# Analysis of parameters indicating a loan default

ANILABH & SAMIK

# Loan Default Case Study

Borrowers who default cause the largest amount of loss to the lenders.

This analysis aims to identify patterns that 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.

# What techniques and tools were used

We used our learnings from the Exploratory Data Science course to identify gaps in data

We cleaned up or imputed the missing data based on the impact and size of the gaps

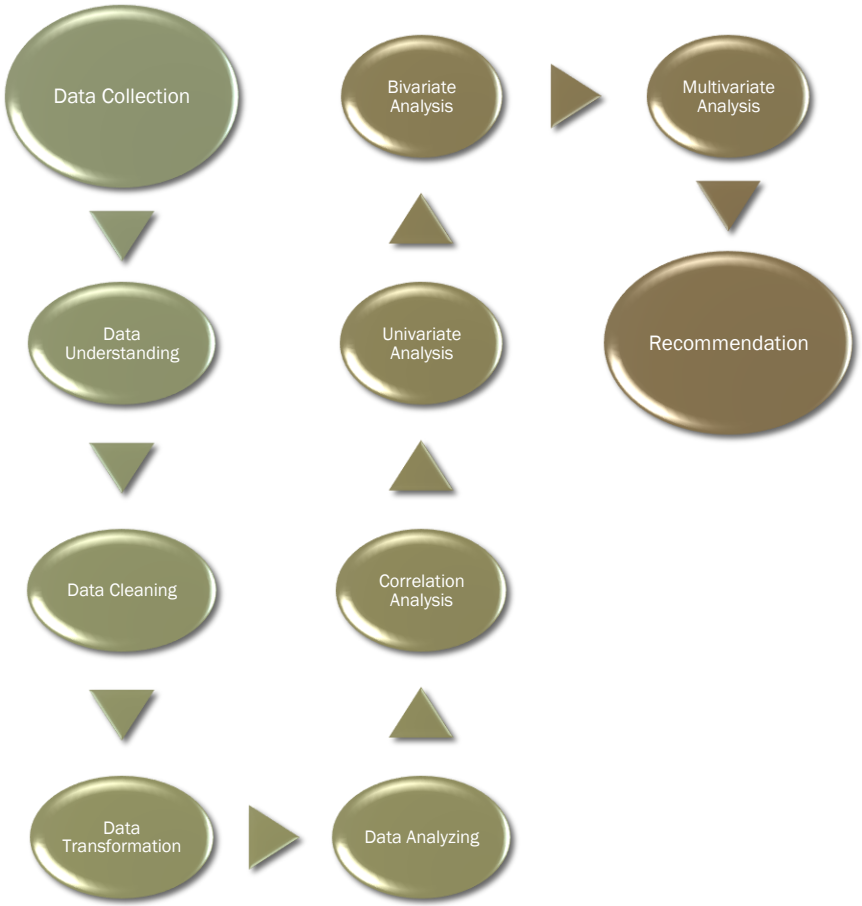
Used NumPy and pandas libraries to identify and fix the gaps and find patterns in the data

Analysed the data in columns to find the spread and spans of the data in columns.

Analysed how the default rate is impacted by categorical and continuous variables. For continuous features, we used quartile-based binning.

Used Bivariate analysis to find interesting cross patterns in the default rate concerning two variables

Used matplot and seaborn libraries to visualize the results of our analysis



# What are the key observations?

We analysed the default rate across multiple univariate and bi-variate analyses. We will focus on the

Parameters that have the biggest impact on the default rate. Among them are:

