

**Presentation  
on  
Lending Club  
Exploratory Data Analysis**

**Pavani Gangula**

**Naveen Kumar**

**10 MAY 2023**

# PURPOSE

Lending club analysis will help the company to make a decision for loan approval based on the applicant's profile. Which control loss of business to the company and avoid financial loss of the company.

# STEPS

❖ Data Understanding

❖ Data Cleaning

❖ Data Analysis

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

❖ Recommendations

# DATA UNDERSTANDING

- ❖ Import necessary libraries such as Numpy, Pandas, Matplotlib and Seaborn.
- ❖ Read the input file.
- ❖ Check the number of rows and columns in the data.
- ❖ Check the columns names in the data.

# DATA CLEANING

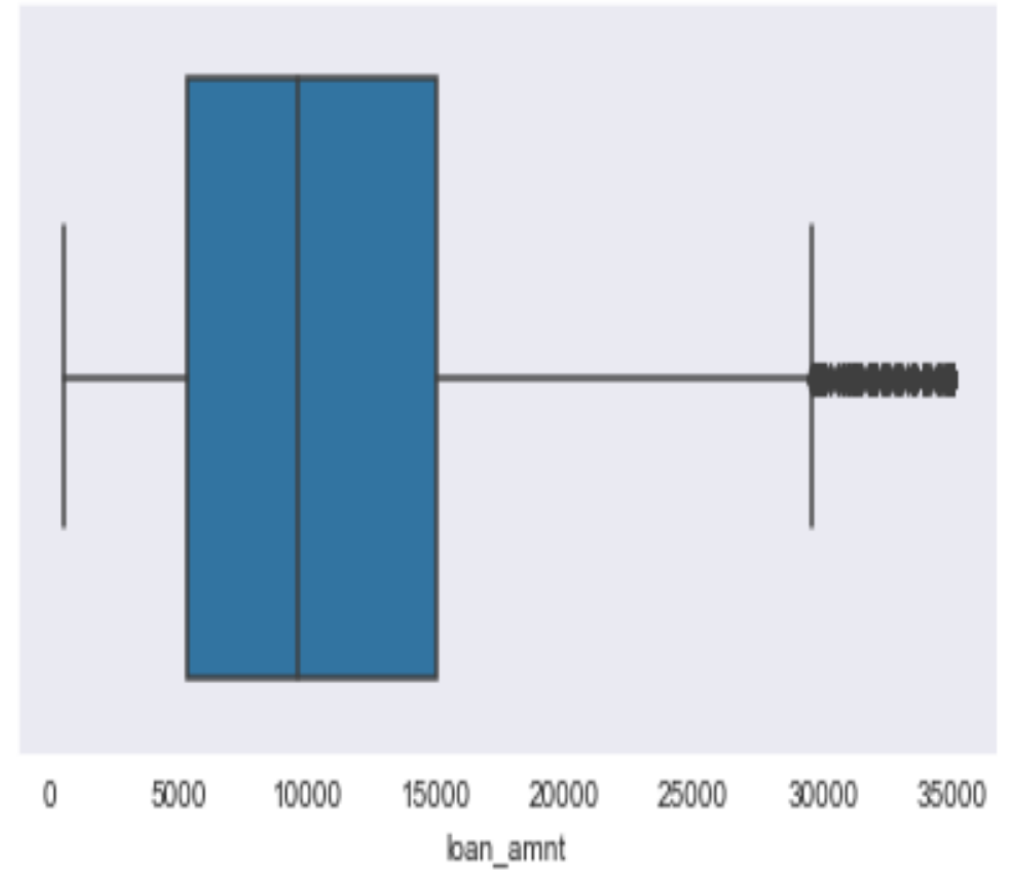
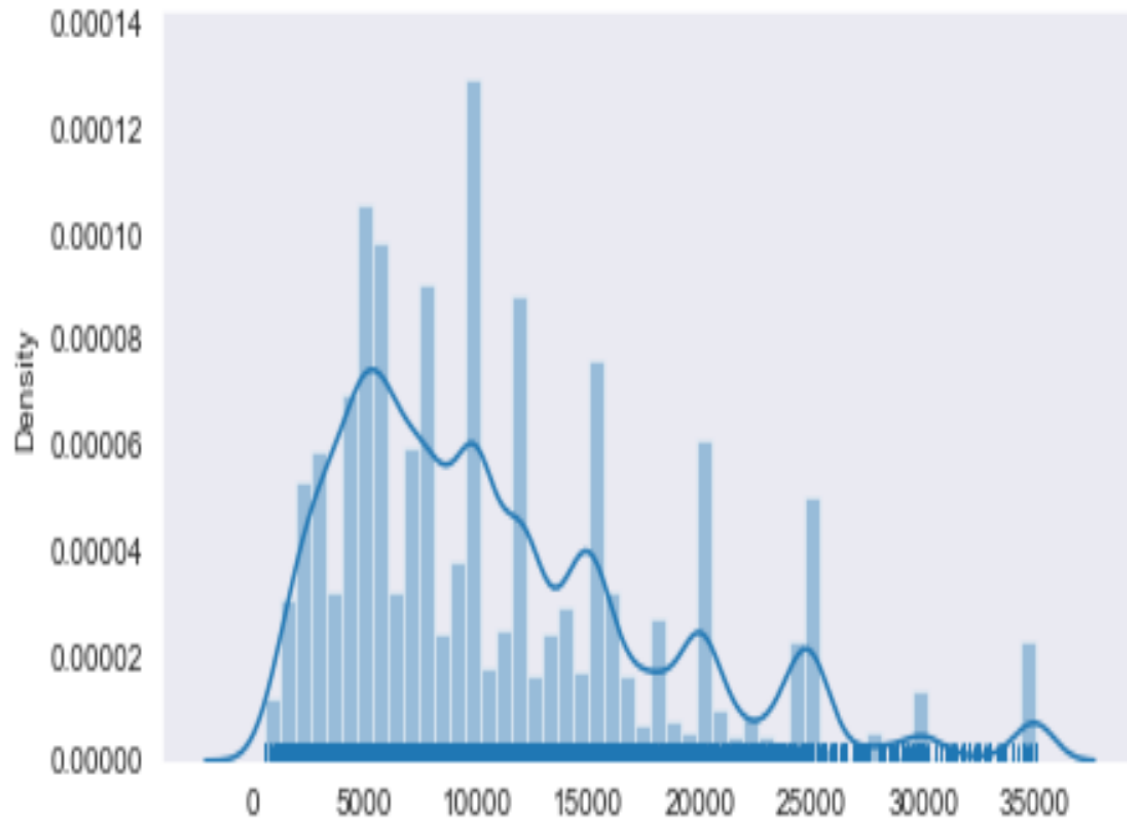
- ❖ Check the number of missing values in all columns.
- ❖ Dropping the columns with all null values.
- ❖ Checking percentage of missing data.
- ❖ Removing columns with greater than 30% missing values.
- ❖ Checking missing values in remaining columns.
- ❖ Checking distribution of "revol\_until"

# DATA CLEANING

- ❖ Checking the distribution of " emp\_length“
- ❖ Removing % symbol so that we can use the values for analysis.
- ❖ Converting the data to create categories.
- ❖ Creating bins for int\_rate,open\_acc,revol\_util,total\_acc for ease of working with the data.
- ❖ Checking the distribution of "loan\_status“
- ❖ Remove the rows with "current" as loan status because they cannot add any value to the analysis.

# UNIVARIATE ANALYSIS OF LOAN AMOUNT

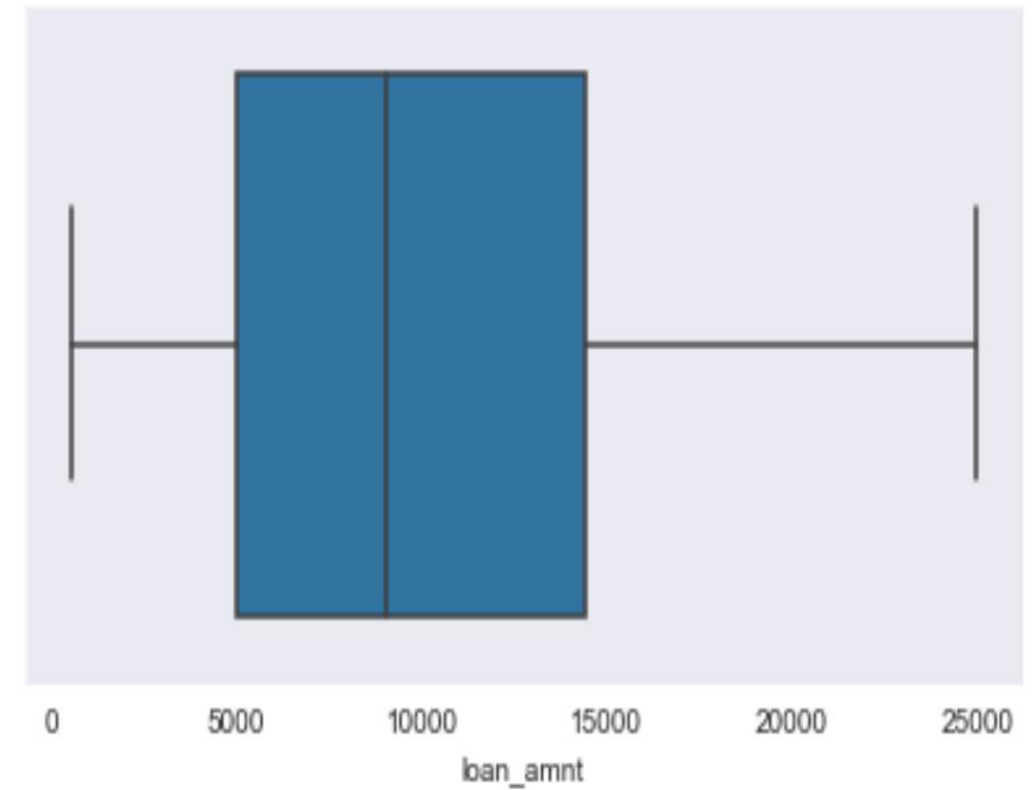
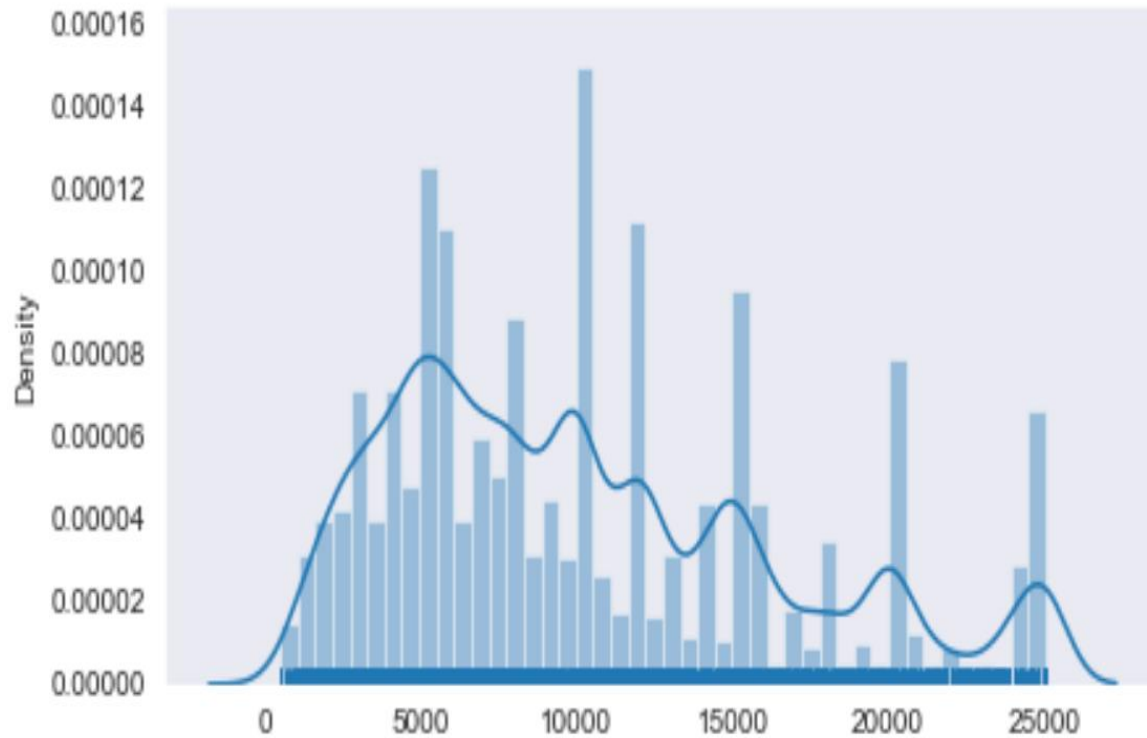
## With Outliers



There are many outliers in "loan\_amnt", so we need to remove them.

