Lending Club Case Study

Team:

1)Nischal Bathula 2)Kevin Shah

Scenario:-

- Given the dataset of Lending Club is about When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile
- The data given 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.

Types of Decisions to make by loan company

- When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

1. **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
    1. **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
    2. **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
    3. **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan
2. **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Problem Statement

- We have to identify the factors that are responsible for a loan company to make decision about the default applicants which will be very risky for company.

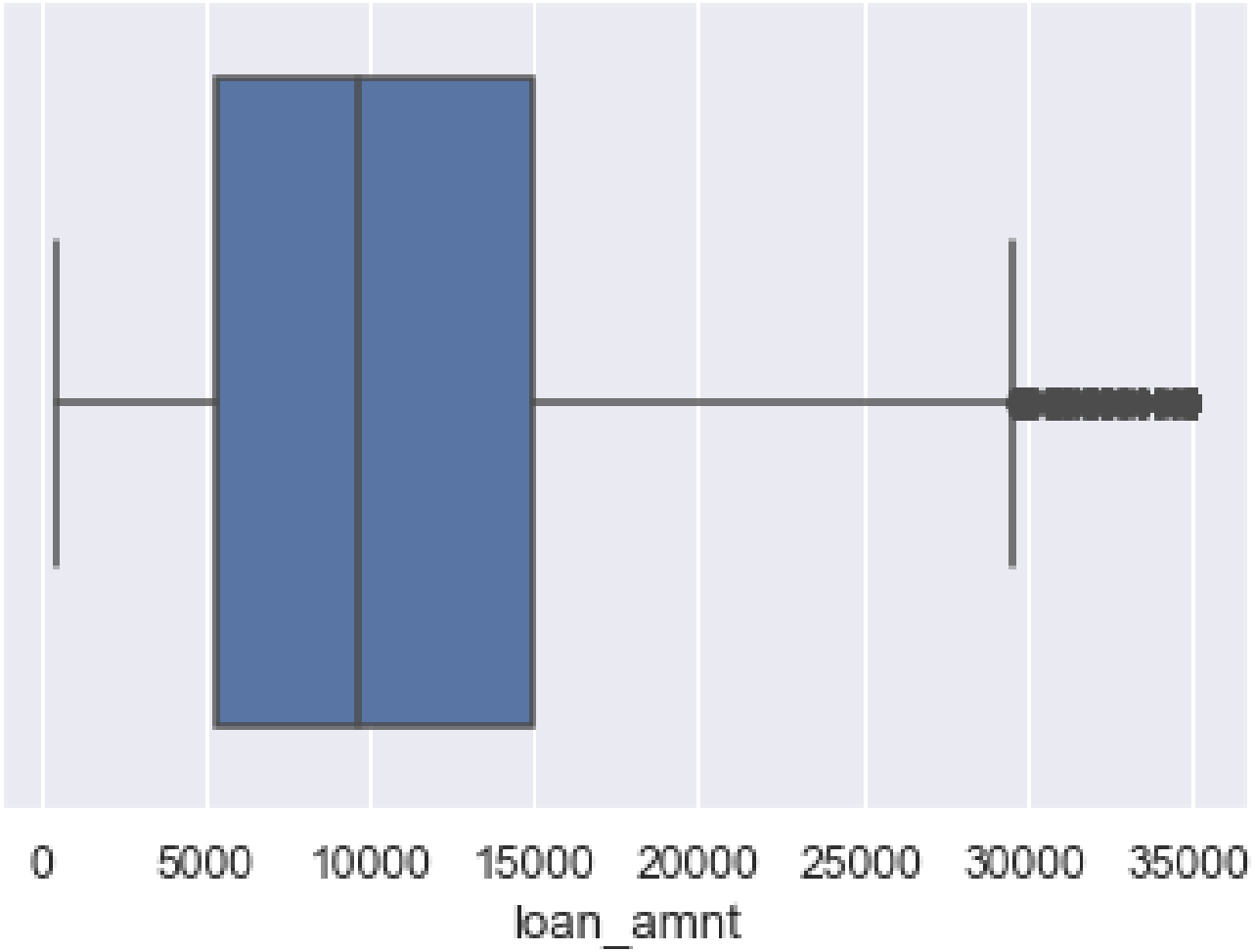
Observations/Recommendations/Insights:-

- Major Factors affecting the defaulters is:
  - Loan_amnt
  - Funded_amnt
  - Funded_amnt_inv
  - Terms_in_months
  - Int_rate_percent
  - Grade
  - Subgrade
  - Emp_length
  - Home_ownership
  - Purpose
  - Revol_bal
  - Loan_status
  - Revol_util_percent
  - Pub_rec_bankruptcies_b
  - Total_pymnt

Loan_amount Boxplot

Observations:

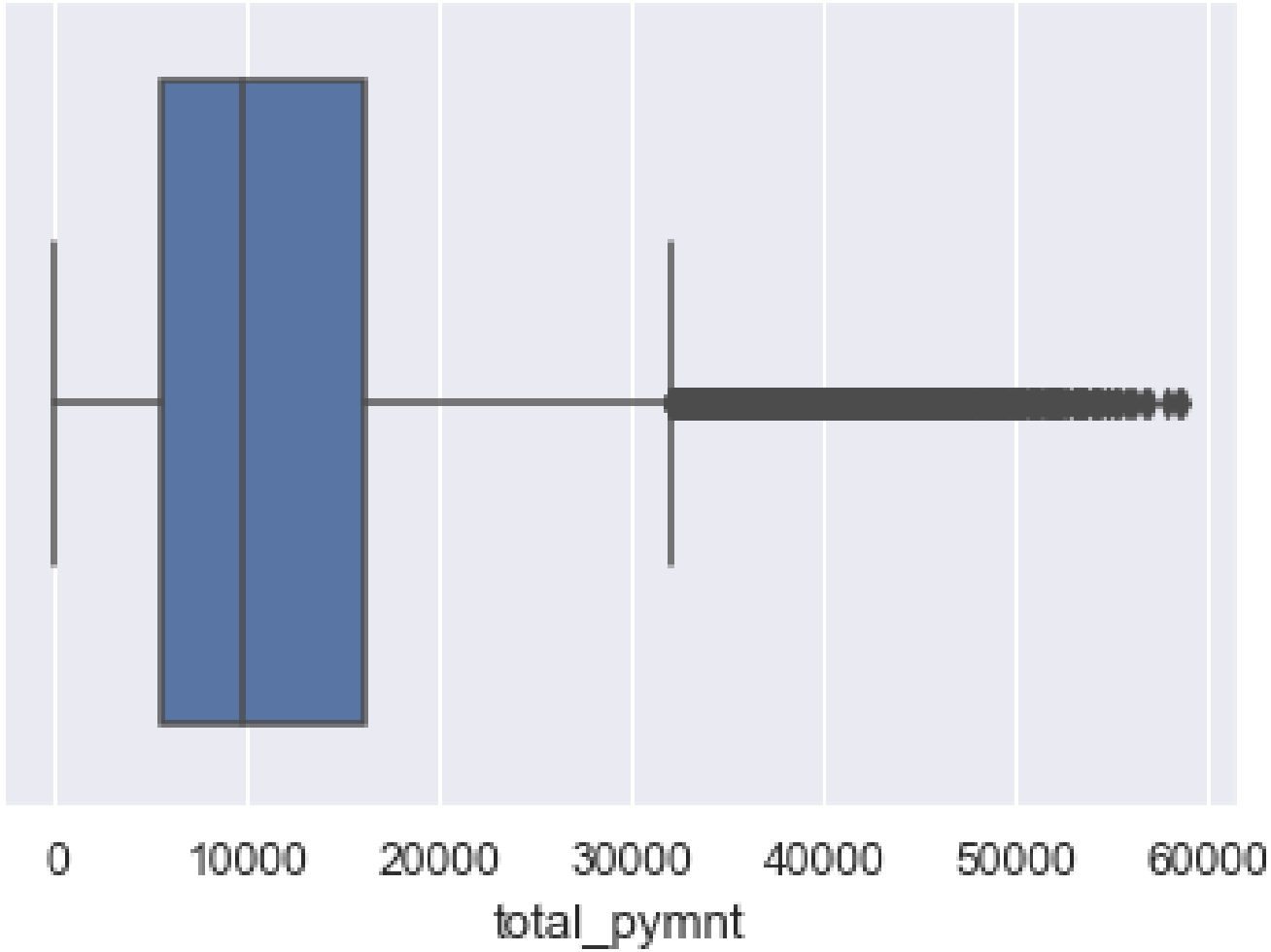
median is around $10000 and outliers is treated in range between $28000-$35000.