- Loan Grade

- Loan Term

- Funded Amount

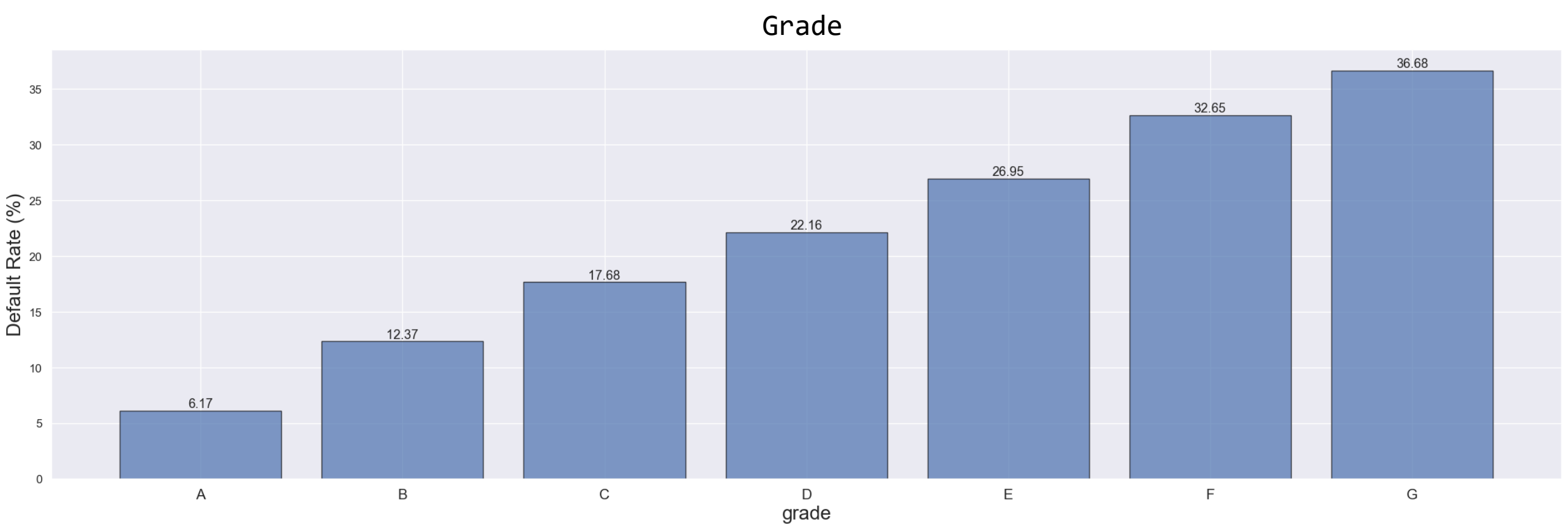
- DTI

- Purpose of loan

- Interest Rate

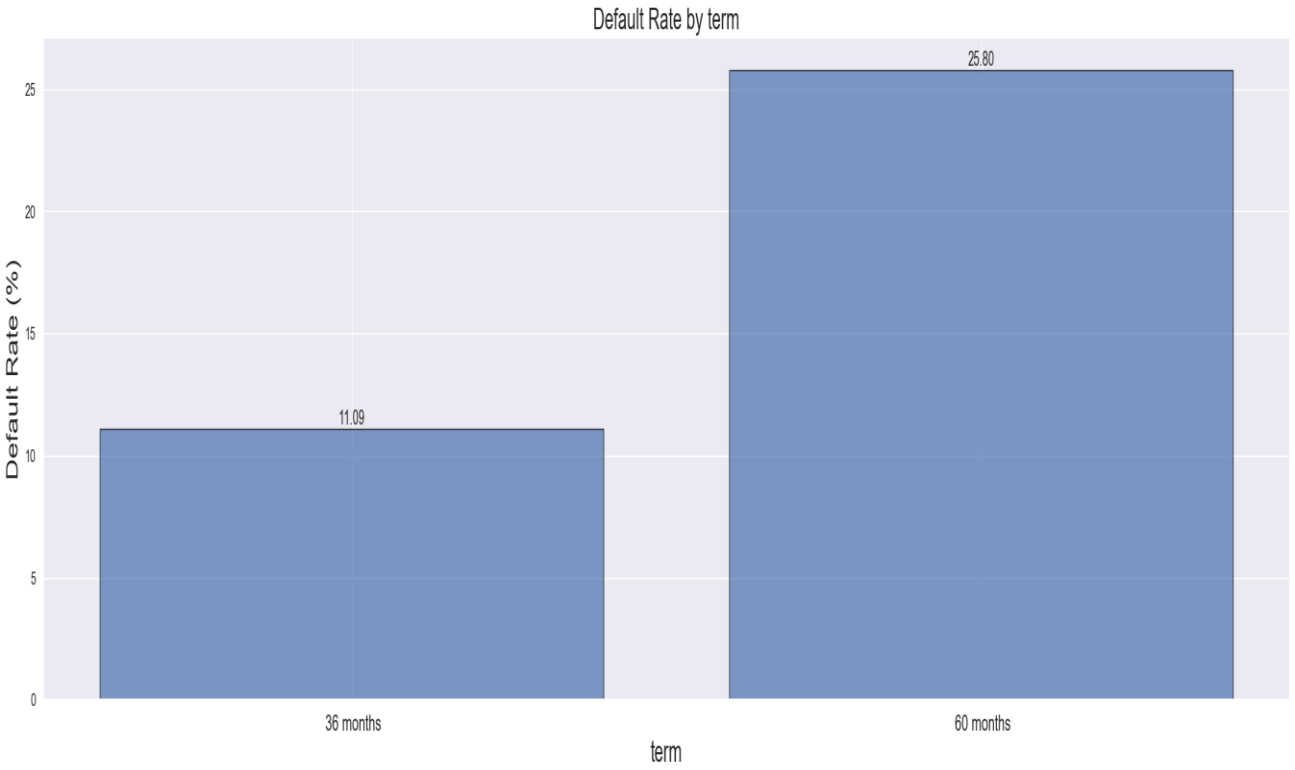
- Verification Status

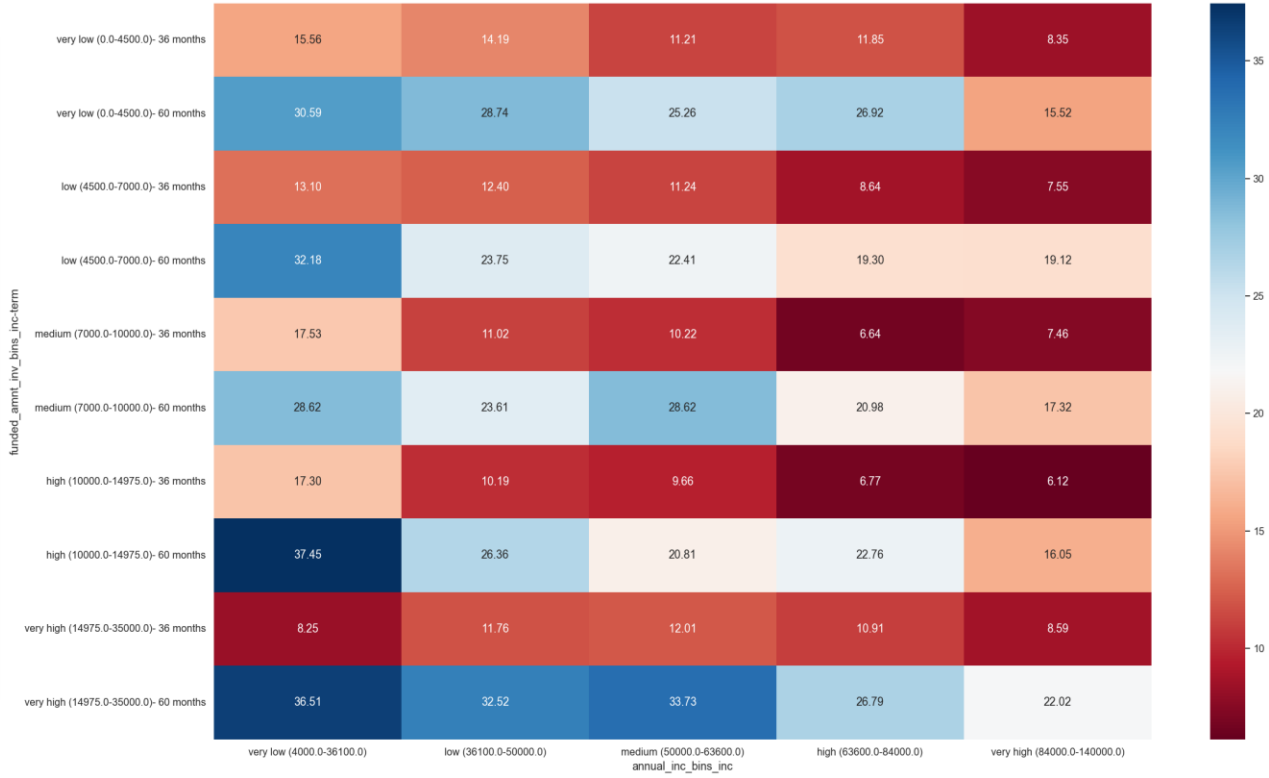
We will look at these in detail in the next slides…