# UNIVARIATE ANALYSIS OF LOAN AMOUNT

## Without Outliers

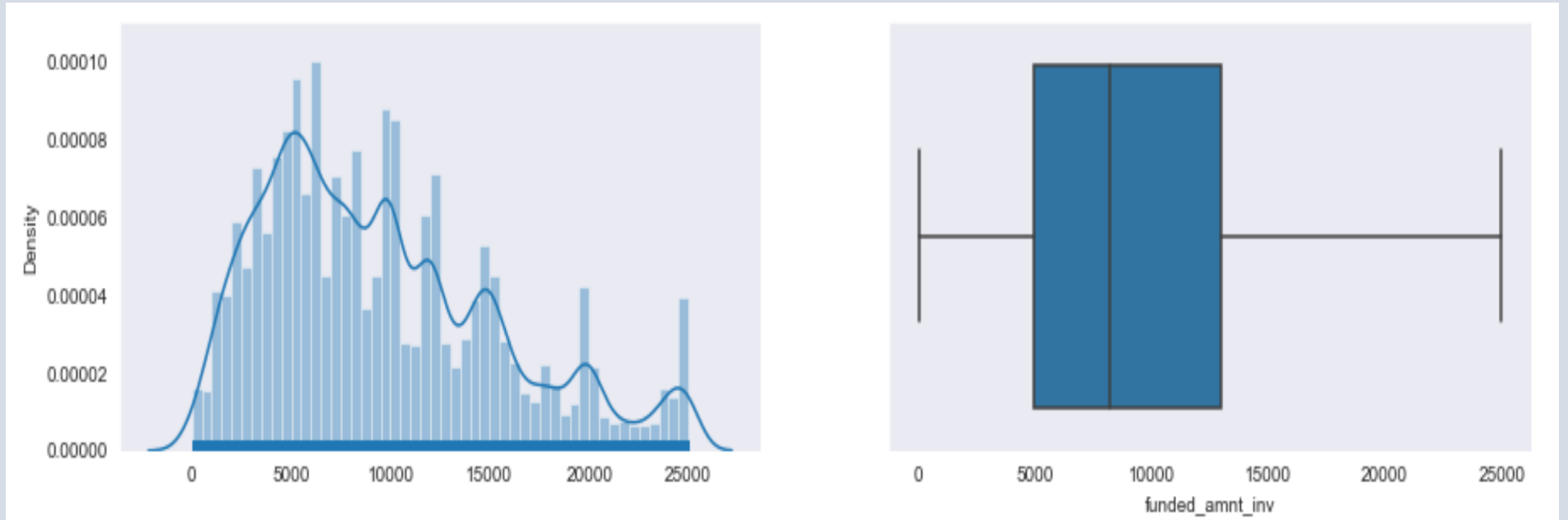


It is observed that the loan amount ranges between 0 - 25000 after removing outliers



# UNIVARIATE ANALYSIS OF FUNDED AMOUNT

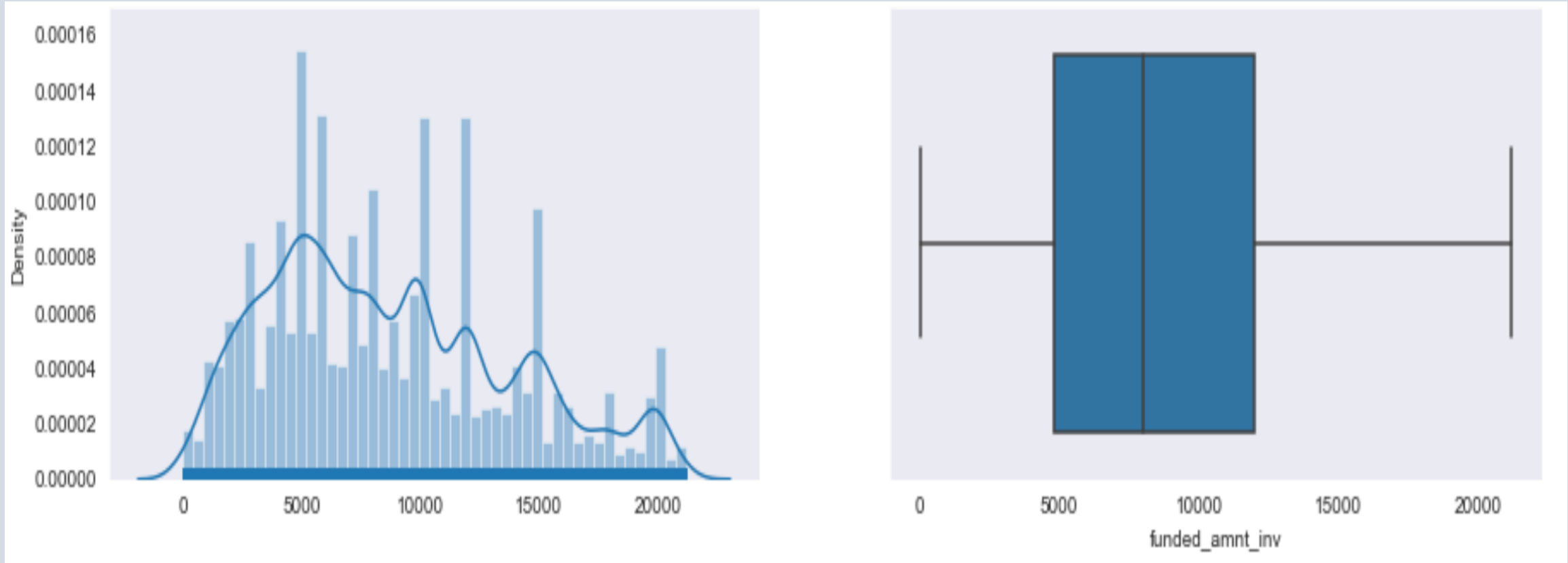
## With Outliers



There are many outliers in “funded amount”, so we need to remove them.

# UNIVARIATE ANALYSIS OF FUNDED AMOUNT

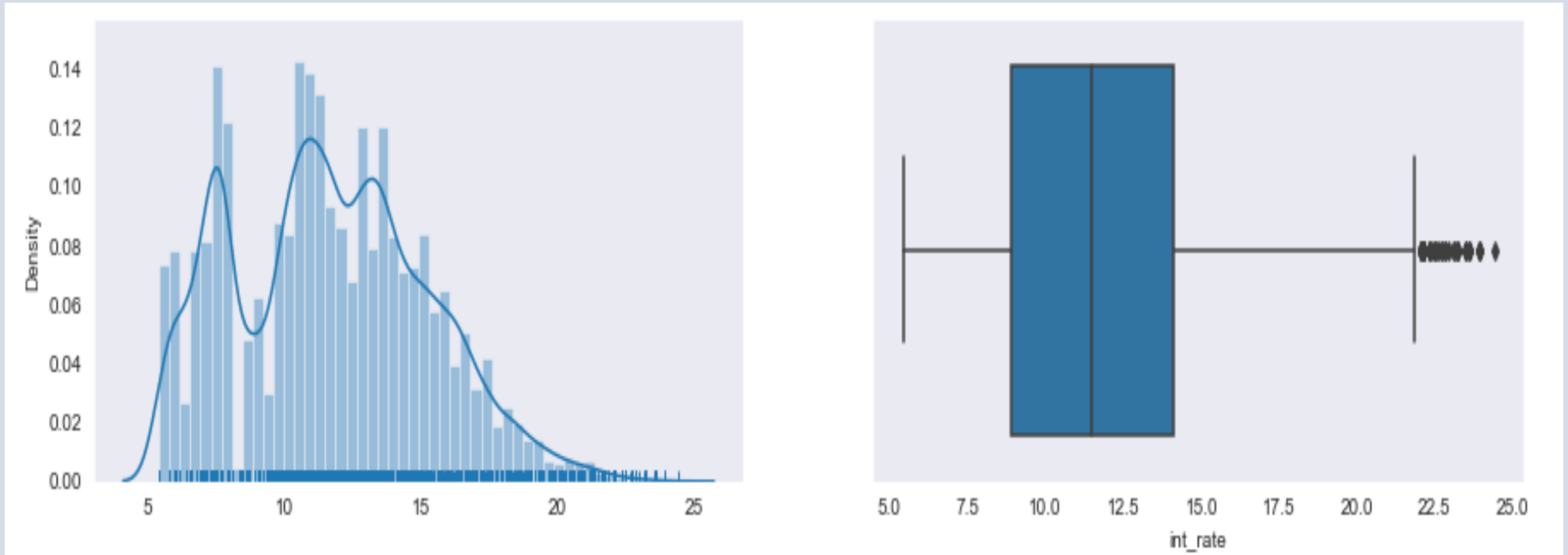
## Without Outliers



It is observed that the funded amount by investors ranges between 0 - 20000 after removing outliers.

# UNIVARIATE ANALYSIS OF INTEREST RATE

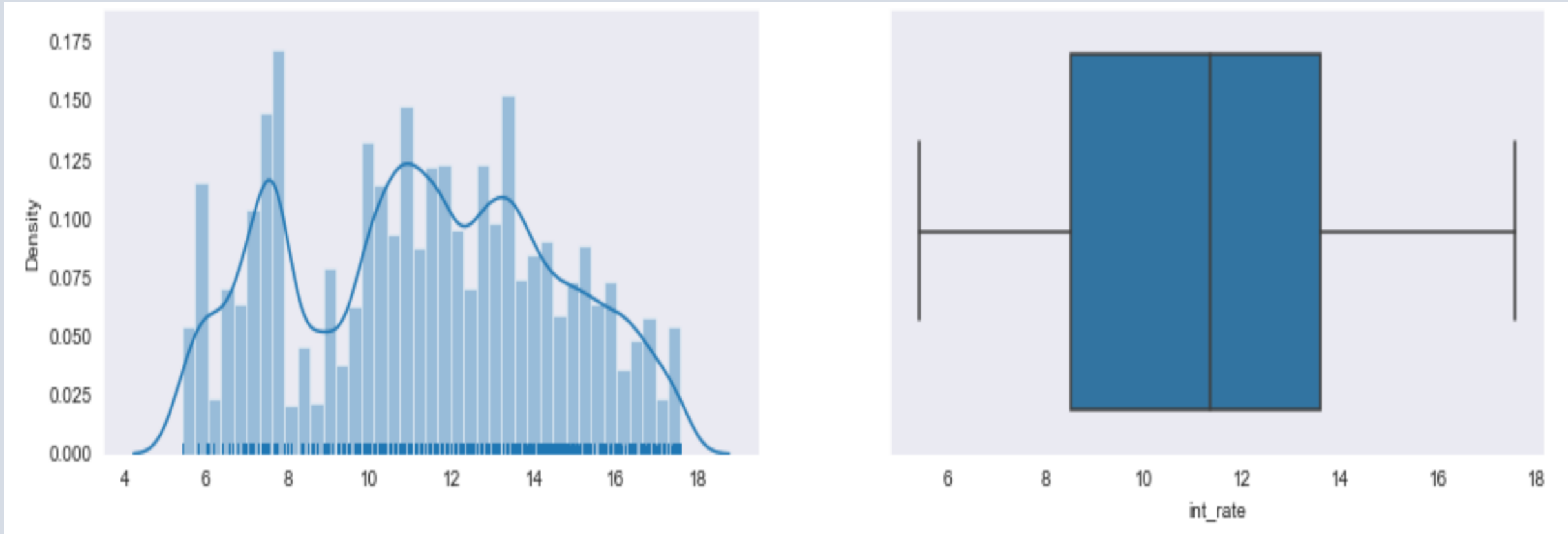
## With Outliers



There are many outliers in “Interest rate”, so we need to remove them.