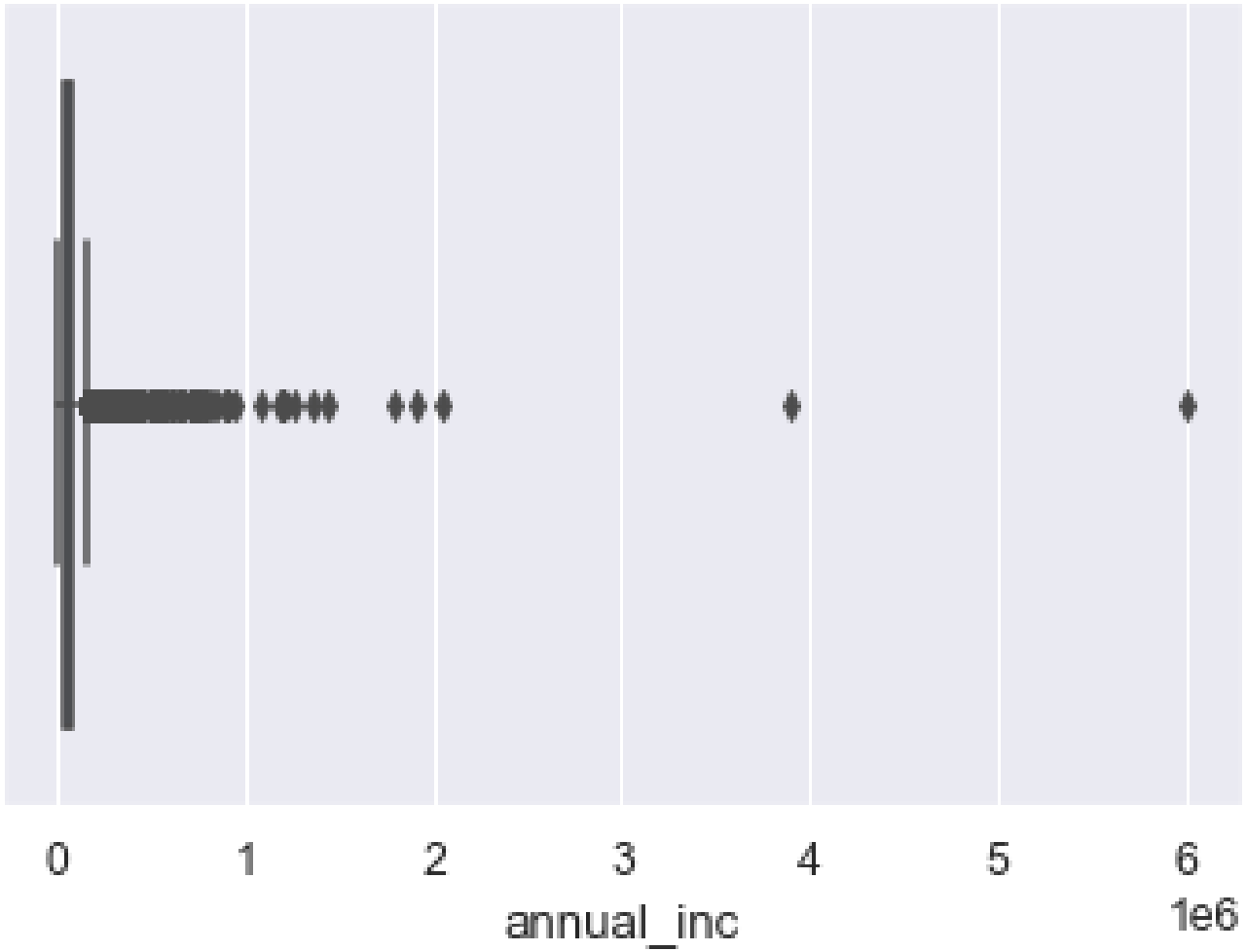
Total_payment BoxPlot

Observations:

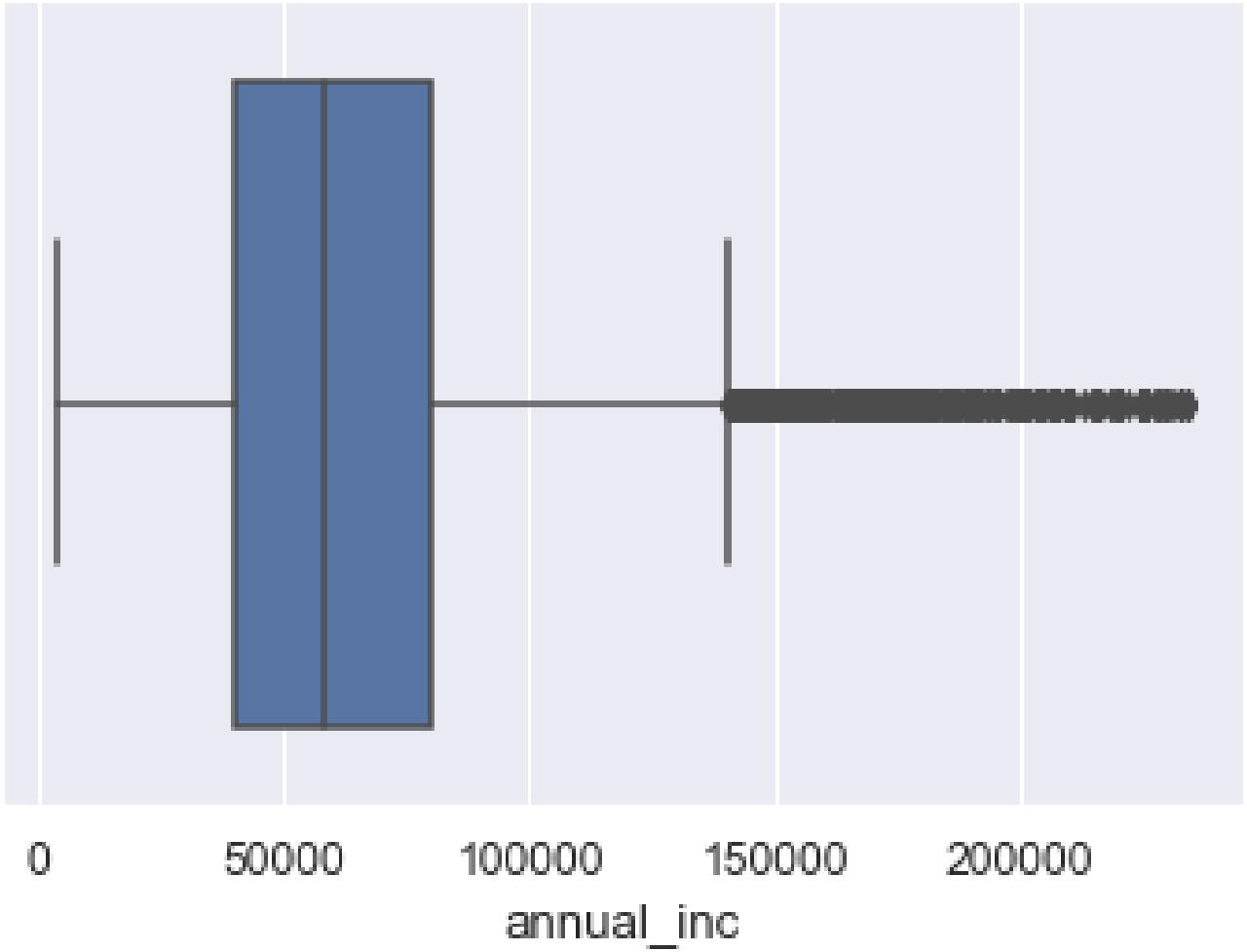
median is around $10000 and outliers is treated in range >$30000.