The default rate continuously increases from Loan grade A to G and is exceptionally high beyond E which is at 26%. Grades F and G have more than 32 % defaults.
Loans in this category need special scrutiny. Sub Grades of F5, G3 and G5 show an exceptionally high default rate of 40%+.

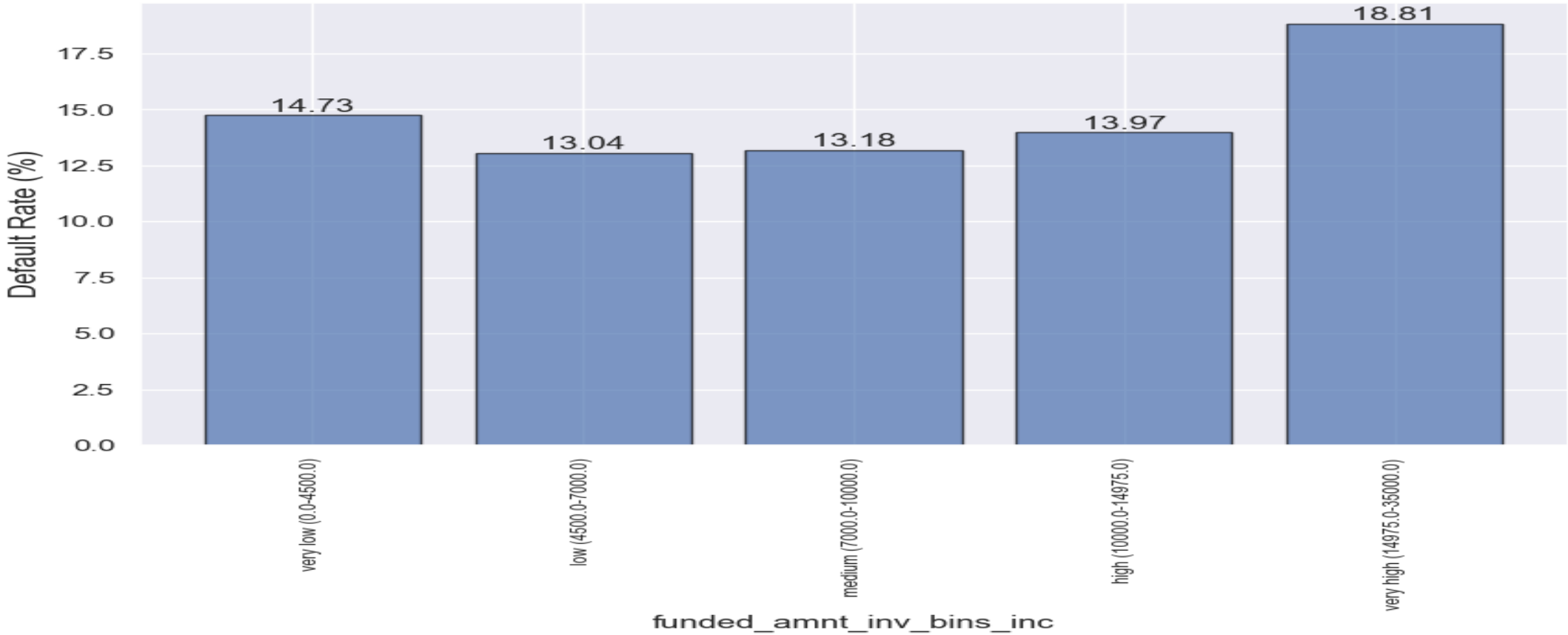
# Term



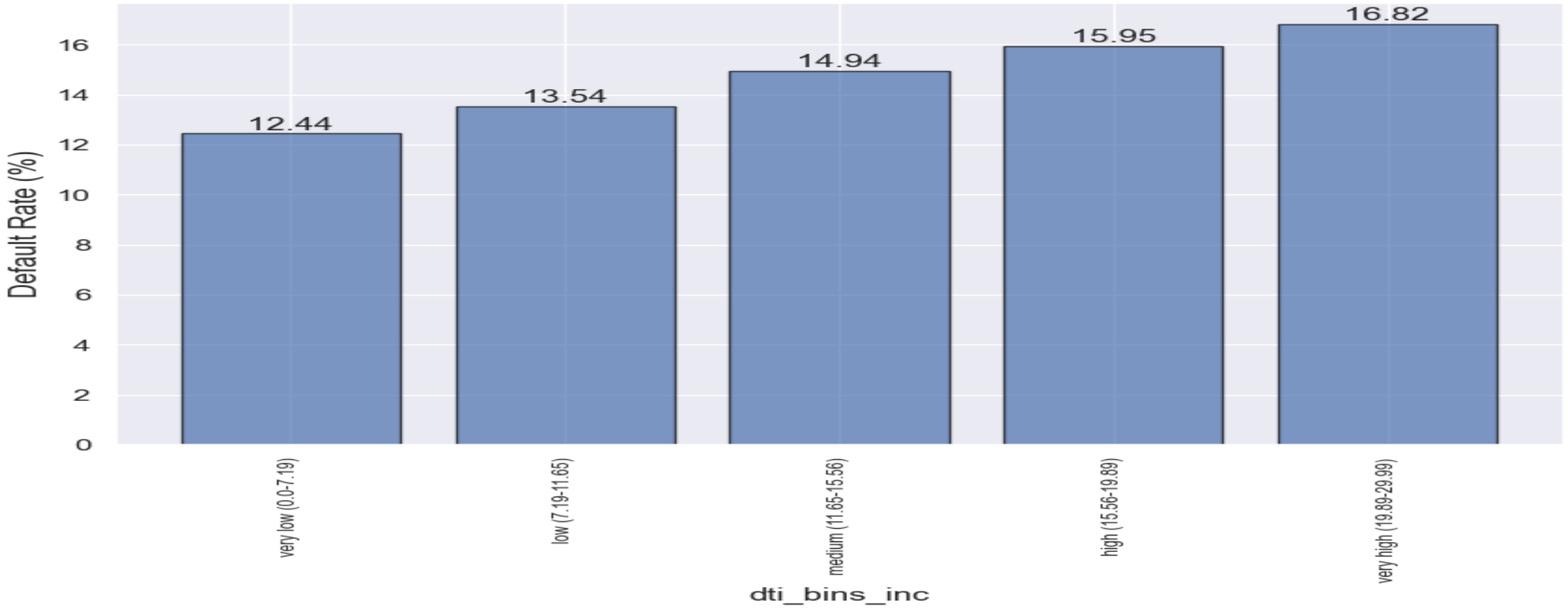

Loans with Term of 60 months have a significantly higher default rate of 25% than those of 36 months which are only at 10% which we can see in the heatmap as well. Any long