# UNIVARIATE ANALYSIS OF INTEREST RATE

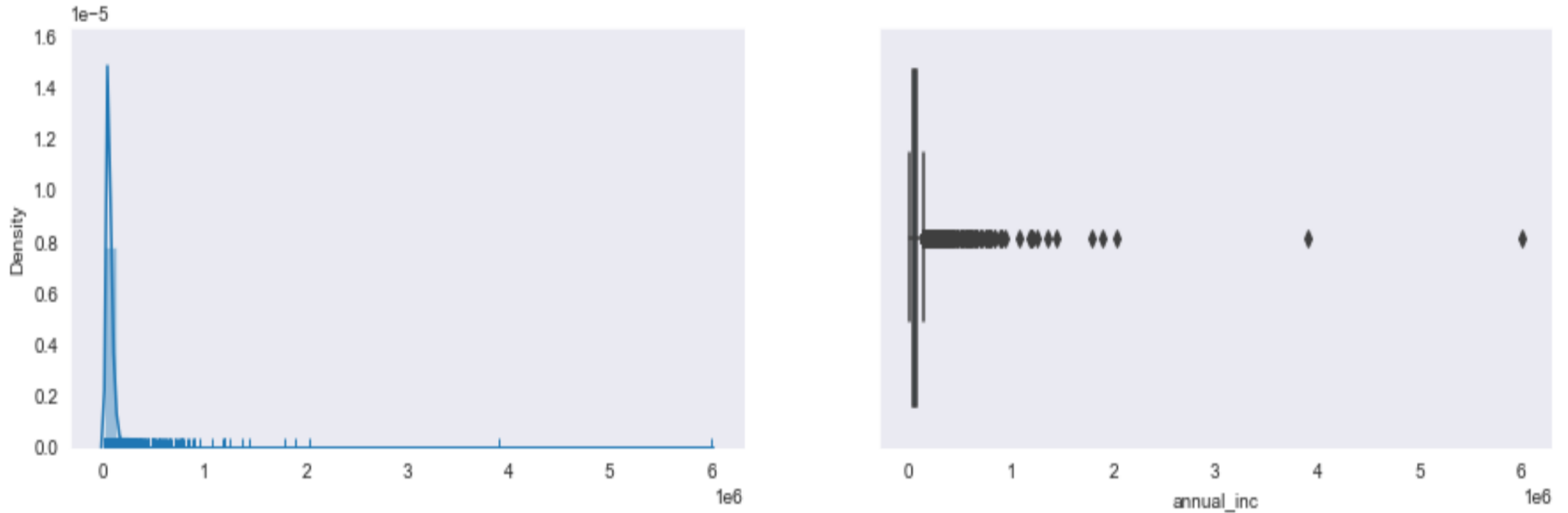
## Without Outliers



It is observed that the interest rates ranges from 4-18% after removing outliers.

# UNIVARIATE ANALYSIS OF ANUAL INCOME

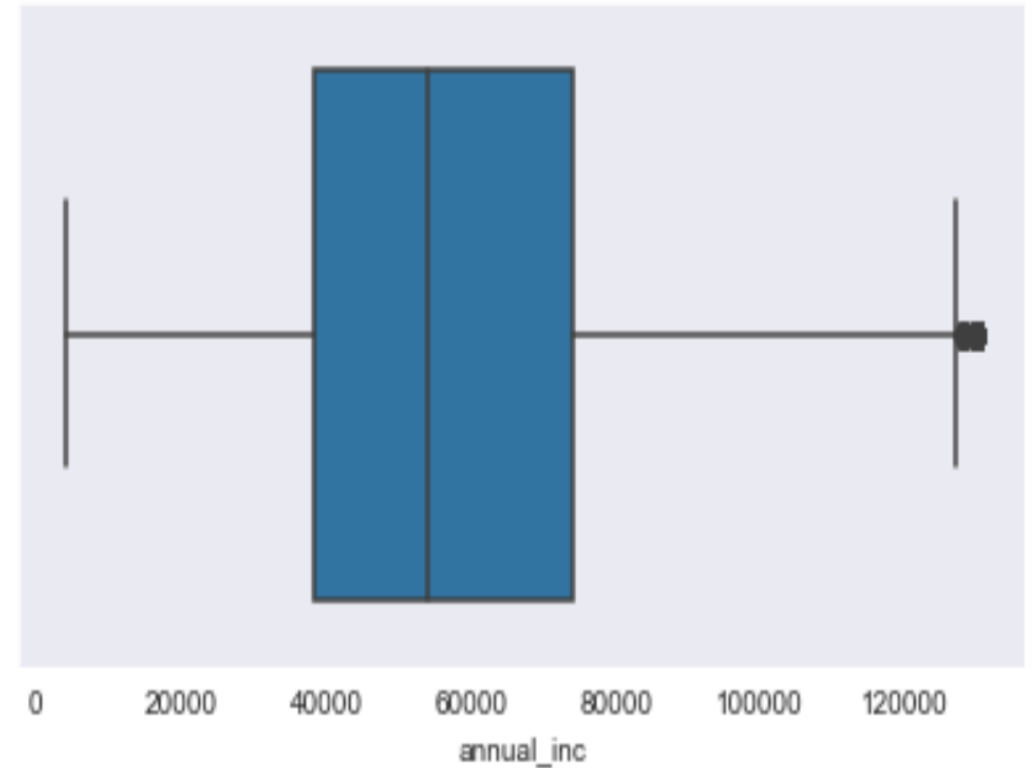
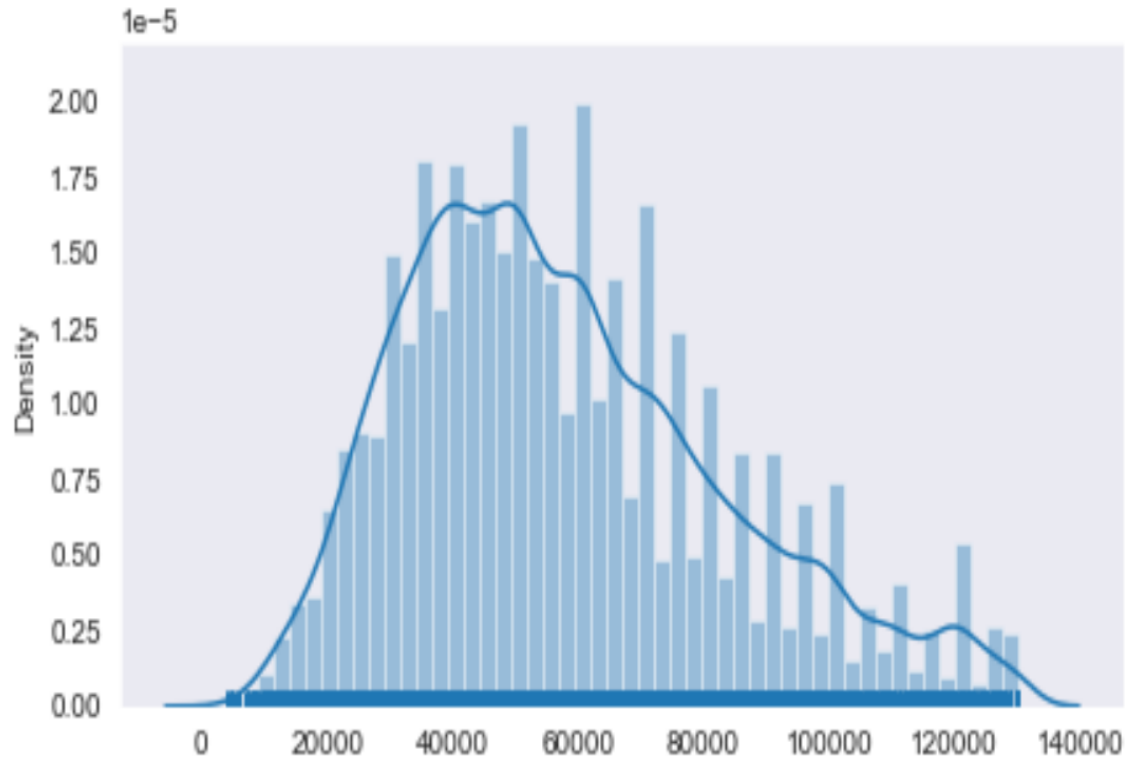
## With Outliers



There are many outliers in “annual income”, so we need to remove them.

# UNIVARIATE ANALYSIS OF ANUAL INCOME

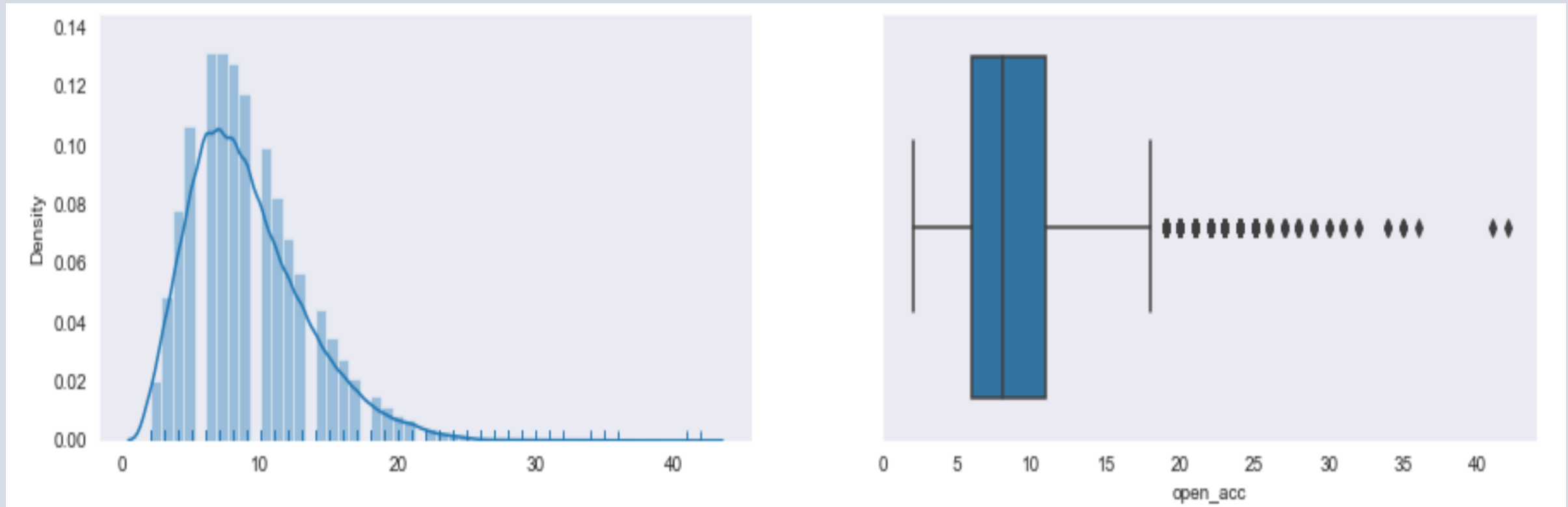
## Without Outliers



Annual income has many outliers in the data. After removing the values considering 95th percentile, it ranges from 0 to 140,000.