Annual_Income's Boxplot before removing outliers



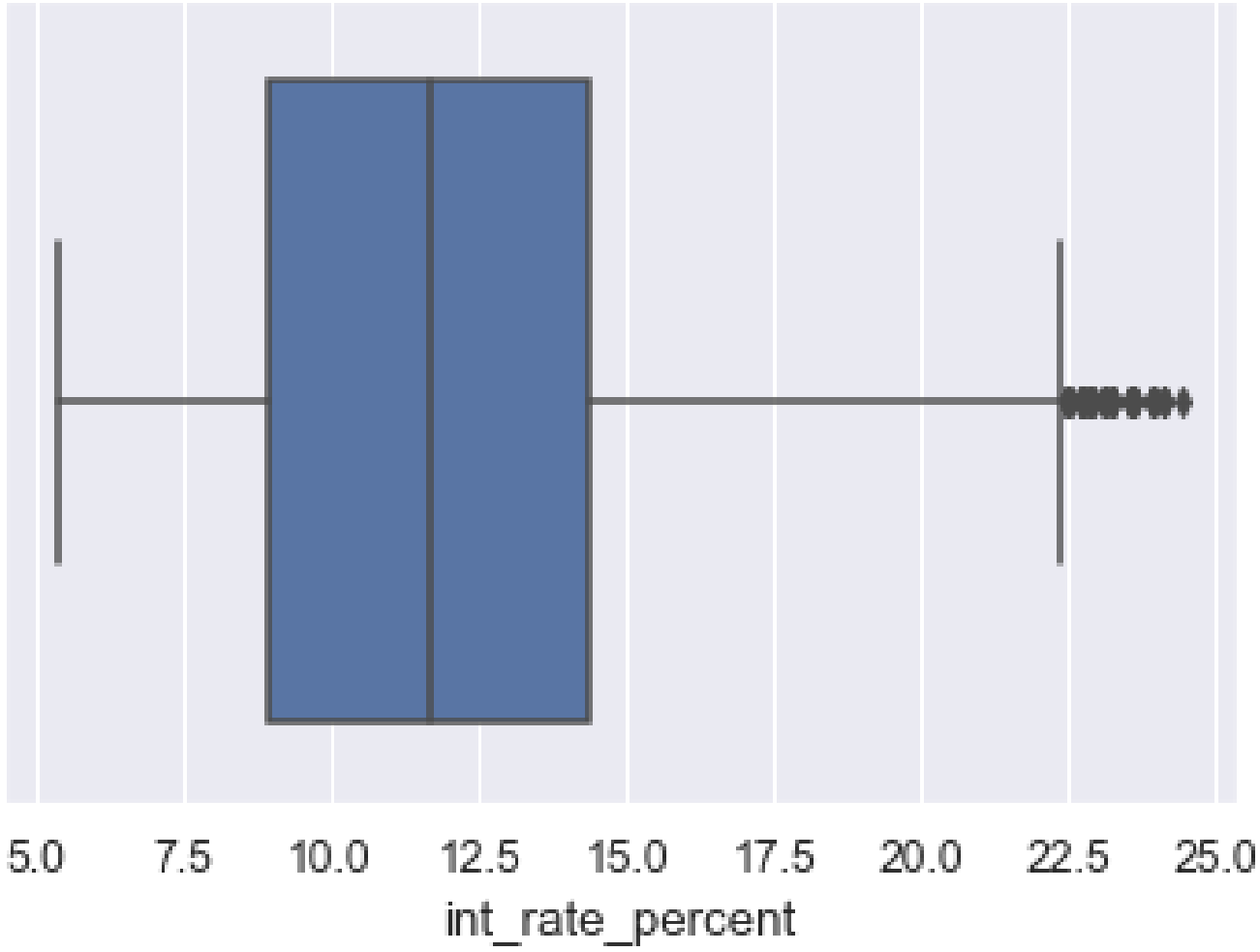
Annual_Income's Boxplot after removing outliers



Interest rate boxplot

Observations

median is around %12 Intrest rate and outliers is treated in range greater than 22.5 means the applicants whose intrest rate is greater than 22.5% will be defaulter.



Barplot for Grade VS Annual Income

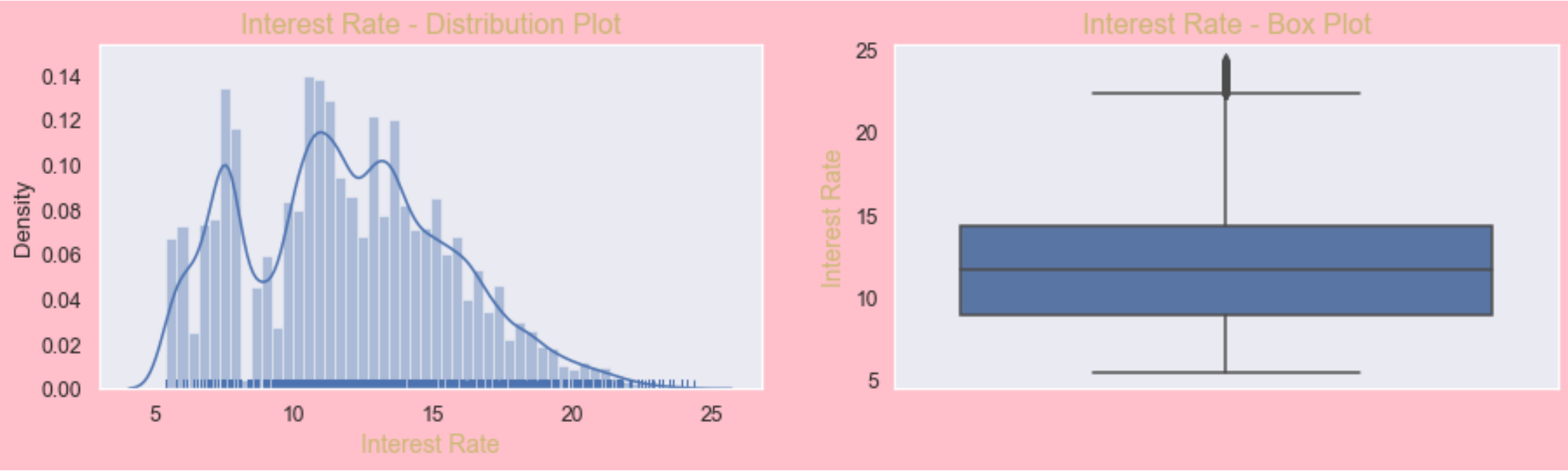
Insights:

- 0 = fully Paid , 1 = charged_off

Observations:

- From this we can conclude that the ones getting 'charged off' have lower annual incomes than the ones who'paid fully' for each and every grade (i.e. at same interest range)

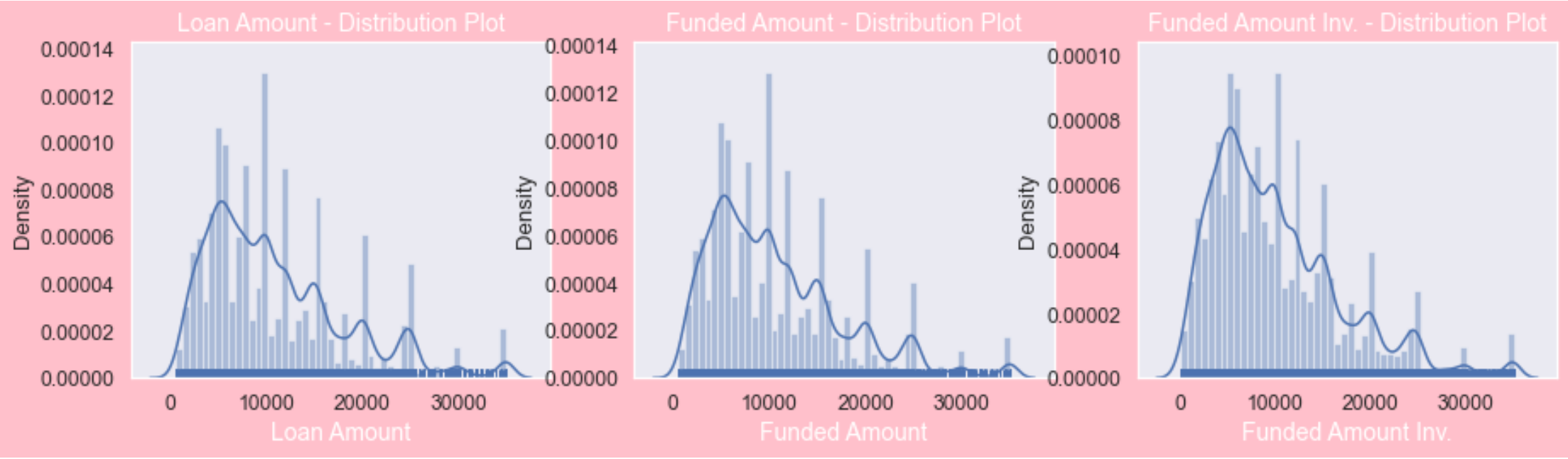
Distribution and Box Plot for Interest Rate