term loan with any range of loan amount by investor is having higher chances of defaulting regardless of any income range. The company may need to put more stringent checks on loan approvals for terms of 60 months.

## Funded Amount by Investor



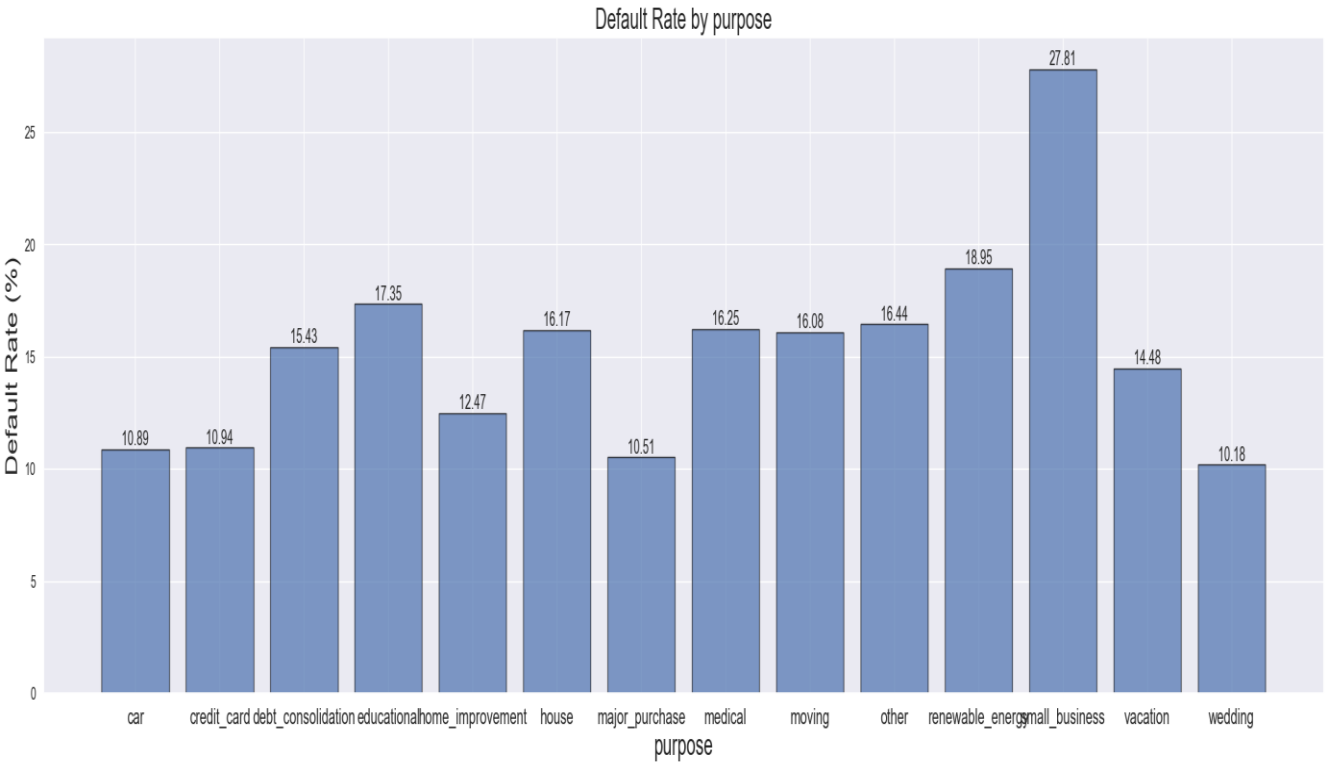

Funded_amnt_inv more than 15000 has a significantly high default rate of 18%.
The default rate decreases with the decreasing funded amount by investors but increases at the lowest bin i.e. below 4500. Investors are advised to limit funding loans beyond 15000.