# UNIVARIATE ANALYSIS OF OPEN ACCOUNT

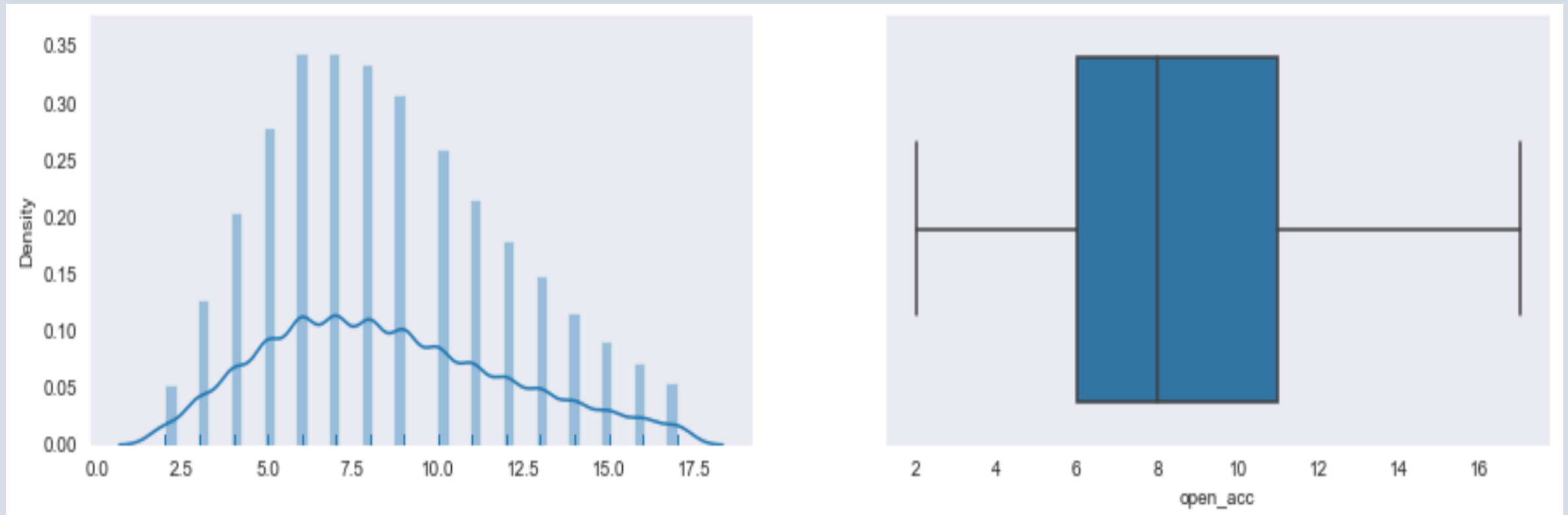
## With Outliers



There are many outliers in “open account”, so we need to remove them.

# UNIVARIATE ANALYSIS OF OPEN ACCOUNT

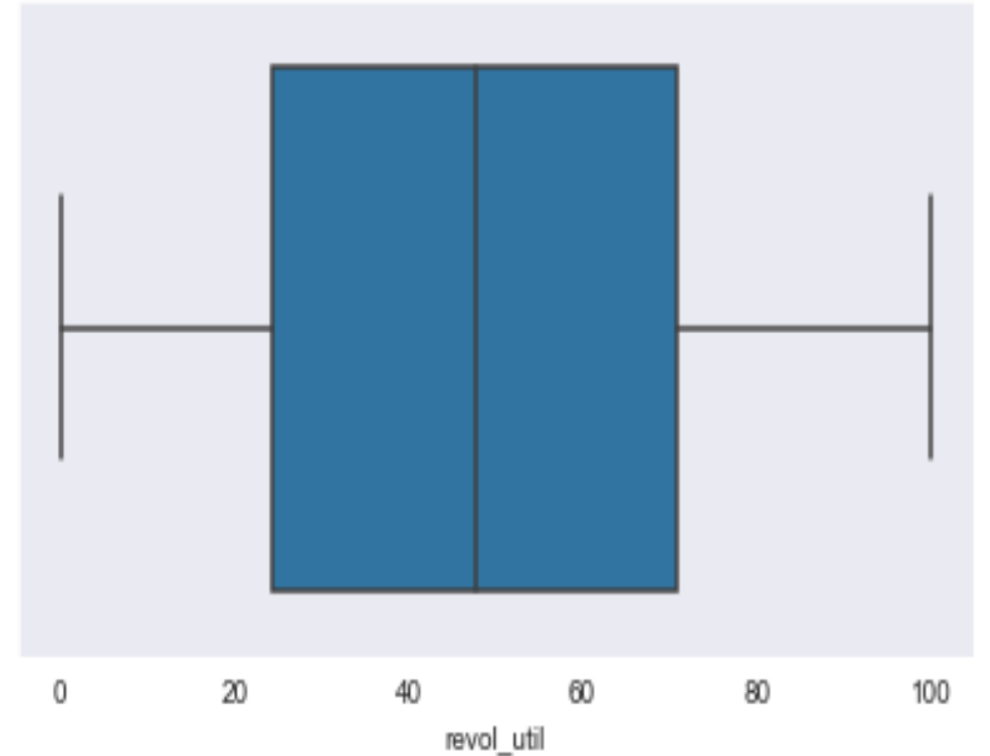
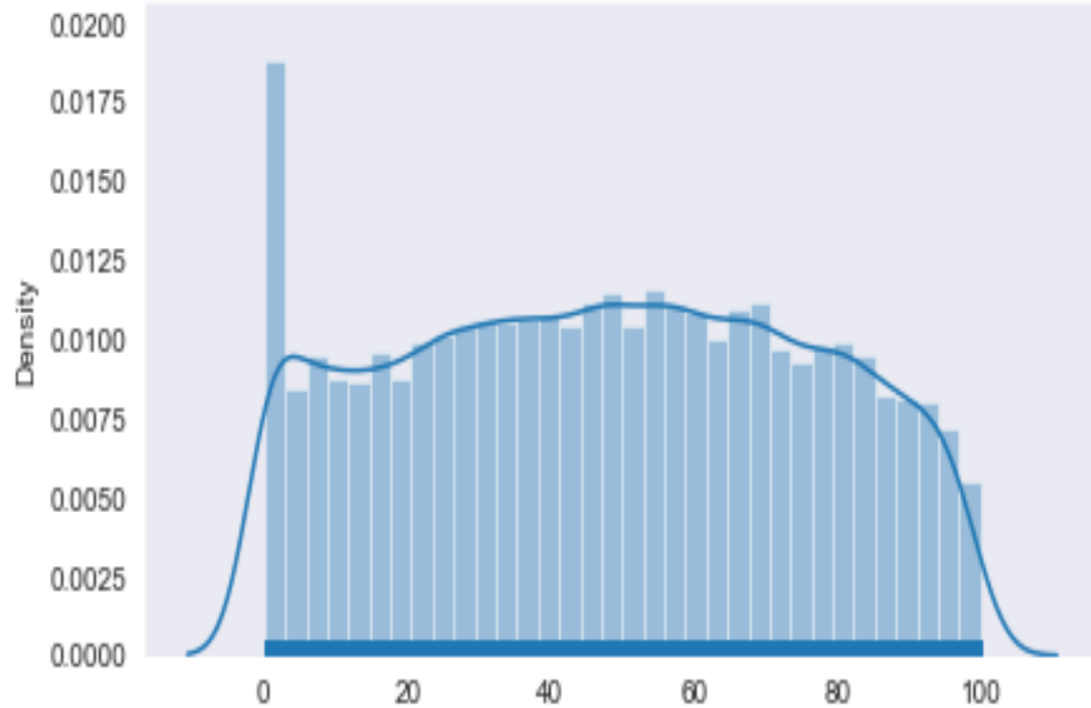
## Without Outliers



It is observed that the column "open\_acc" ranges from 2 to 17.5 after removing outliers.



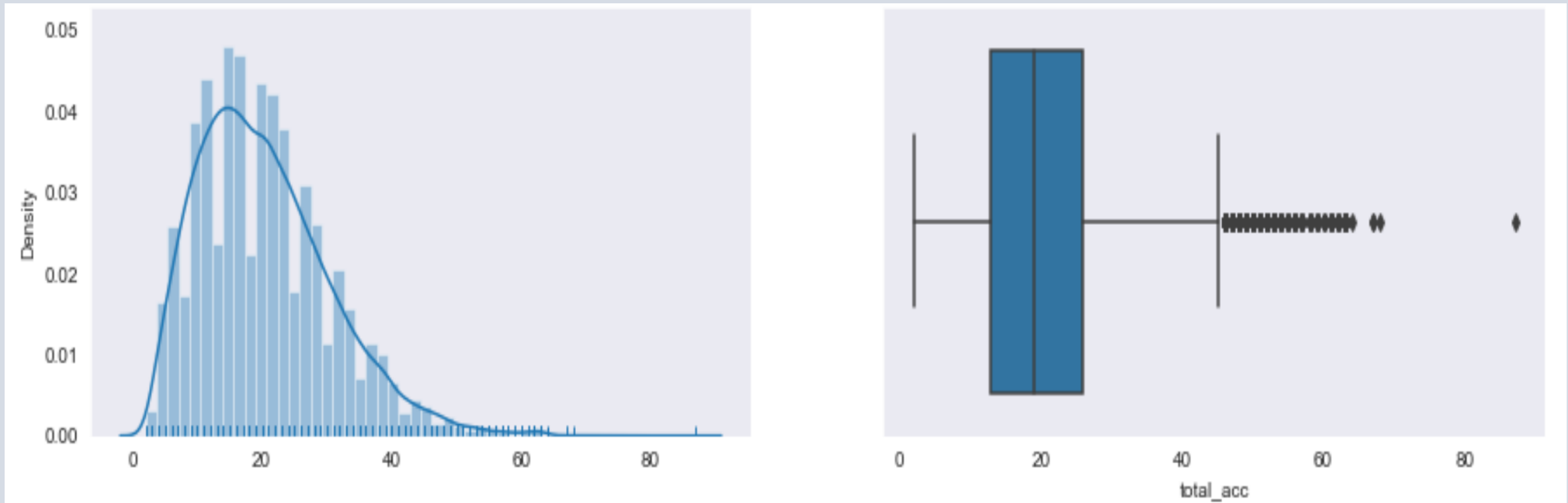
# UNIVARIATE ANALYSIS OF REVOL\_UTIL



The column "revol\_util" has no outliers so, we can consider the data as is for the analysis:

# UNIVARIATE ANALYSIS OF TOTAL ACCOUNT

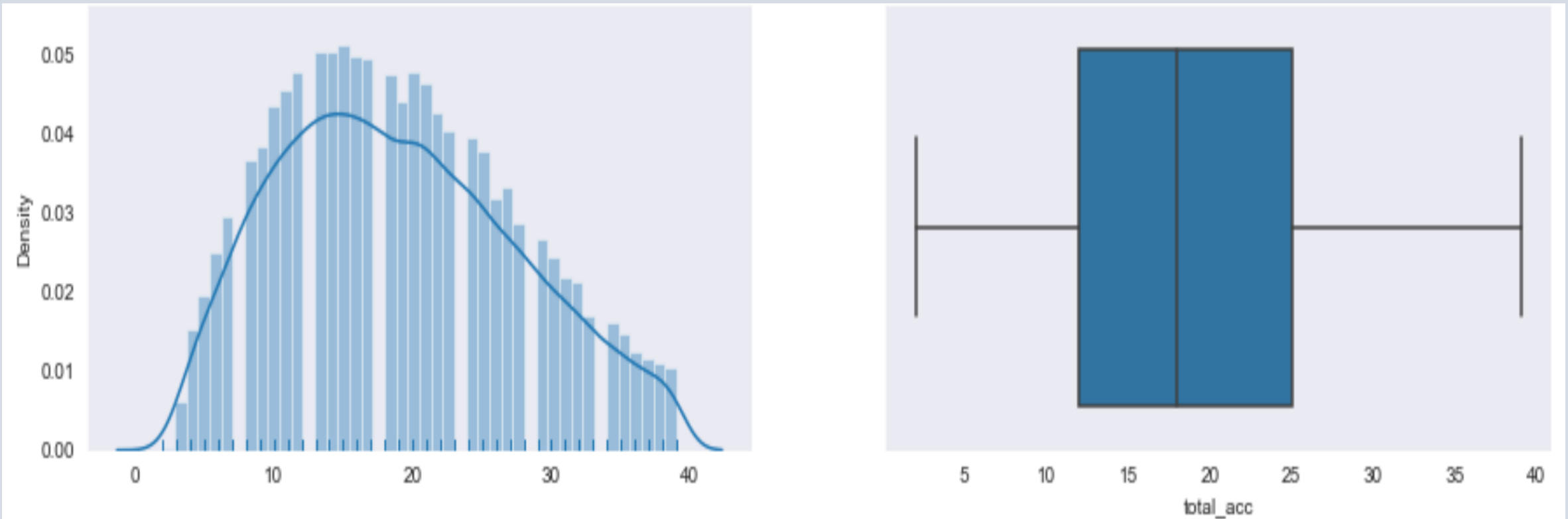
## With Outliers



There are many outliers in “total account”, so we need to remove them.