Insights:

- The plots show that most of the Interest Rates on loans are in range of 10% - 15%

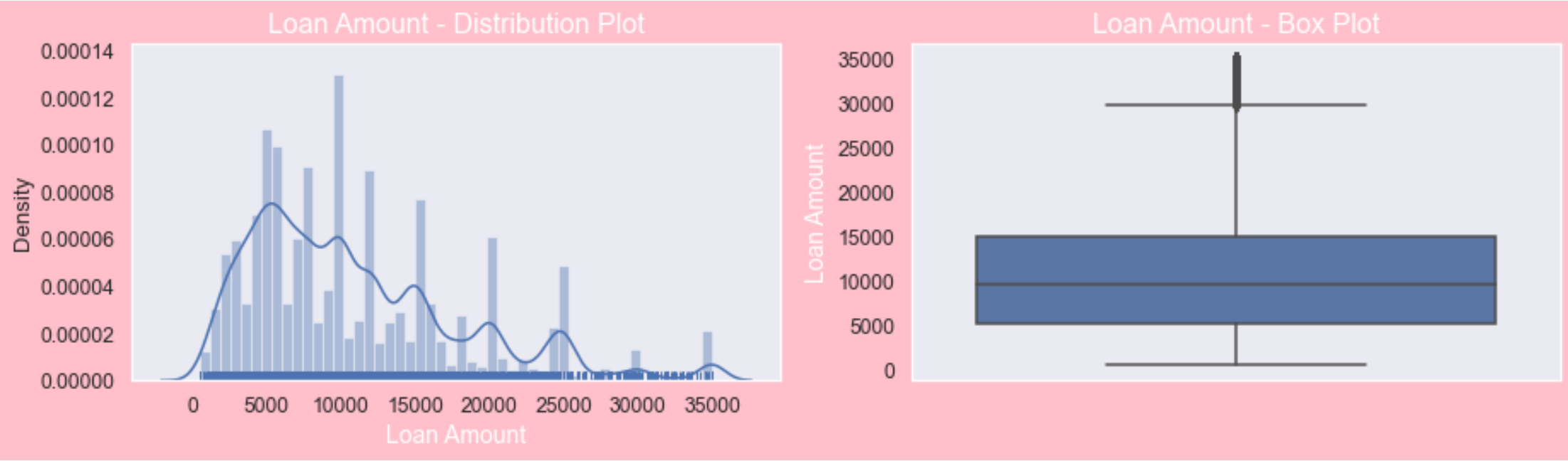
Distribution Plot for loan amt,funded amt,funded_amt_inv



Insights:

- Distribution of amounts for all three looks very much similar.

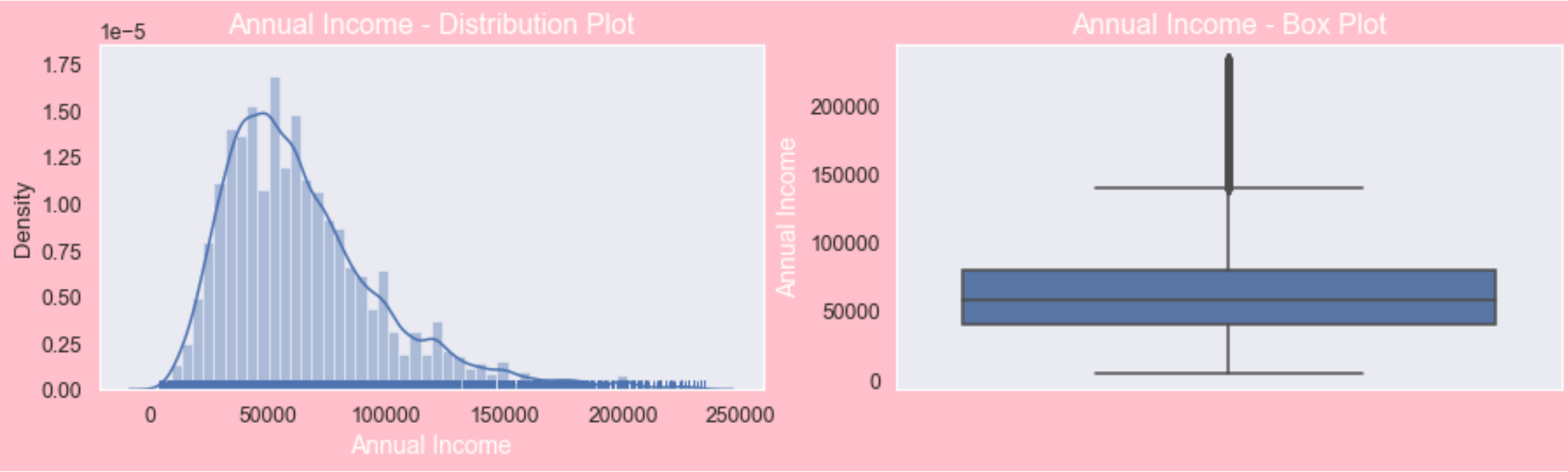
Distribution Plot & Box Plot for Loan_amnt



Insights:

- Below plots show that most of the Loan amounts are in range of 5000 - 15000

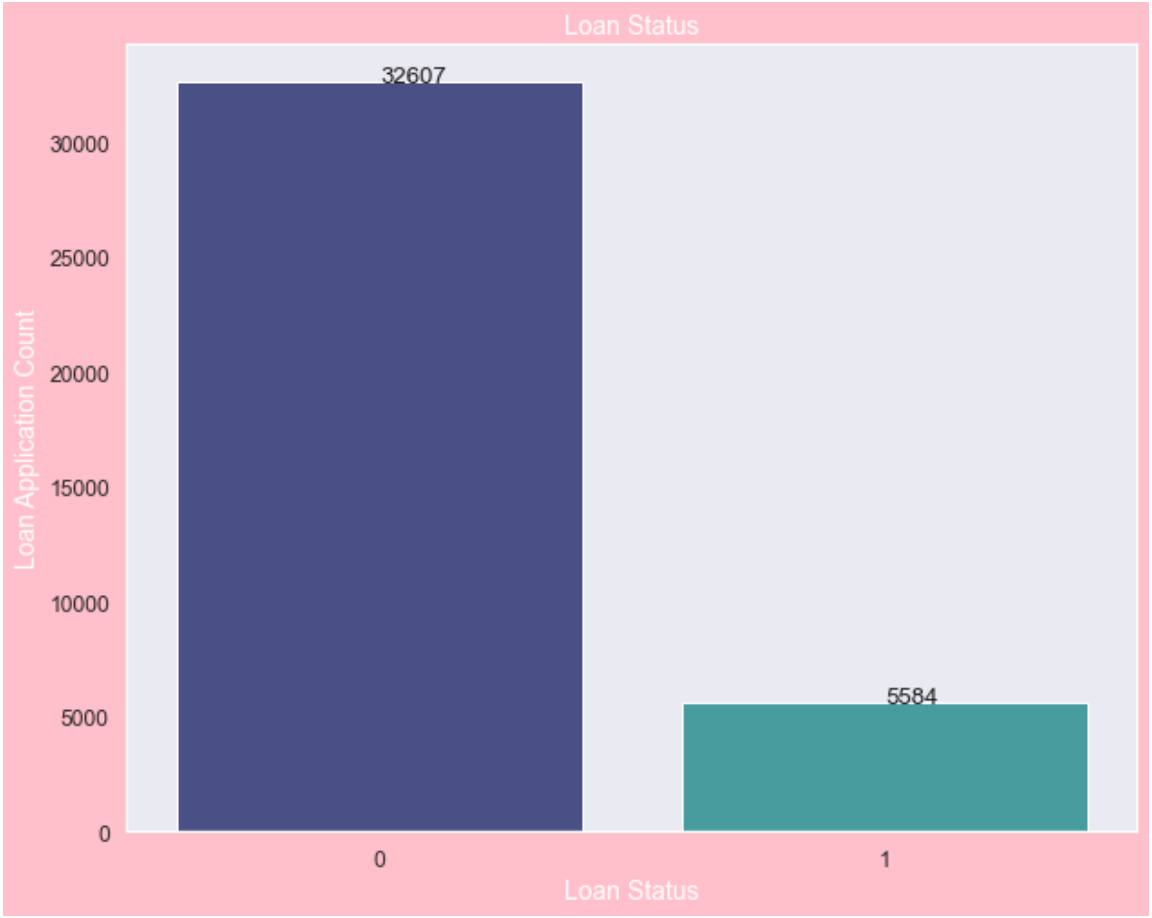
Distribution and Box Plot for Annual_Income