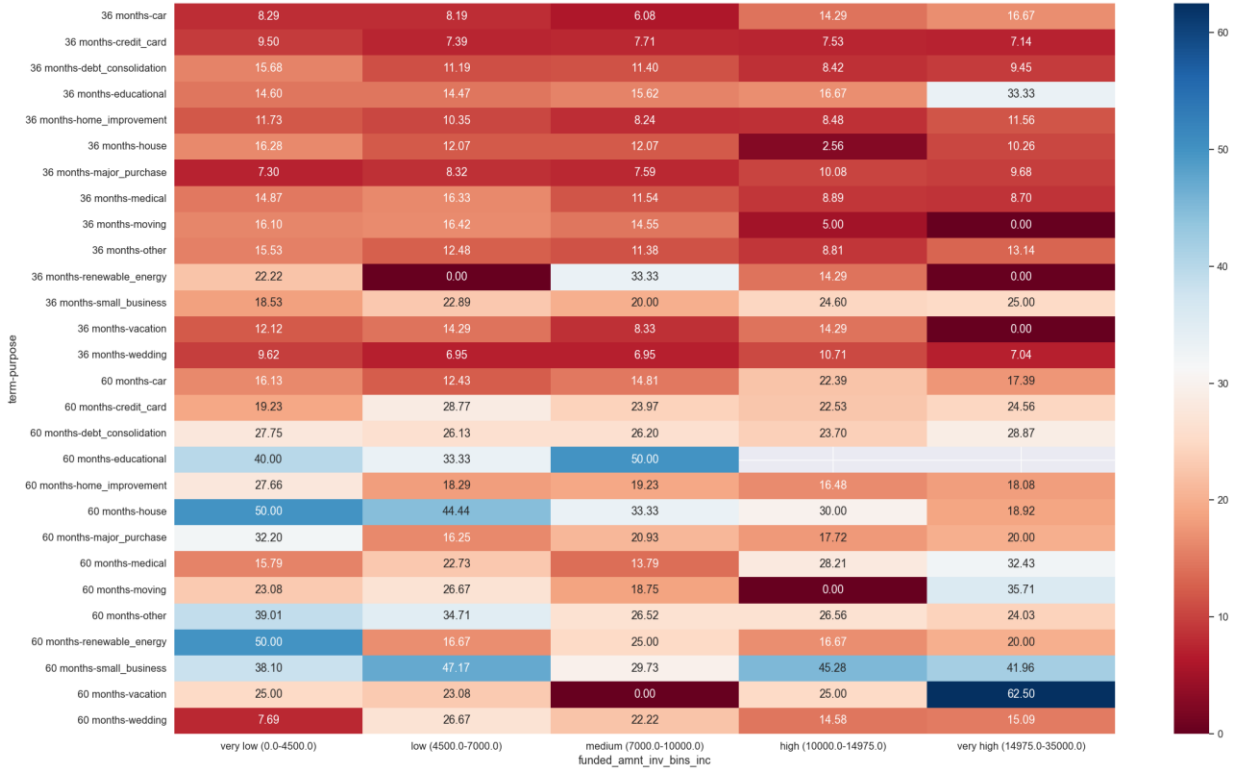
# DTI




DTI has a direct correlation with the default rate. The default rate increases to nearly 17% for dti above 19. This is expected as customers with high debt payments to their monthly pay are possibly under financial strain. This needs to be checked during approval. And DTI above 19 should be scrutinized.