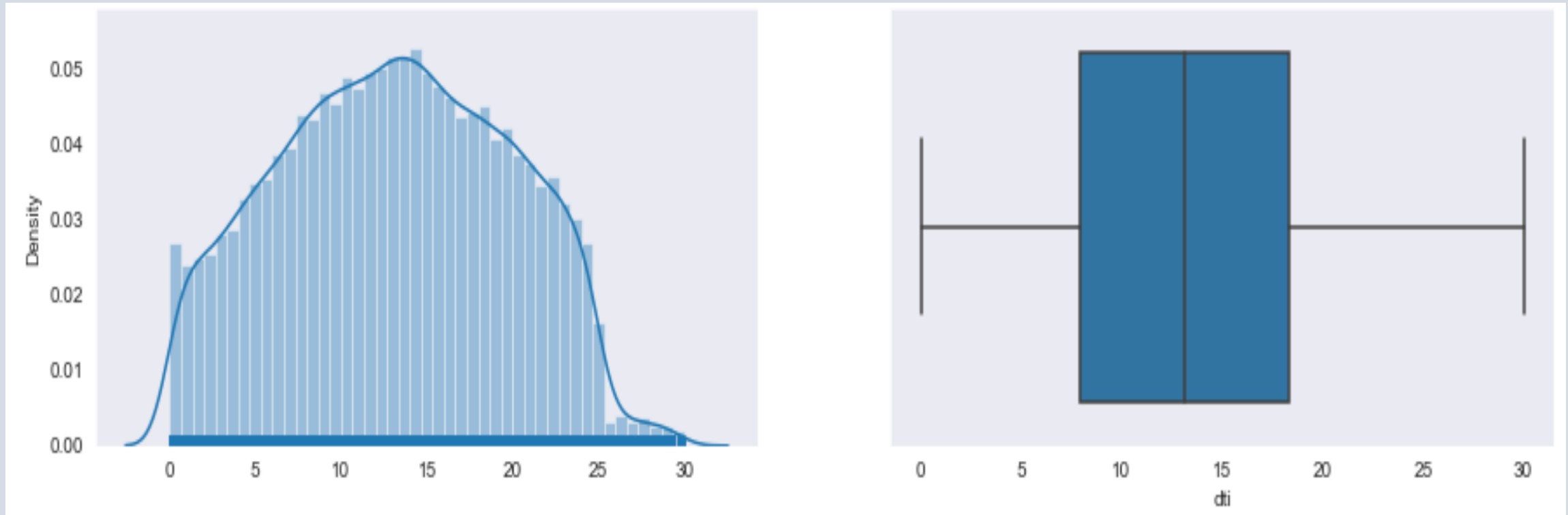
# UNIVARIATE ANALYSIS OF TOTAL ACCOUNT

**Without Outliers**



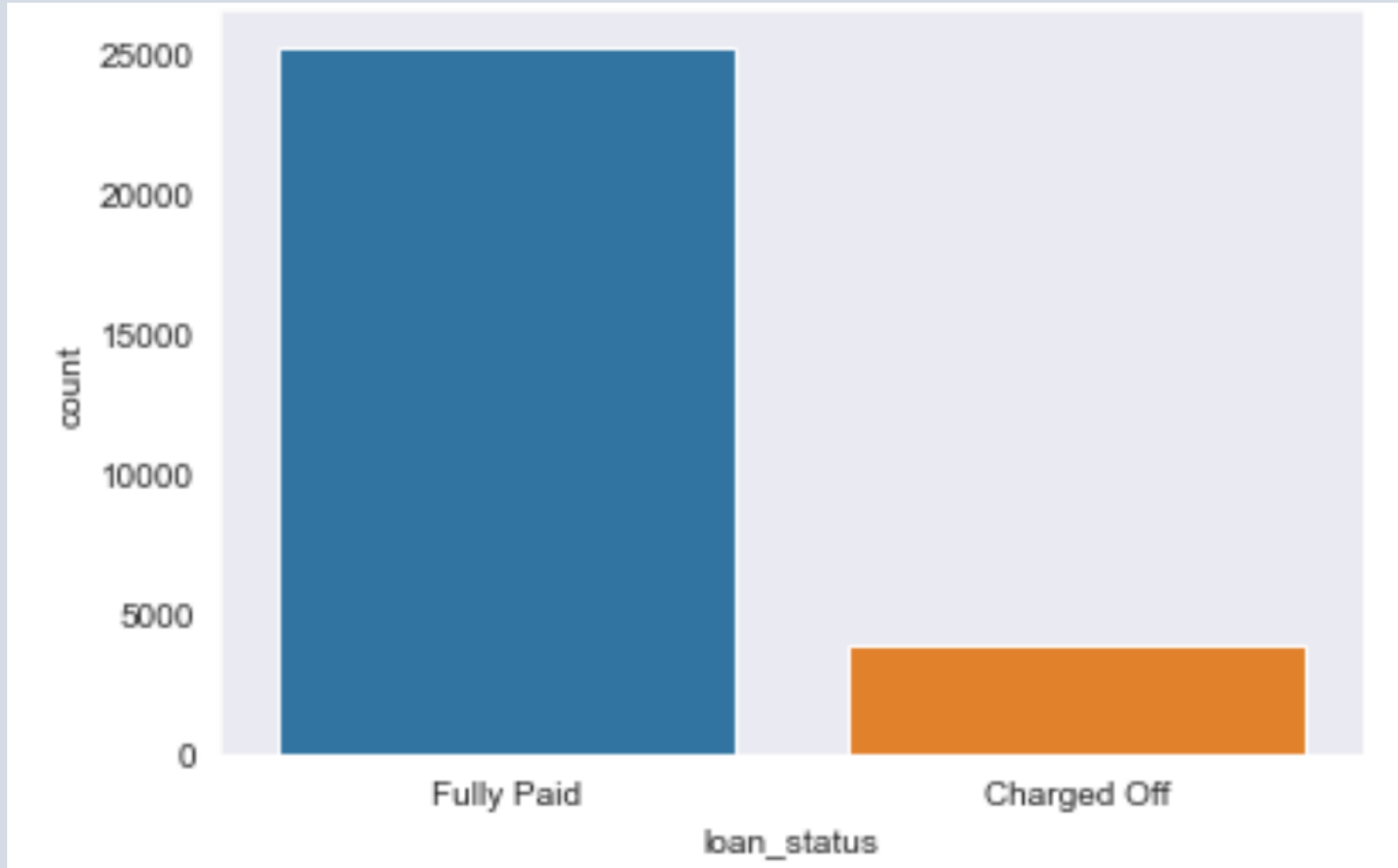
"total\_acc" has many outliers in the data. After removing them, the values range from 0 to 40.

# UNIVARIATE ANALYSIS OF DEBT TO INCOME RATIO



"dti" or popularly known as "Debt to Income Ratio" does not have any outliers and ranges from 0-30.

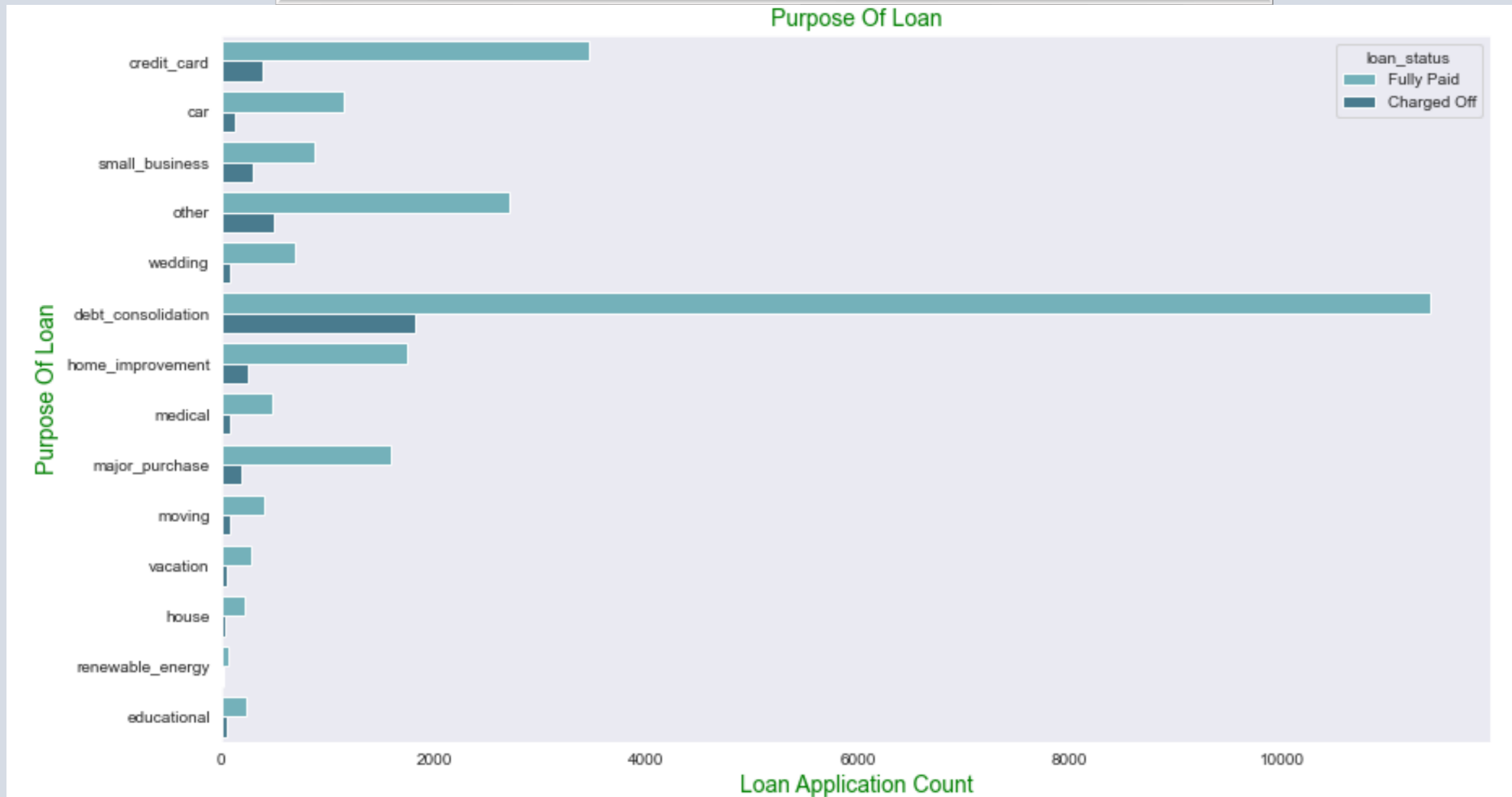
# DISTRIBUTION OF LOAN STATUS



It is observed that the majority class in the dependent column i.e "loan\_status" is "Fully Paid".

