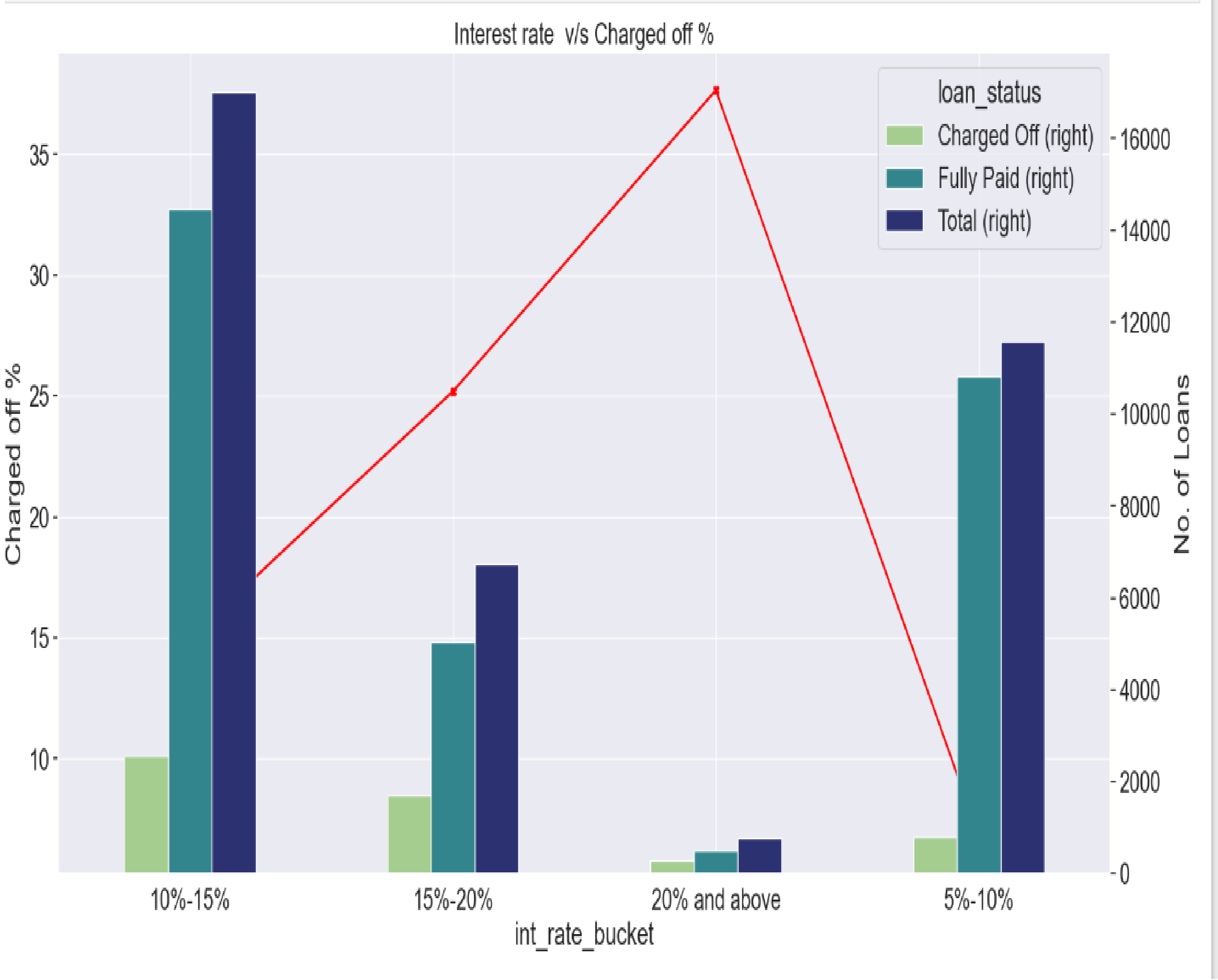
Multivariate analysis – Annual income

The charge-off rate is highest for employees earning less than 40K, at 18.1%. Borrowers in lower income groups have the highest tendency to default on loans, and this tendency generally decreases as annual income increases.