# Purpose

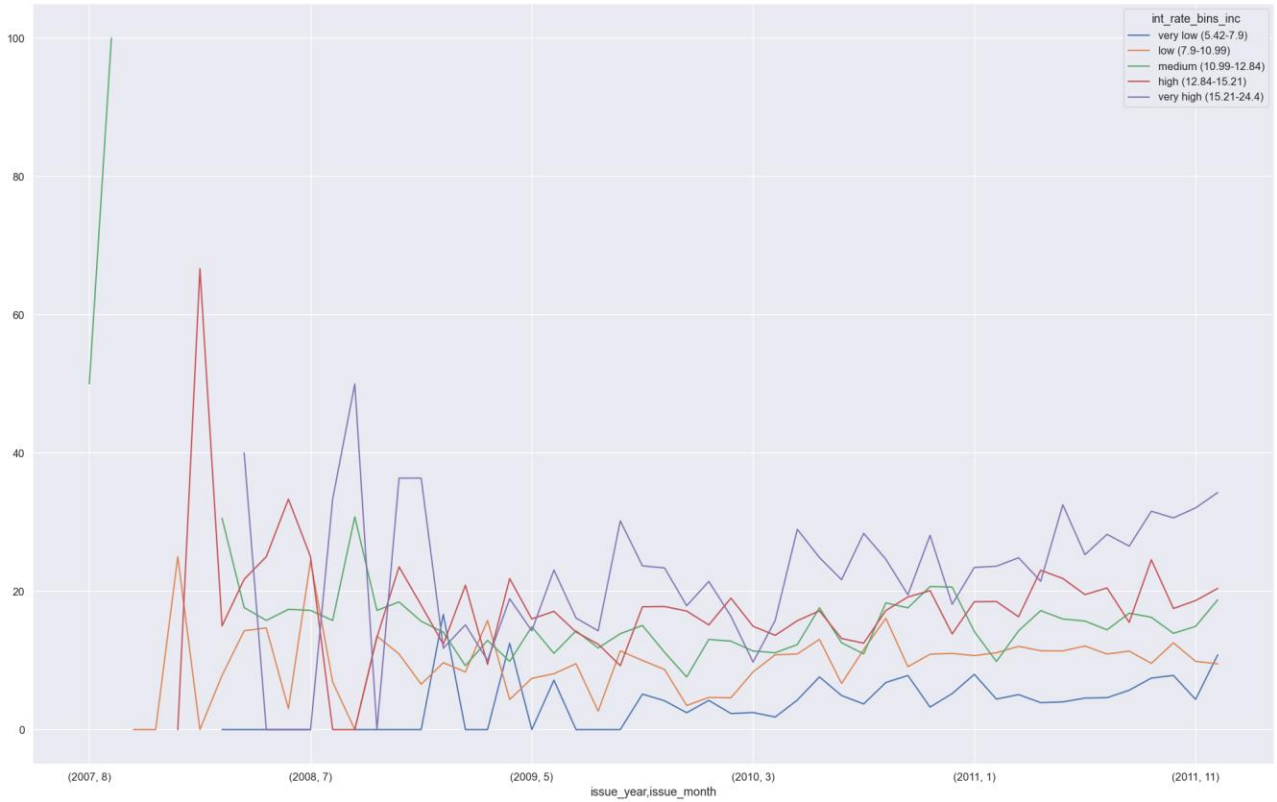


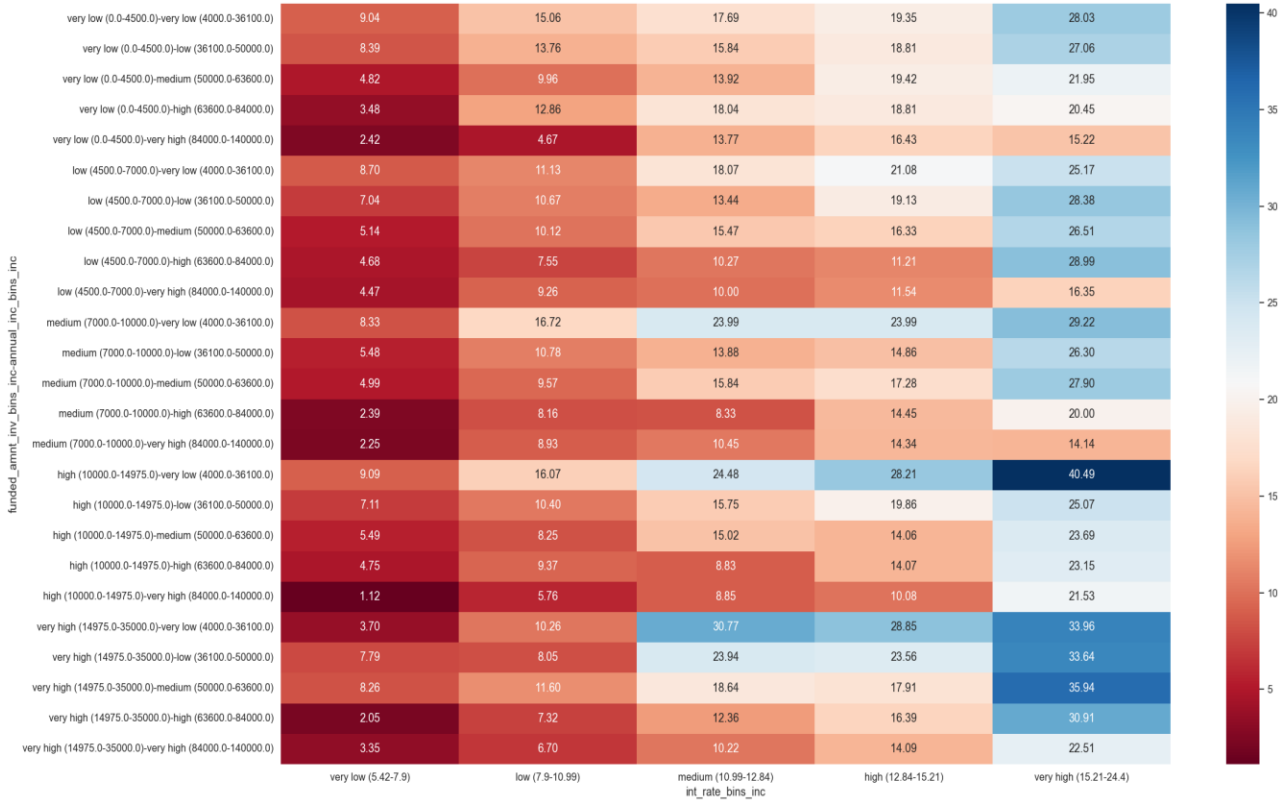




Purpose marked Small_business have a very high default rate of 27%. This shows that small business run a risk of not being able to pay as they might not survive tough business situations.