# BIVARIATE ANALYSIS

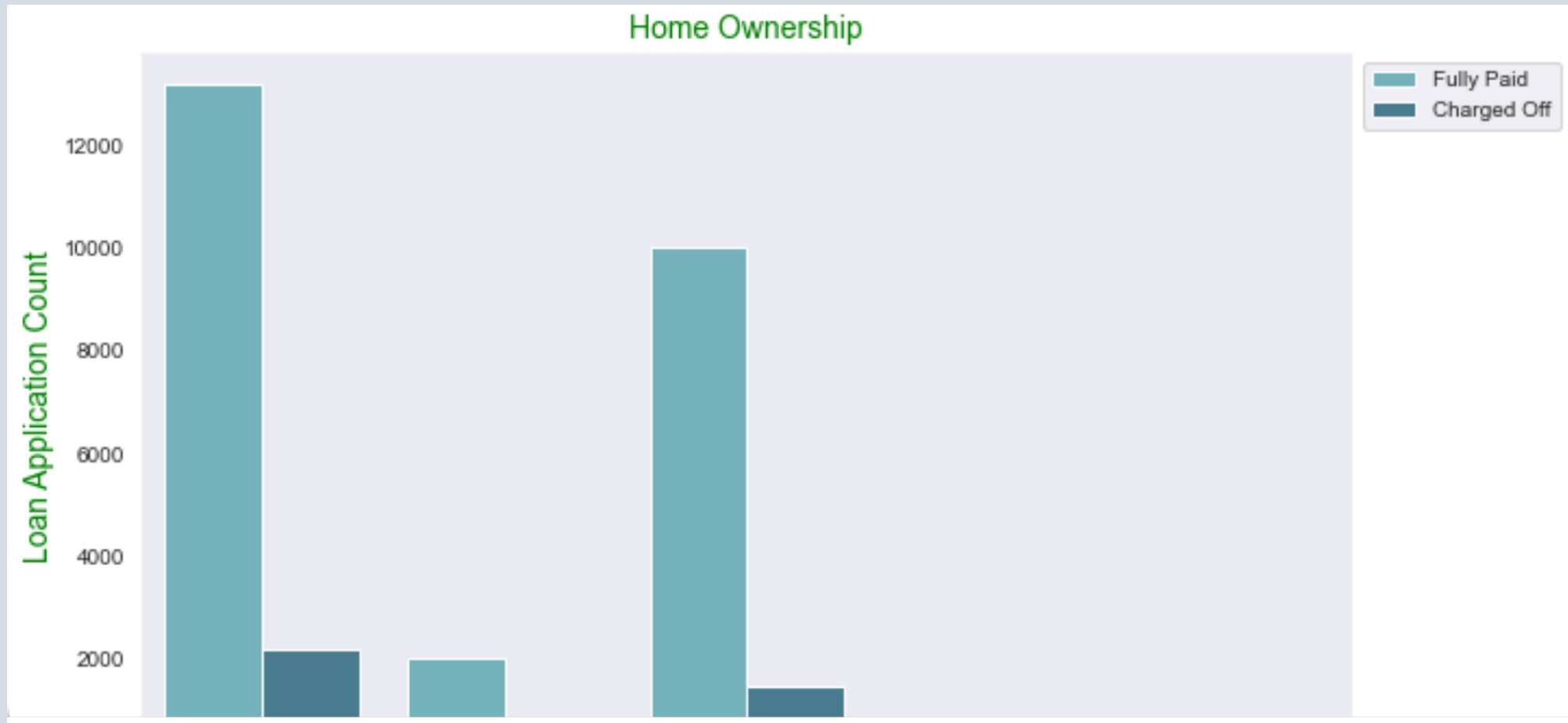
## Purpose of loan vs loan application count



From the plot above, it can be inferred that the "Fully paid" loans are highest for "debt\_consolidation" followed by "credit\_card". The "Charged off" loans are also highest for "debt\_consolidation", followed by "other".

# BIVARIATE ANALYSIS

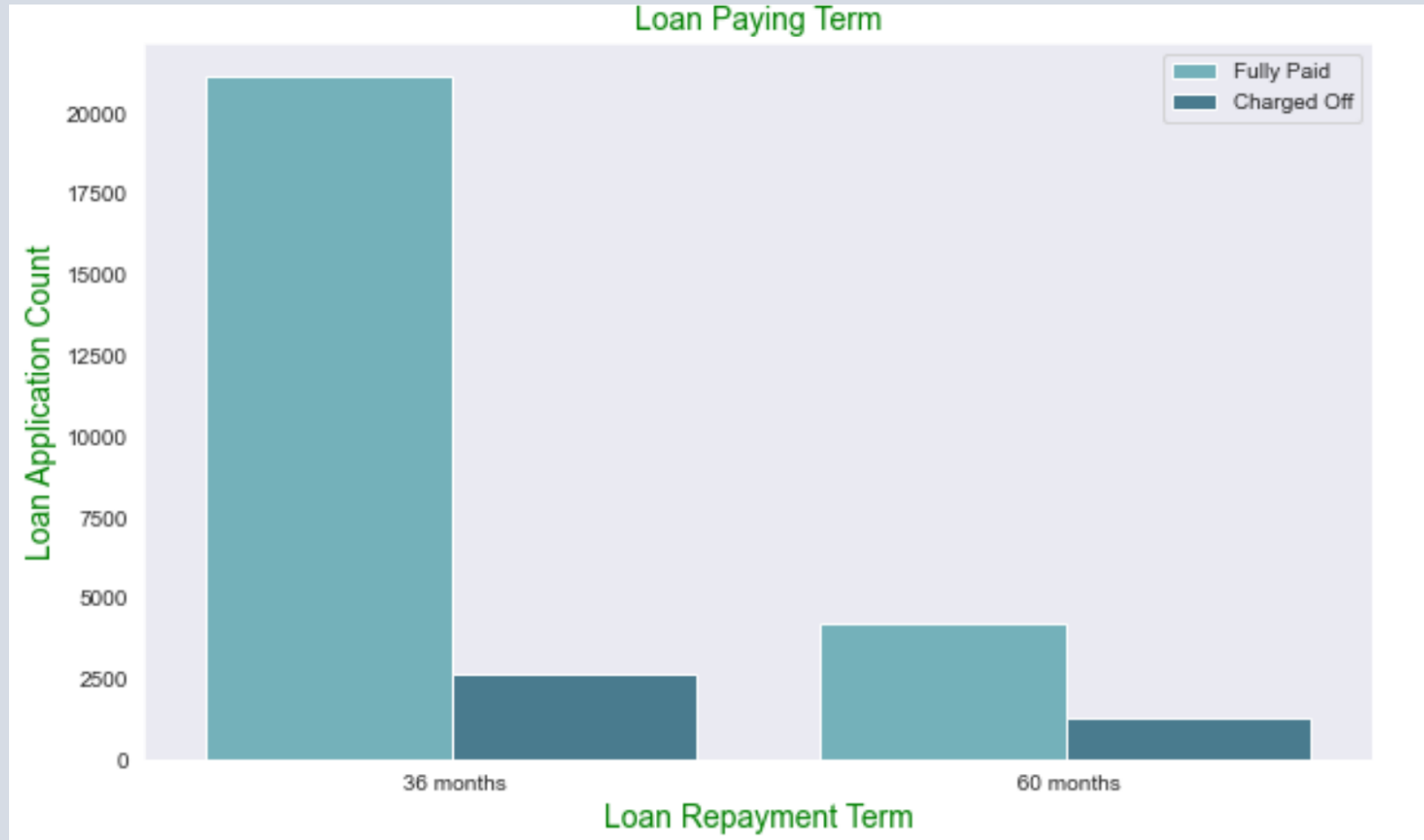
## Home ownership vs loan application count



From the above plot, we can observe that the number of "Fully Paid" loans are high for "Rent" and "Mortgage". The "Charged off" loans are also high for these two categories.

# BIVARIATE ANALYSIS

## Loan paying term vs loan application count

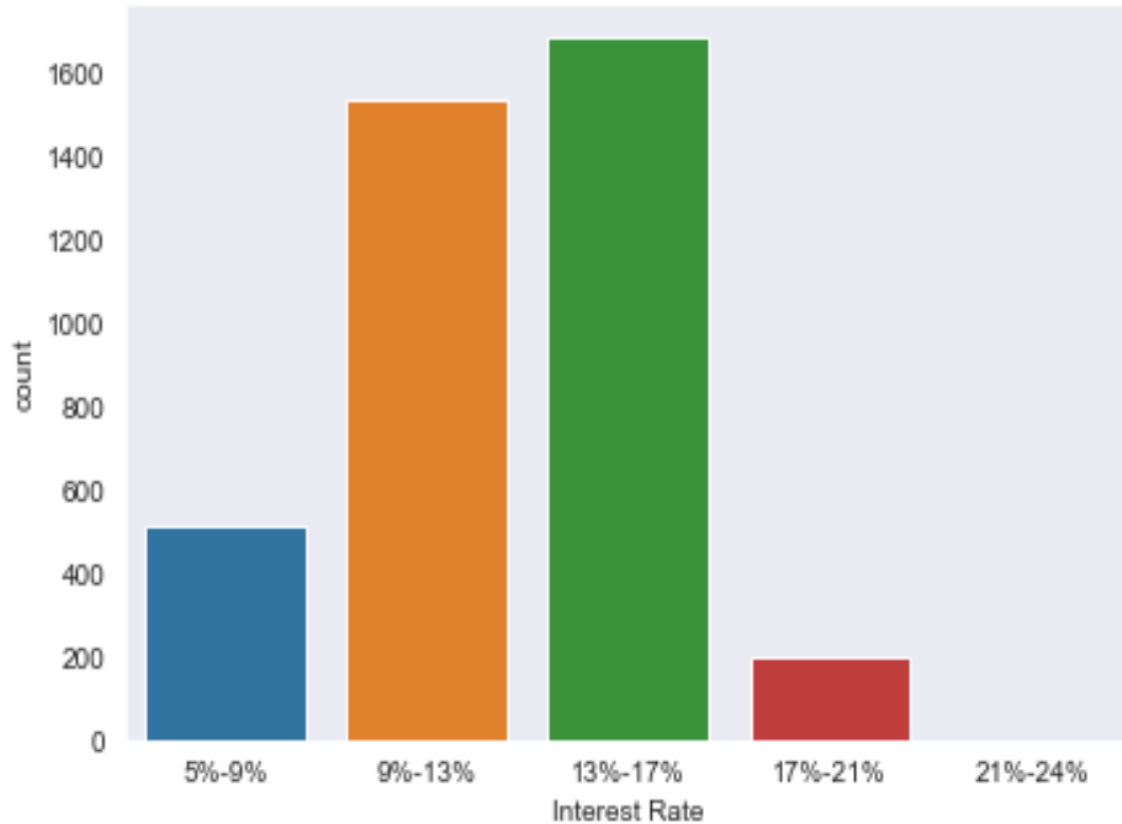


It can be observed from the above plot that the "Charged off" loans are high for "36 months" tenure when compared to "60 months".