Insights:

- Below plots show that most of the borrower's Annual incomes are in range of 40000- 80000

CountPlot for Loan_status



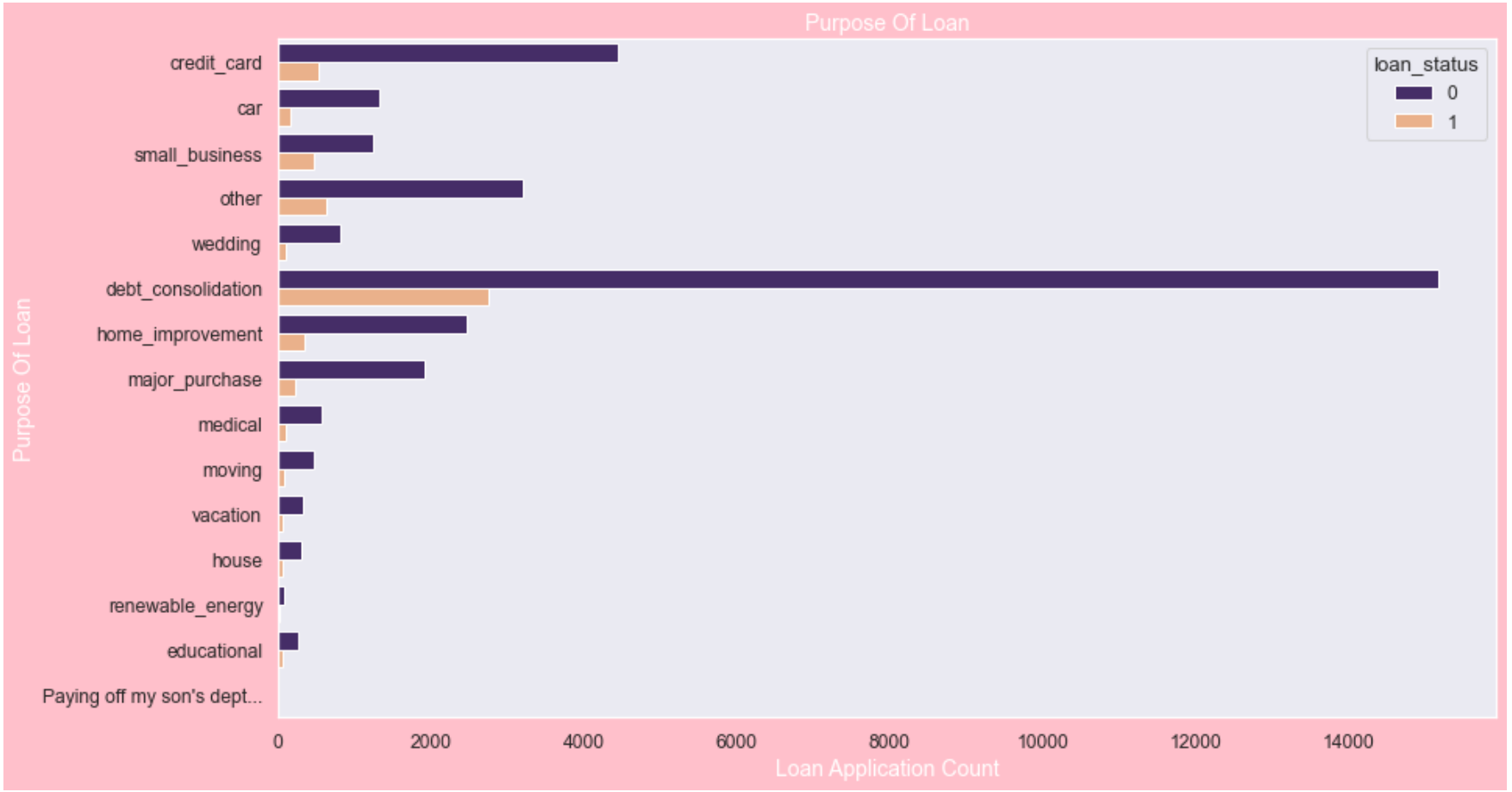
Insights:

- Assumptions: loan_status-0:Fully Paid, 1-Charged_off

Observations :

- The plot shows that close to 14% loans were charged off out of total loan issued.

Count Plot for Purpose of Loan

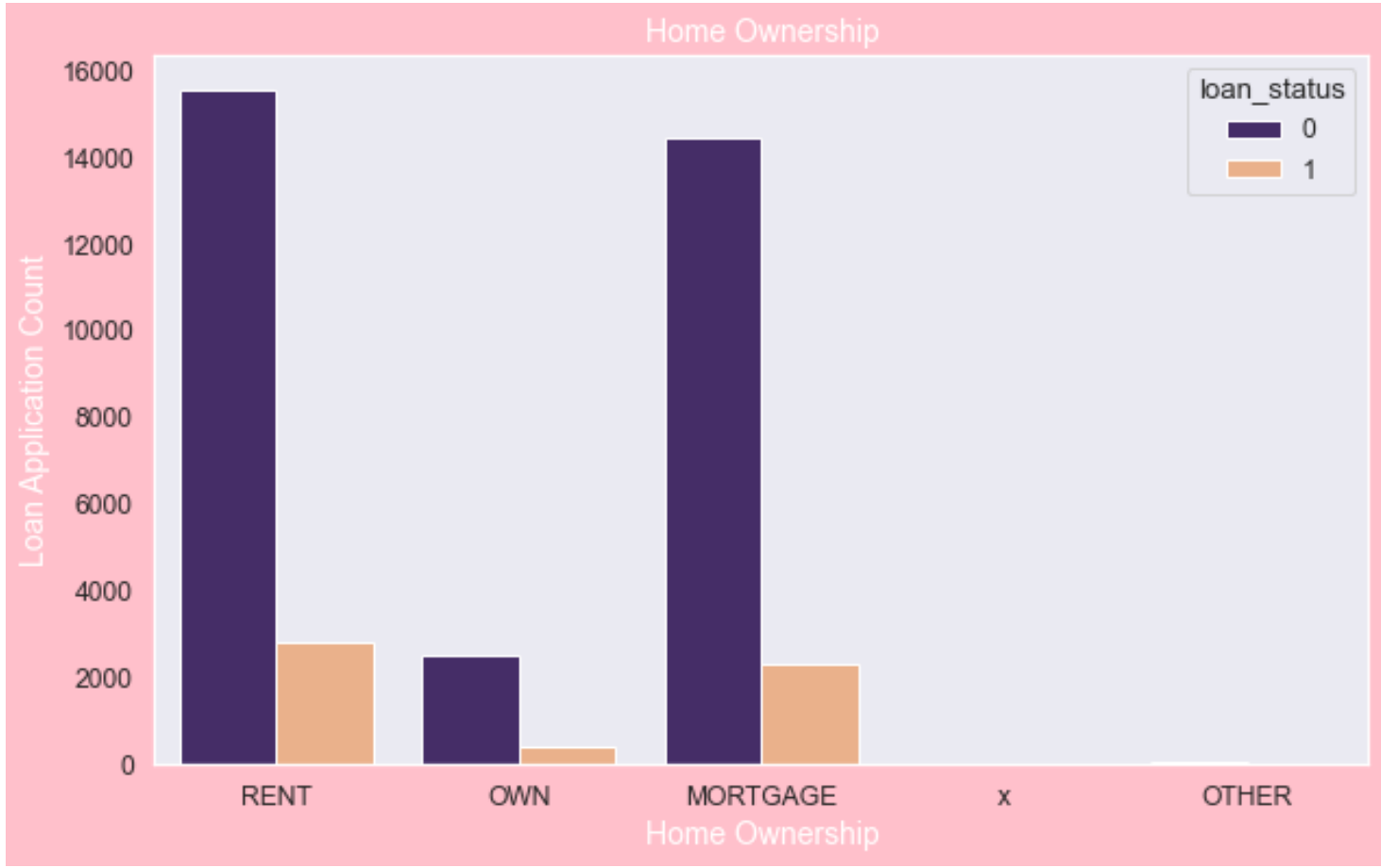


Insights:

- Most of the loans were taken for the purpose of debt consolidation & paying credit card bill.