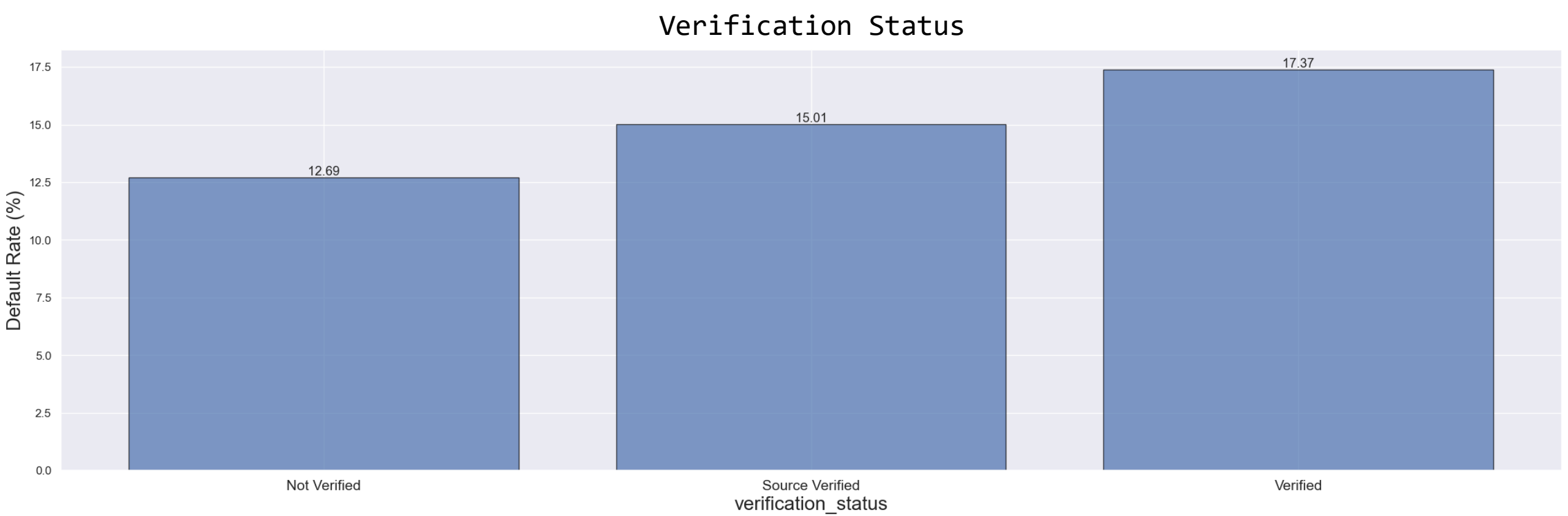
Note - Very high range loans(15000+) with term of 60 months for the purpose of vacation are most likely to default(60%+).