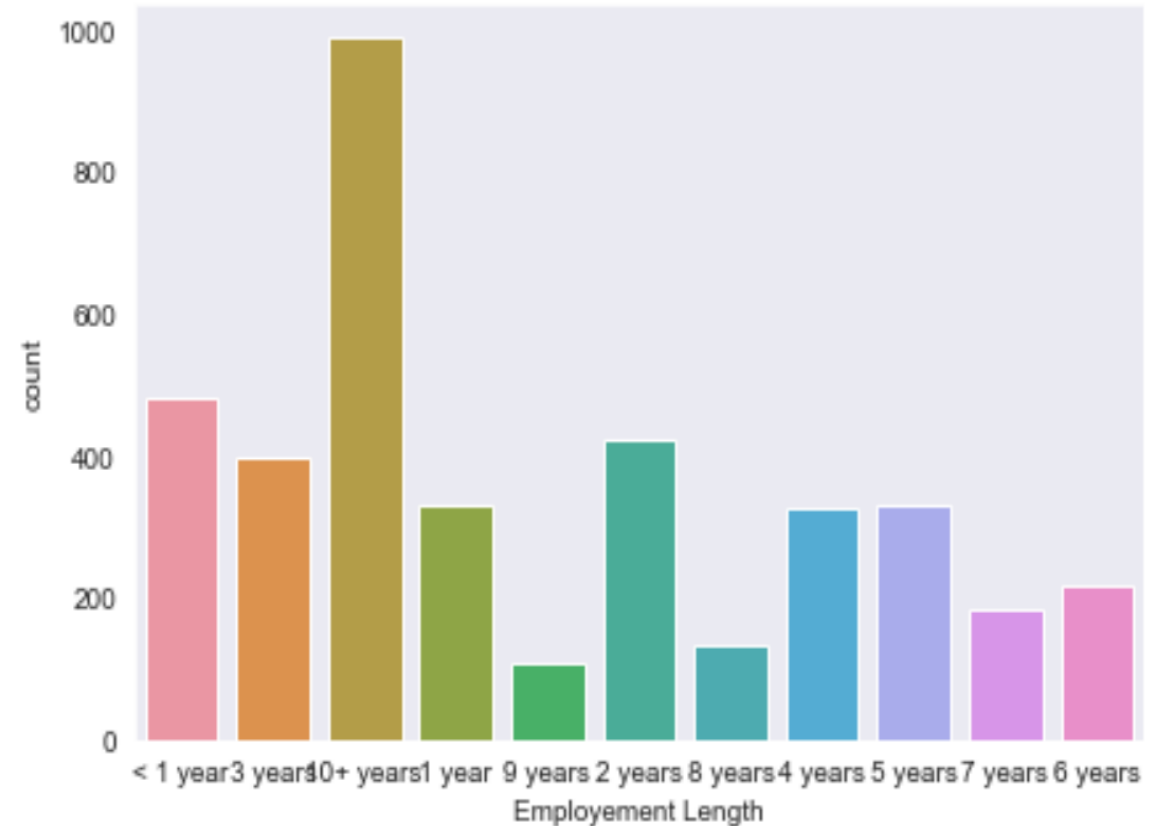


# BIVARIATE ANALYSIS

Interest rate vs loan application count



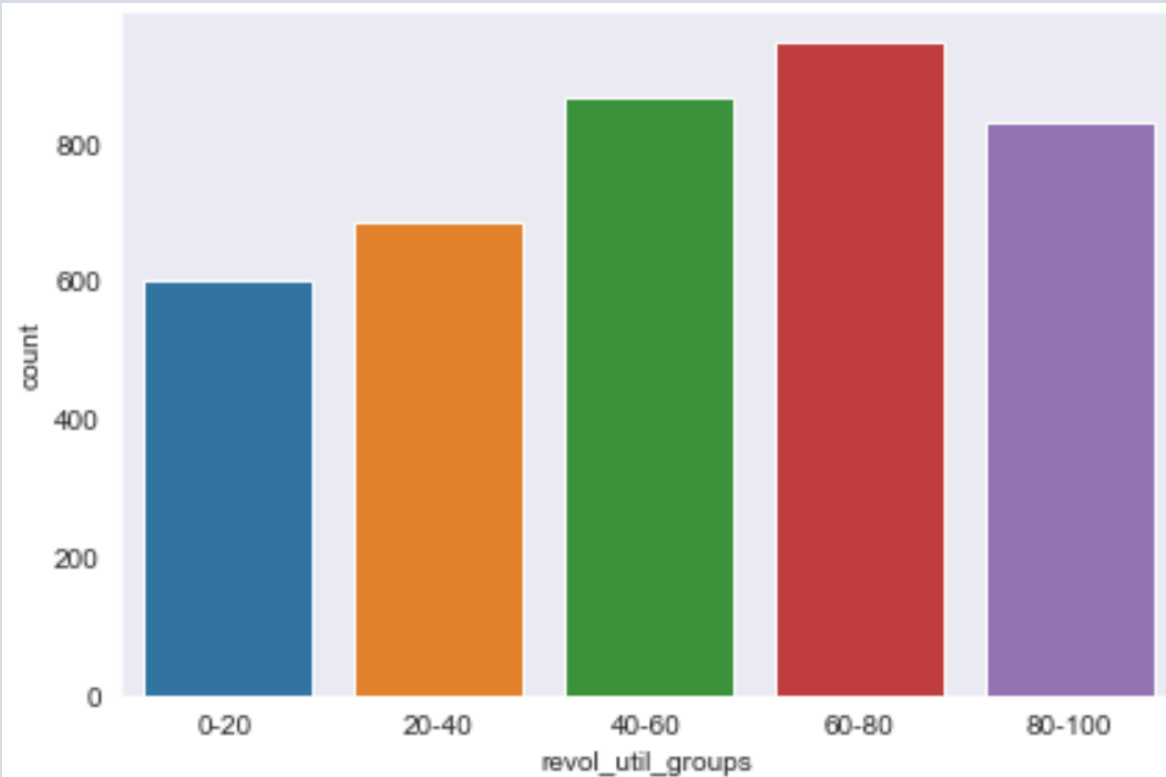
Employment length vs loan application count



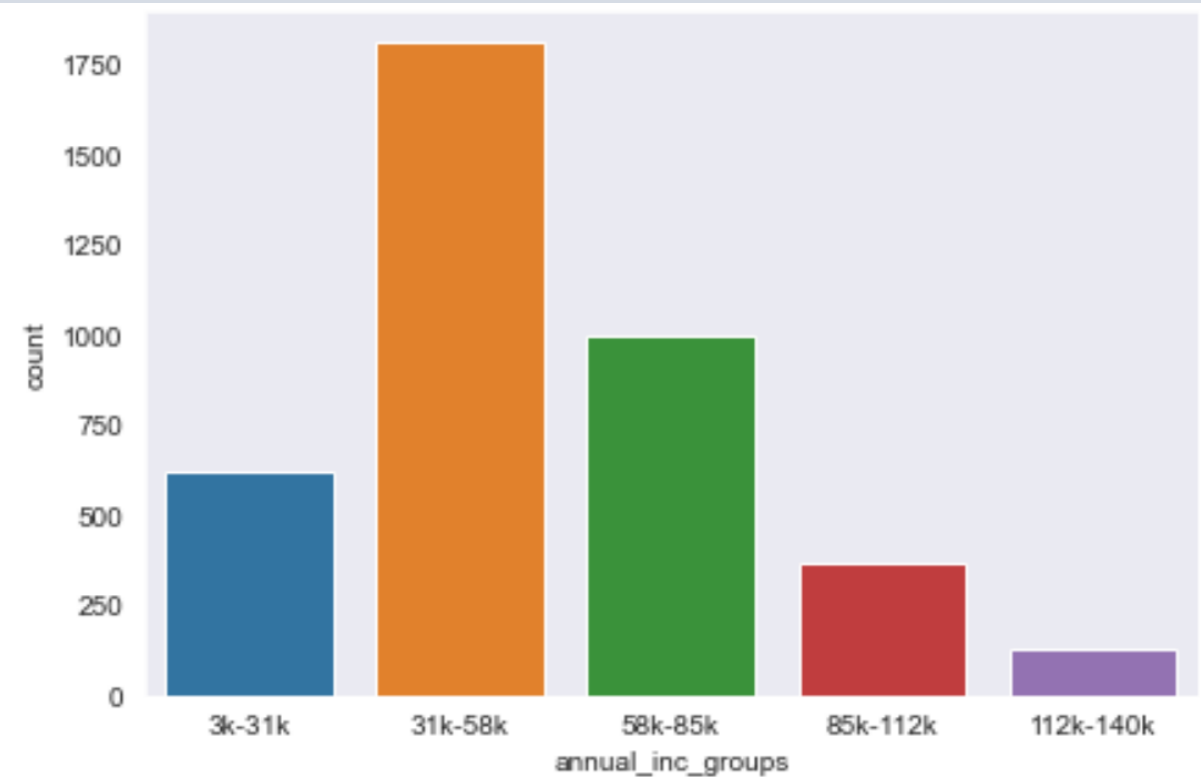
From the above two plots, it can be inferred that the "Charged Off" loans are higher with interest rates in the range 13%-17%, followed by 9%-13%. For employment length, the number of "Charged off" loans are highest for applicants with 10+ years of experience.

# BIVARIATE ANALYSIS

**Revol\_util vs loan application count**



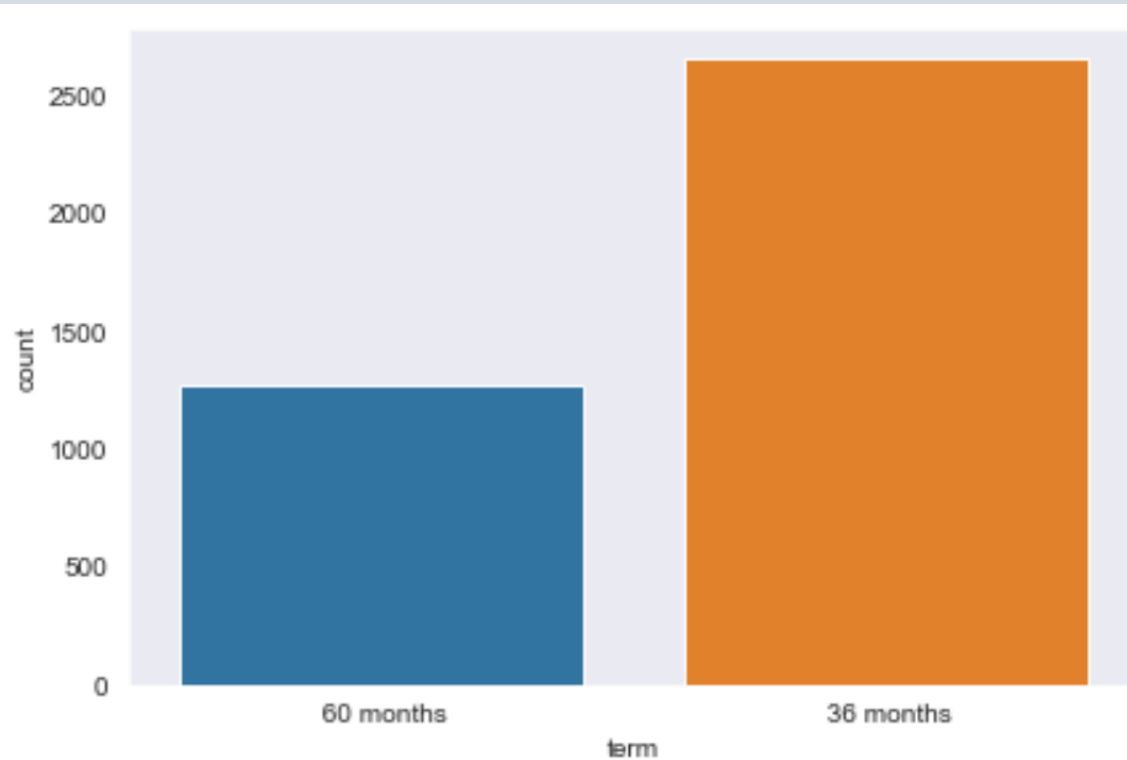
**Annual income vs loan application count**



It is observed that the most "Charged Off" loans are in the range 60-80, followed by 40-60. And the most "Charged Off" loans are in the annual income range of "31k-58k".

# BIVARIATE ANALYSIS

**Term vs loan application count**



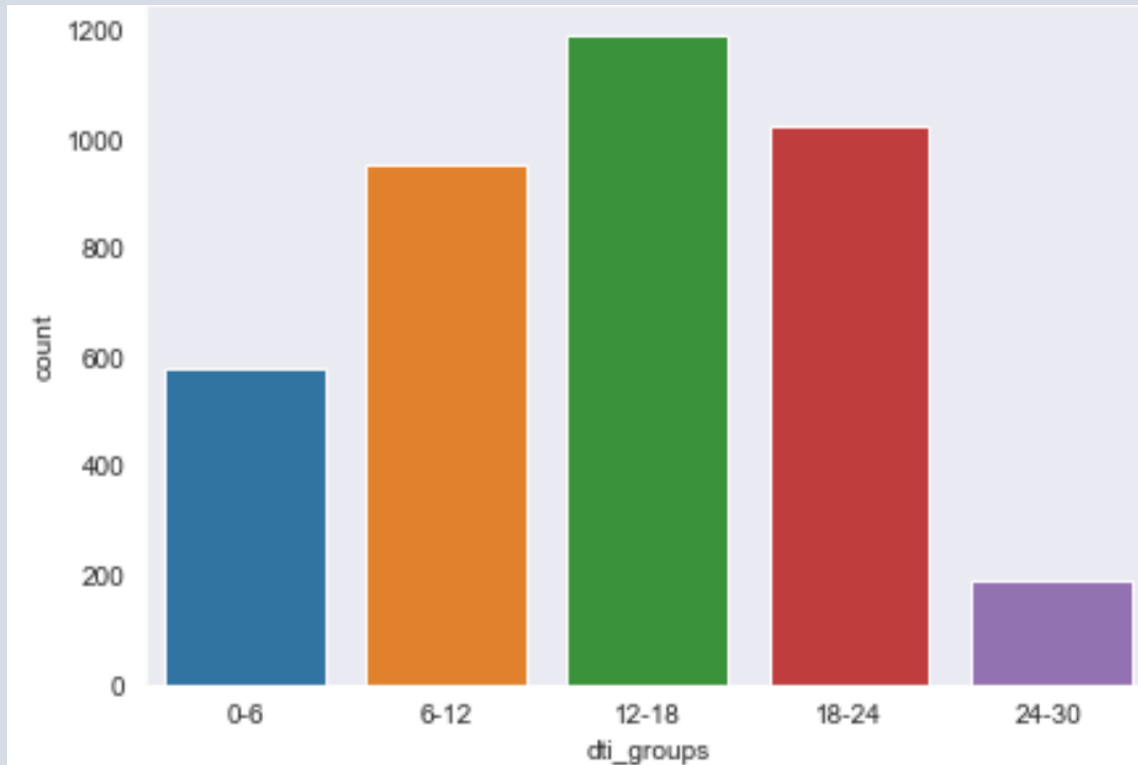
**Verification status vs loan application count**