- Number of charged off count also high too for these loans.

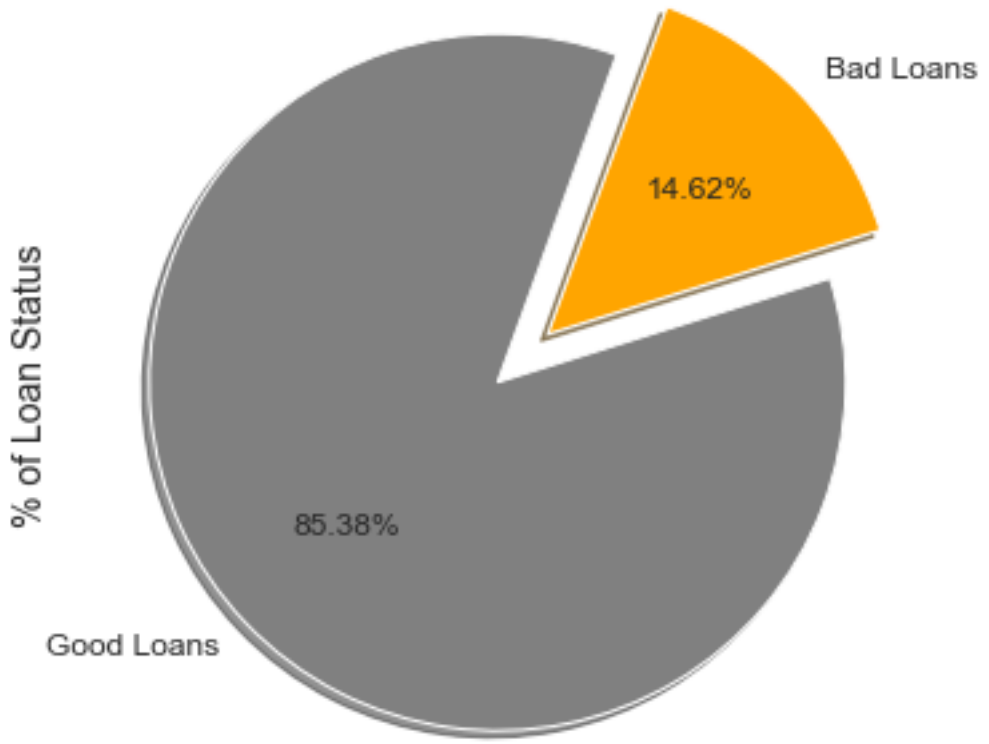
Count Plot for Home_Ownership



Insights:

- The plot shows that most of them living in rented home or mortgazed their home.
- Applicant numbers are high from these categories so charged off is high too.

PieChart for Loan Status

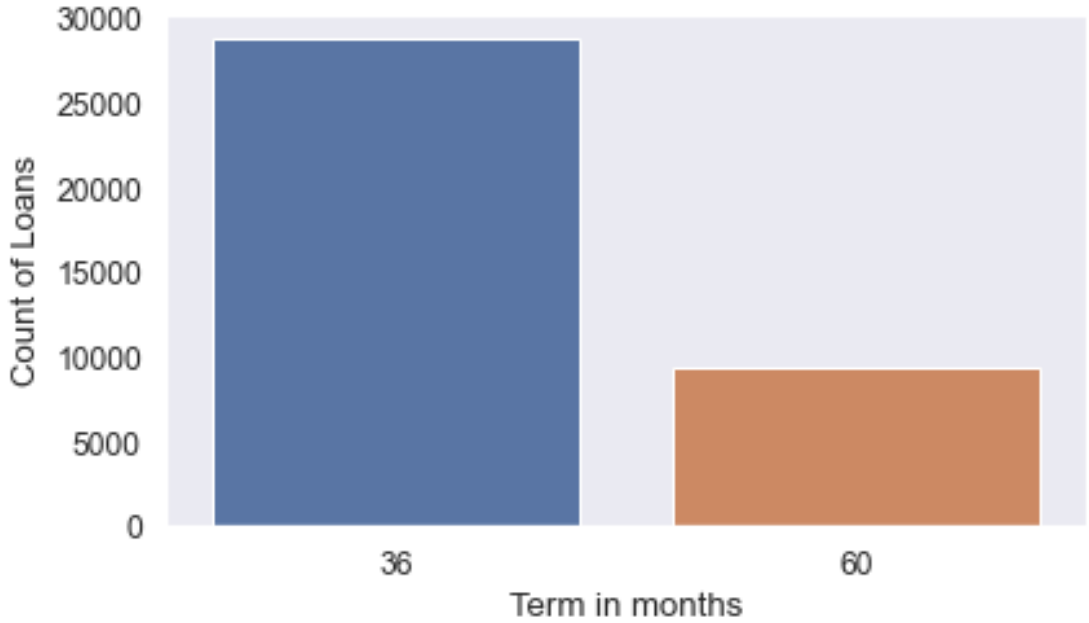


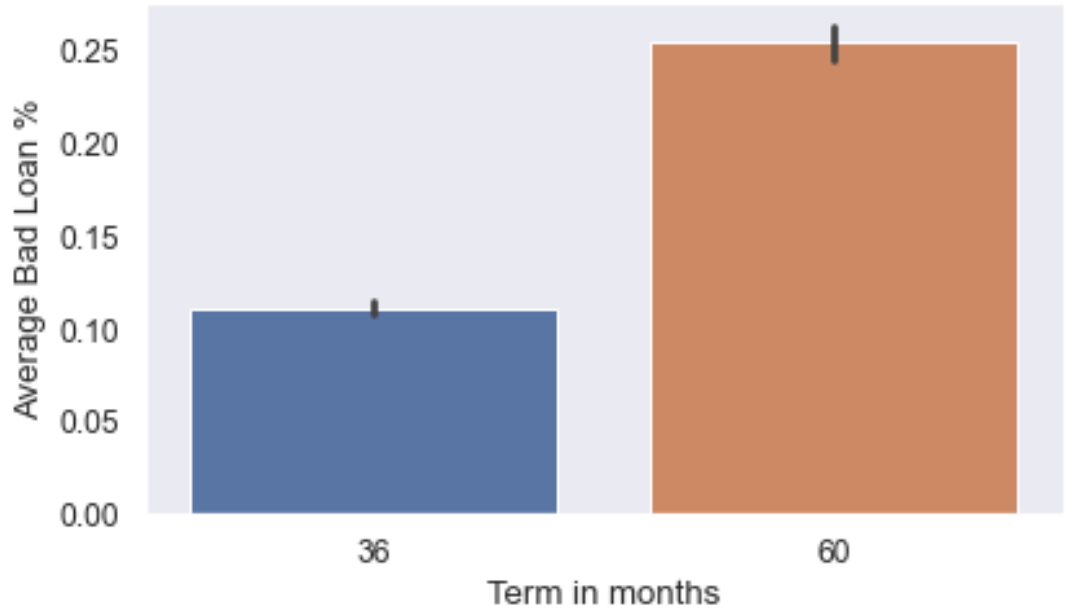
Insights:

- Fully paid or good loans consist 85.38% of total loans in the cleaned data frame.

- Charged off or bad loans consist 14.62% of total loans in the cleaned data frame.