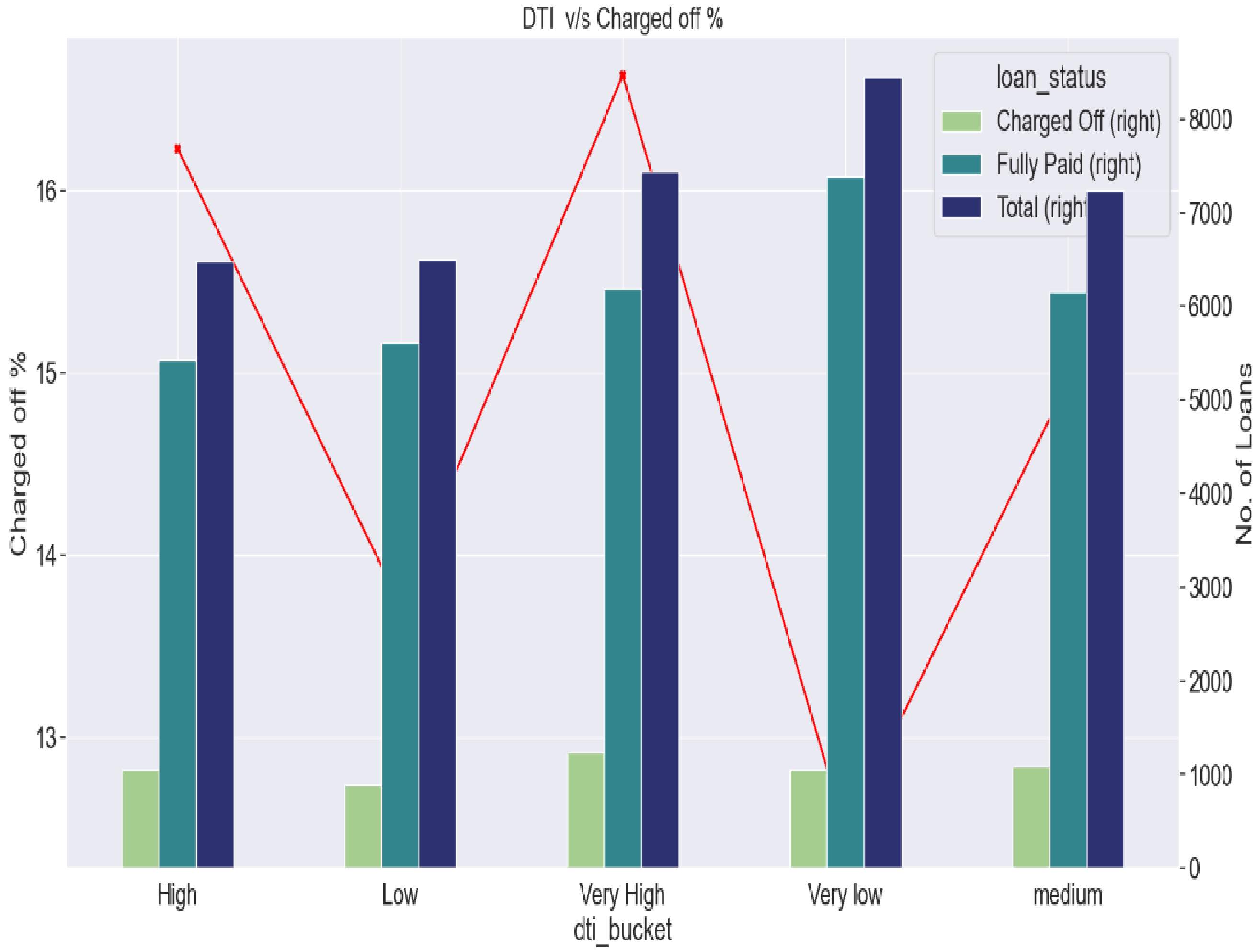
Multivariate analysis – interest rate

Rise in interest rate will lead to higher default percentage.



Multivariate analysis – Debt-to-income

Increase in DTI will lead to higher default percentage. Before approving loans, the company should verify the open debts and assess the payment capability of loan applicants relative to their income.



Multivariate analysis - Summary

• The tendency to default is higher for Grades B, C, and D. However, the charge-off percentages relative to the total number of loans are highest for Grades G and F, at 36.6% and 32.8%, respectively. Despite having fewer loans in Grades G and F, caution should be applied when approving loans on these grades too as the likelihood of default is higher in these grades as well.

•Borrowers in sub-grades B3, B4, B5, C1, C2, and D2 are likely to default. However, in terms of charge-off percentage relative to the total number of loans, F5 and G3 have the highest rates. The best performance in this category is seen in grades A1 to A5, with less than 10% of borrowers defaulting. Company should also provide special attention to subgrades F5 and G3 while approving loans.

•Employees with over 10 years of service had a charge-off rate of 16.5%.This clearly indicates that the length of service is not a reliable criterion for loan approval.

•Although the majority of loan defaulters are from CA, FL, and NY, there are 17 states with a charge-off rate more than 15%. The company should implement an efficient methodology to thoroughly evaluate loan applicants from these states before approving loans.

•Even though majority of defaulters are in debt consolidation , considering Charge off percentage, small Business contribute the maximum which is 27.7%. Lending company should have proper measures in place for approving small business loans.

Multivariate analysis – Summary   Contd.

- The charge-off rate is highest for employees earning less than 40K, at 18.1%. Borrowers in lower income groups have the highest tendency to default on loans, and this tendency generally decreases as annual income increases.

- Rise in interest rate and increase in DTI will lead to higher default percentage. Before approving loans, the company should verify the open debts and assess the payment capability of loan applicants relative to their income.