# Interest Rate

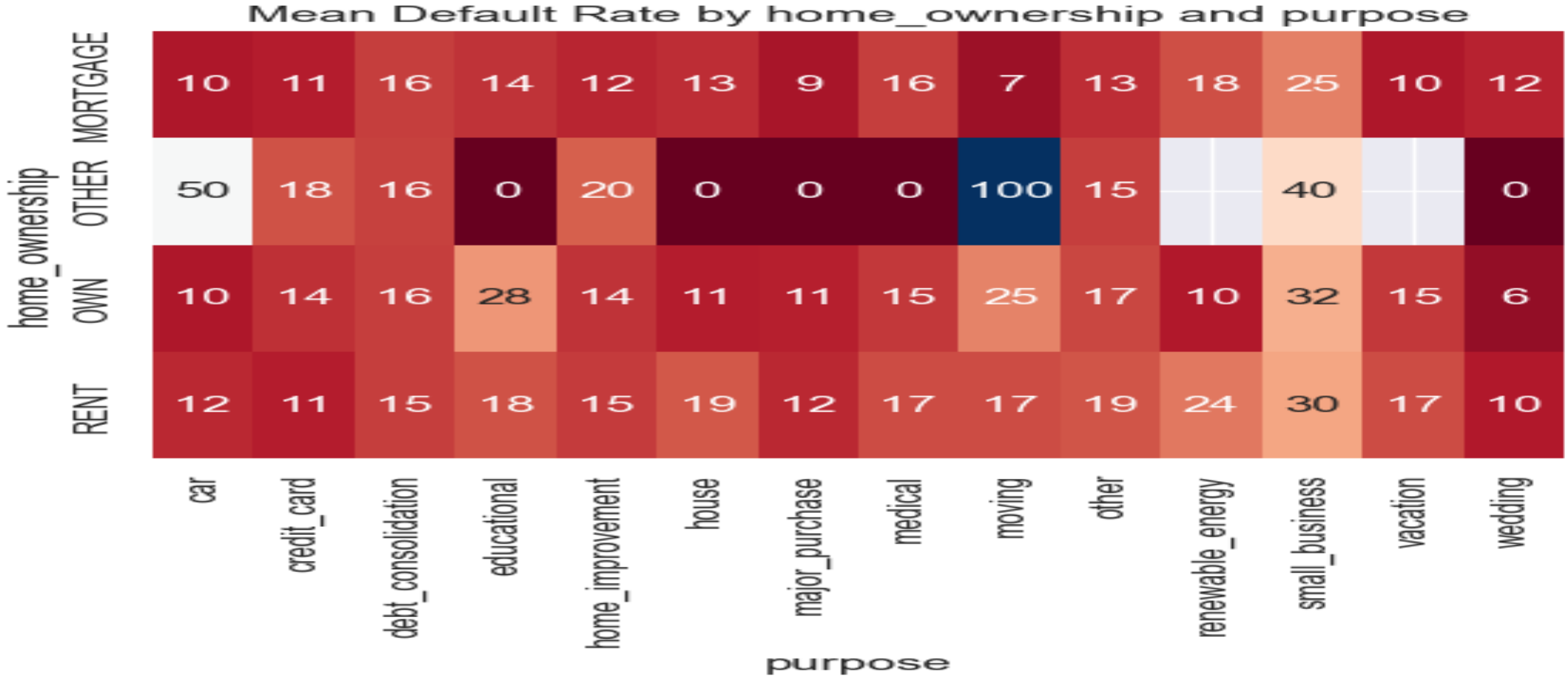


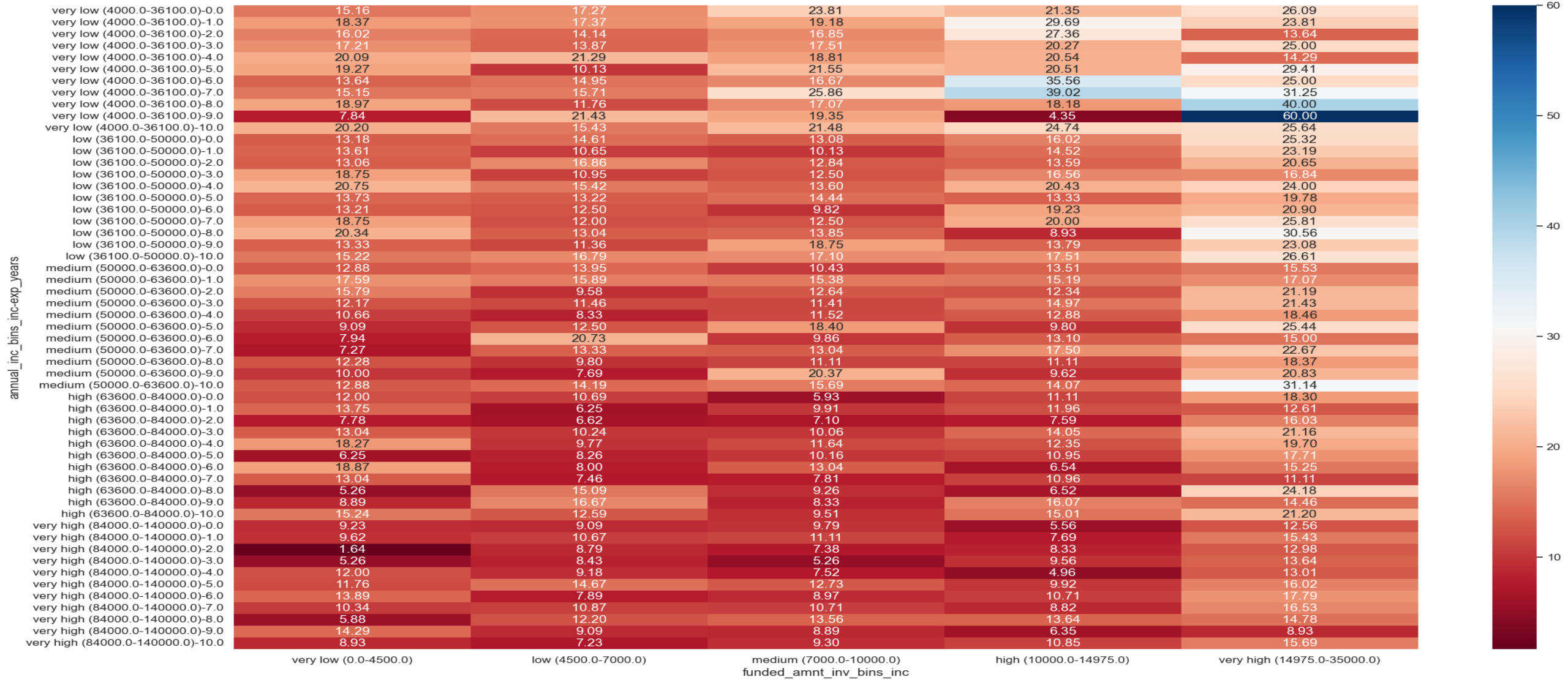

The trend over the years clearly indicates that the loan with very high interest rate have the highest probability of defaulting on the loan irrespective of the loan amount and the income of the borrower.




Non-verified loans have much lower defaults of 12% while Verified ones have a default of 16%. This might be a sign that the verification process has gaps and we are better off with third-party verification [ie Source Verified]. Non-verified loans are possibly already known to be safe bets and hence have an expected lower defaults. These are probably reliable and known customers.




From the Bivariate analysis of home ownership vs purpose heatmap: we see 100% default with the combination of Moving as Purpose and Other as home Ownership and also a default of 50% with a combination of Car as Purpose and Other as Home ownership. Small Business as purpose was a high rate of default in the univariate analysis. But now we can see that in this category even customers who own a home also default to a high rate of 33%. The purpose code of Renewal Energy with customers who are on Rent has a high default rate of 25%.




Borrowers having 7+ years of experience with very low annual income (<36000) are mostly likely to default high-very high range (>10000) loans.

# Prescriptive Insights

1. Grades F and G have more than 32 % defaults. Loans in this category should be offered to known and reliable customers.
2. Loans with 60 months term default with 25% rate. The company should offer only shorter-term loans of 36 months where the default rate is only 10%.
3. Loans where the funded amount is more than 15000 have a high default rate of 18%. Investors are advised to limit their funding to below 15000.
4. Loans with Purpose code of ‘Small Business’ default with a rate of 27%. The company should not offer loans to fund Small Business.
5. Loans with high interest rate are having higher chances of defaulting specially for the higher range of loan amount.
6. Non-verified loans have much lower defaults of 12% while Verified ones have a default of 16%. This might be a sign that the verification process has gaps and we are better off with third-party verification [i.e. Source Verified].