Count Plot for Terms in months & Bar Plot for Terms with Average Bad Loan



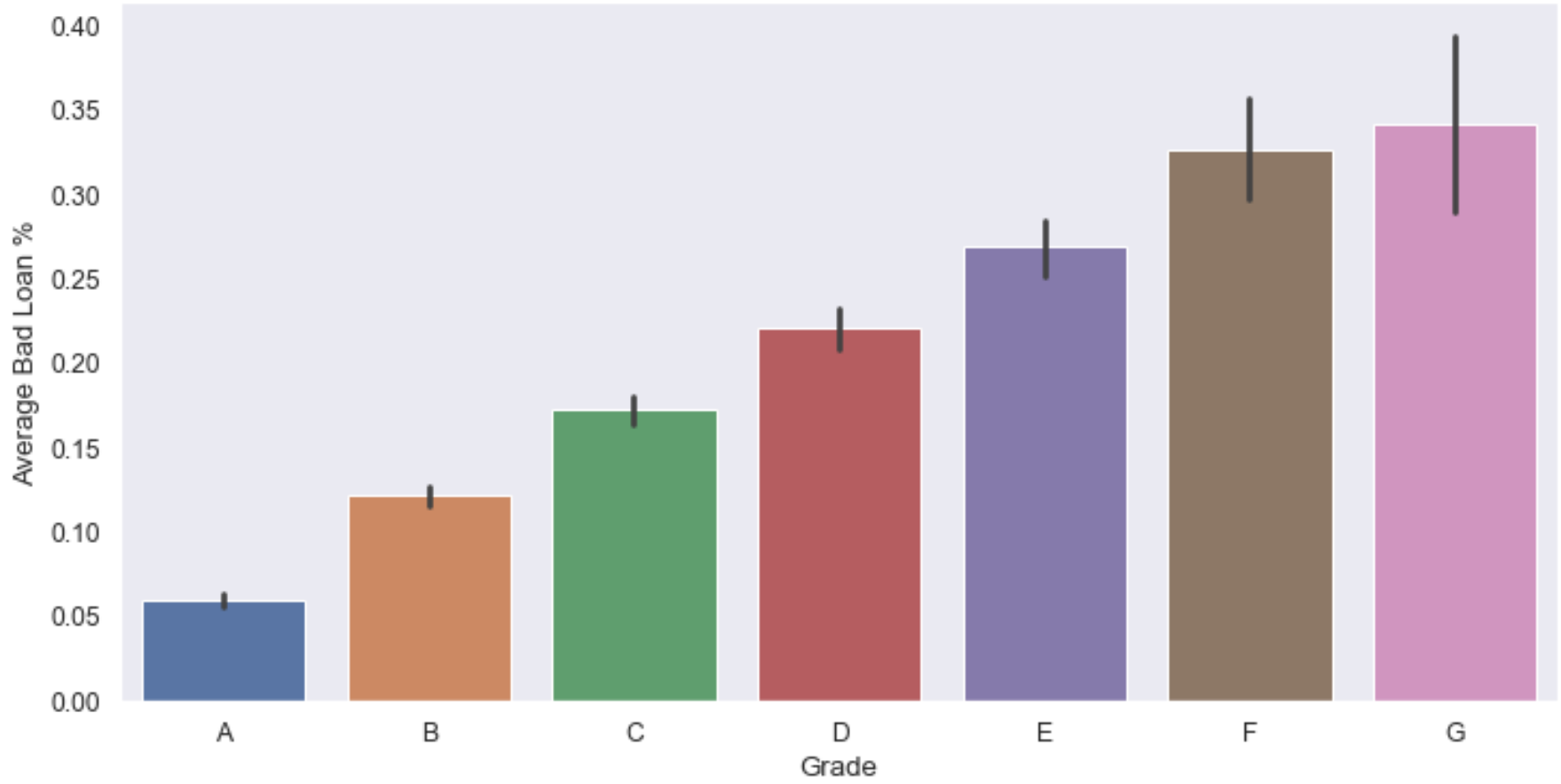


Insights:

- If loan term is 60 months then it has 25% average who default on loans as compared to loan term of 36 months which has 11% average who default on loans.

- Also from the count plot, one can infer that number of entries of 36 months is approx 3 times number of entries of 60 months. Still it has 11% average who default on loans.

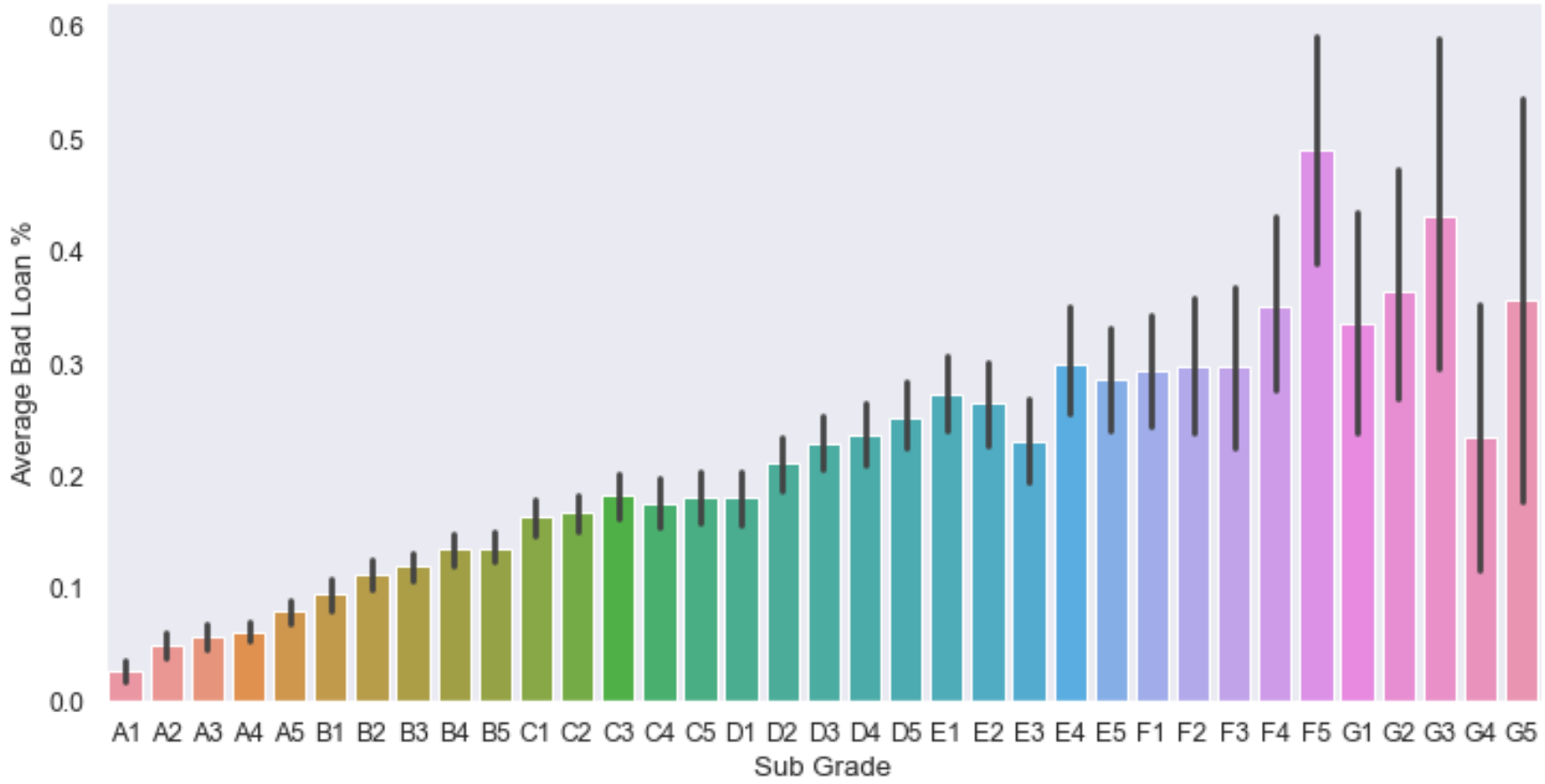
Analyzing Grade with Bad Loan % using Barplot



Insights:

- In Grade A-G the bad loan or the defaulters percentage has increased ,Hence Grade F & G has highest chance to get proned to be defaulters.