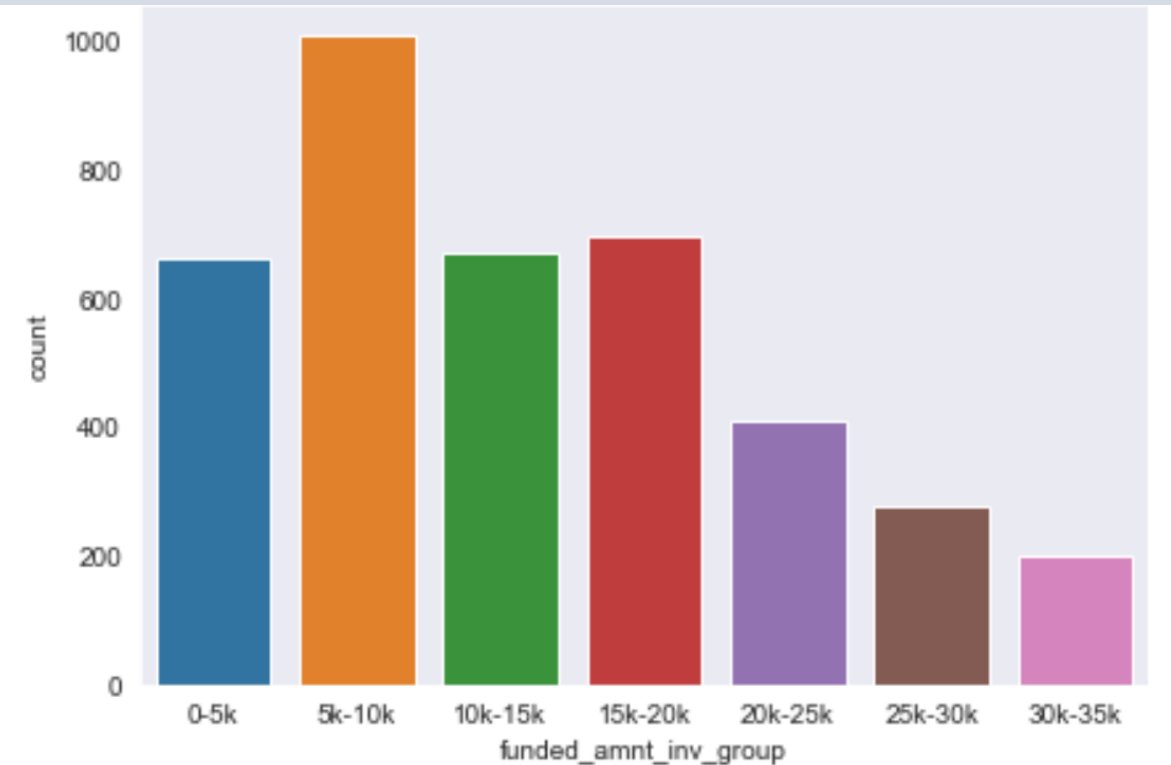
It is observed that the "Charged Off" loans are highest for "36 months". For "verification\_status", the "Charged off" loans are highest for "Not Verified" category.

# BIVARIATE ANALYSIS

**Dti group vs loan application count**



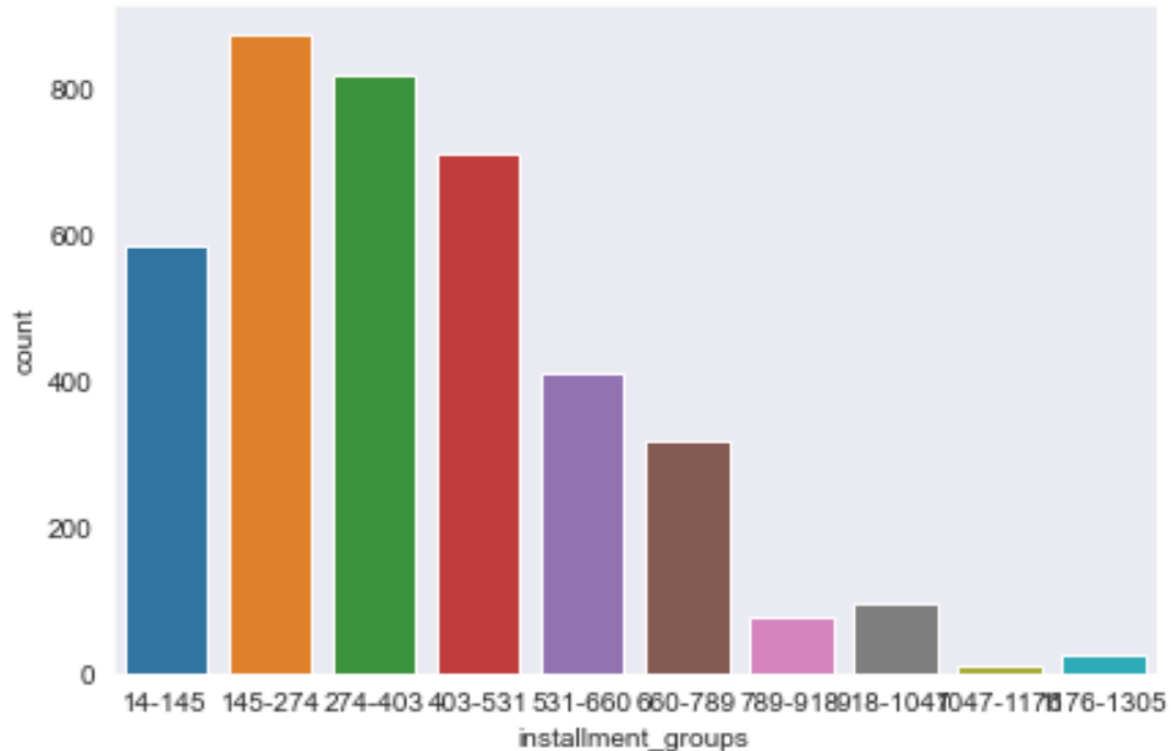
**Funded amount vs loan application count**



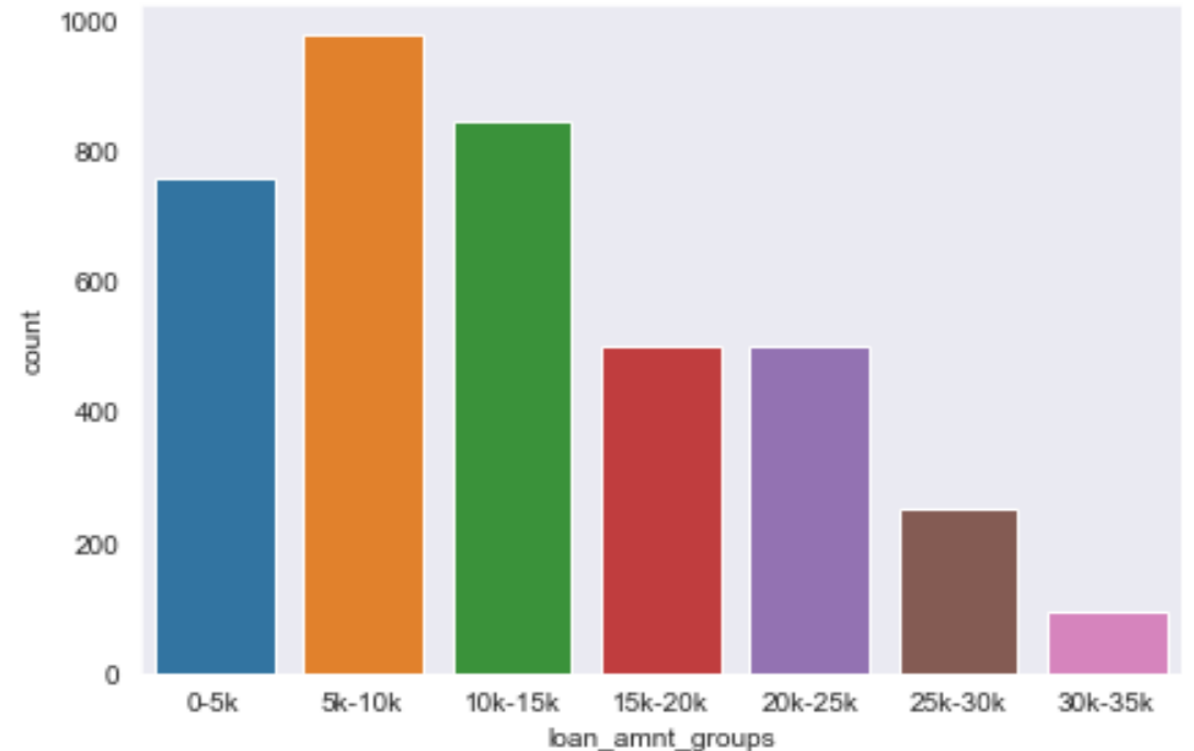
It is observed that the "Charged Off" loans are highest for a dti (Debt-to-Income Ratio) of 12-18, followed by 18-24. The highest number of "Charged Off" loans are in the funded amount range of 5k-10k.

# BIVARIATE ANALYSIS

**Installment group vs loan application count**



**Loan amount vs loan application count**



It is observed that the "Charged Off" loans are highest for installments 145-274. The highest number of "Charged Off" loans are in the loan amount range of 5k-10k.

# CORRELATION MATRIX – HEATMAP



**Following are list of features with high correlation (>0.6)**

1. Funded\_amnt\_int and loan\_amnt
2. Installment and loan\_amnt
3. Installemnt and funded\_amnt\_inv
4. Open\_acc and total\_acc

# RECOMMENDATIONS

- ❖ The probability of getting "Charged off" is lowest for 60 months tenure:
- ❖ For Purpose of loan, The probability of getting "Charged off" is highest for "Debt collection","credit card" and "other" categories. So it is not recommended to approve loans for these categories:
- ❖ For the type of Home ownership, though the number of "Charged off" loans are highest for both "rent" and "mortgage" categories, so are numbers high for "Fully paid" loans. So not much of inference can be found using this column.
- ❖ For Interest rate, it is observed that 21%-24%, 17%-24% are having very less "Charged Off" loans. But these interest rates are high. After 17-24%, the number of "Charged off" loans are less for 5-9%. Hence it is recommended to approve loans with this interest rate category.

# RECOMMENDATIONS

- ❖ For Employment Length, the least "Charged off" loans are for 8-9 years of experience. Hence it is recommended to approve loans for this category of applicants.
- ❖ For "revol\_util" categories, the least "Charged off" loans are for 0-20 category. So it is recommended to approve loans for this category.
- ❖ The least number of "Charged Off" loans for "annual\_income" ranges are 112k-140k followed by 85k-112k. But the number of applicants with these income ranges could be very less. The next category with least "Charged Off" loans was in the category 3k-31k. Hence it is recommended to approve loans for these categories.



# RECOMMENDATIONS

- ❖ It is observed that the least "Charged off" loans are for "Status\_verified" category. Hence it is recommended to verify the sources before approving the loan.
- ❖ The least "Charged Off" loans are for "dti" of 24-30 followed by 0-6. So it is recommended to approve loans for these categories.
- ❖ The "Charged Off" loans for funded amount by investor is least for 30-35k, followed by 25-30k and 20-25k. So, loans for these categories can be approved. But the number of loans with these categories may be less compared to others. Next best categories to approve loans are 0-5k and 10k-15k as the "Charged Off" loans are less.
- ❖ 11. The "Charged Off" loans for installements of 1047-1176, and 1176-1305 are very less compared to other categories. Hence it is recommended to approve loans with these categories

Thank You