Bar Plot for Subgrades vs Average Bad Loan



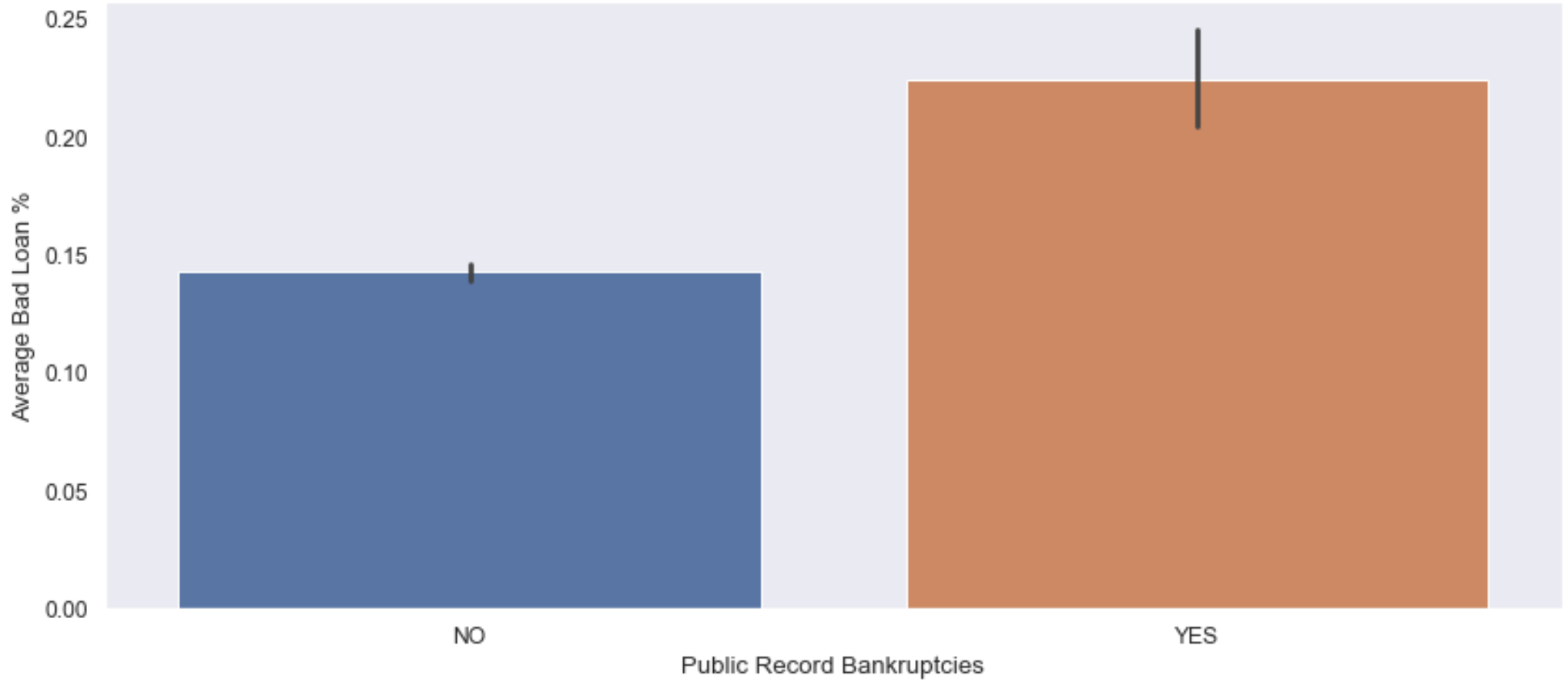
Insights:

- Among the subgrades too of lower Grades A to D - bad loan % increments in a consistent manner between the sub grades in each grade.

- Grades E to G already determined as higher risks.

- F5 subgrade has the higher bad loan percentage.

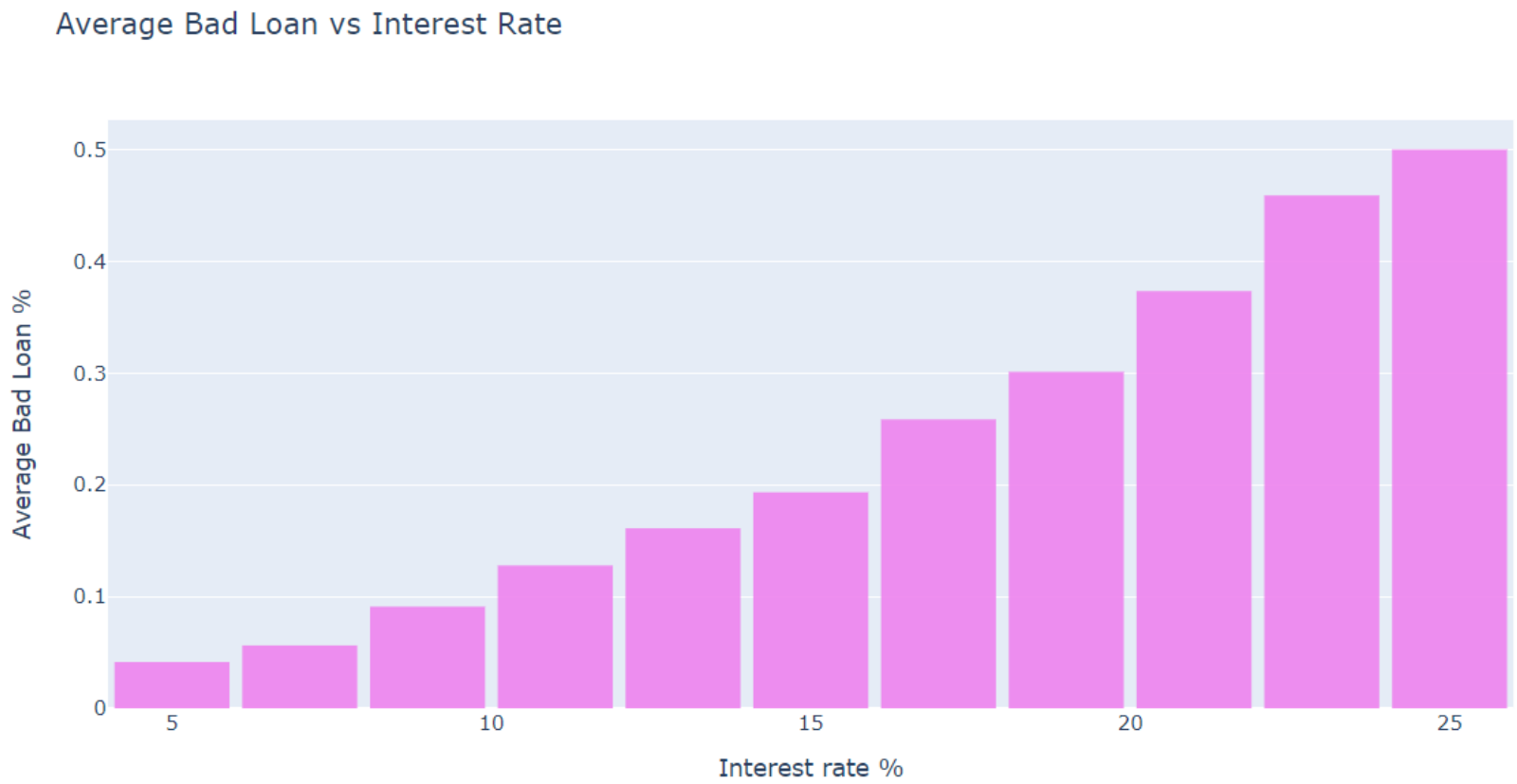
Bar Plot for Bankruptcies_record vs Bad Loan%



Insights:

- The accounts having bankruptcies tends to be defaulters more than the one who doesn’t have bankruptcy record

Histogram for Bad Loan Vs Interest rate



Insights:

- Higher the interest rate , higher the chances to have bad loan and to be a defaulter.

Chart to Count the good and bad loan with respective to grades

Grade	Bad Loan %	Good Loan Count	Bad Loan Count
G	0.342561	190	99
F	0.327044	642	312
E	0.269732	1906	704
D	0.221205	3915	1112
C	0.173094	6411	1342
B	0.122097	10167	1414
A	0.060239	9376	601

Insights:

- Bad Loan Count and its ratio is high for Grade g and it decreases going down to Grade A

Recommendations:-

- The records who is having Grade G,Highest Interest_Rate,Mire Bankruptcy_record,More Terms in months,Having Rent or Mortgage House,Purpose(Debt Consolidation,Credit card),Low Annual_Income will have the affect to be the risky account and more tends to be a default applicant.s