Lending Club Case Study

Exploratory Data Analysis Project

Manupriya Srivastava 12th February 2024

# Lending Club Case Study

Exploratory Data Analysis

- **Study of Data Dictionary** to understand the columns

- **Size of Data** - (39717, 111)

- **Data Cleaning**

  - Multiple columns have null values. Dropping these results in a size of (39717,57)

  - Many columns have single values. Dropping these results in 48 columns

  - Many columns are irrelevant in studying about loan defaulters. These are dropped

    - id, zip code etc

    - post approval attributes

  - Rows related loans that are currently running are removed

# Lending Club Case Study
Exploratory Data Analysis

- **Checking for missing values** shows that 2 columns still have some missing values

  - Replaced with mode values

- **Standardising the Data**

  - % signs removed from some columns and values made numerical

  - 'year' , '<', '+', removed in 'emp_length' and column values made numerical

- **Checking for Outliers & Removing Outliers where required**

  - Outliers found in loan_amnt, funded_amnt and annual_inc

  - Outliers after 90th percentile removed from annual _inc

  - Outliers in loan_amnt and funded_amnt are continuous and left as is
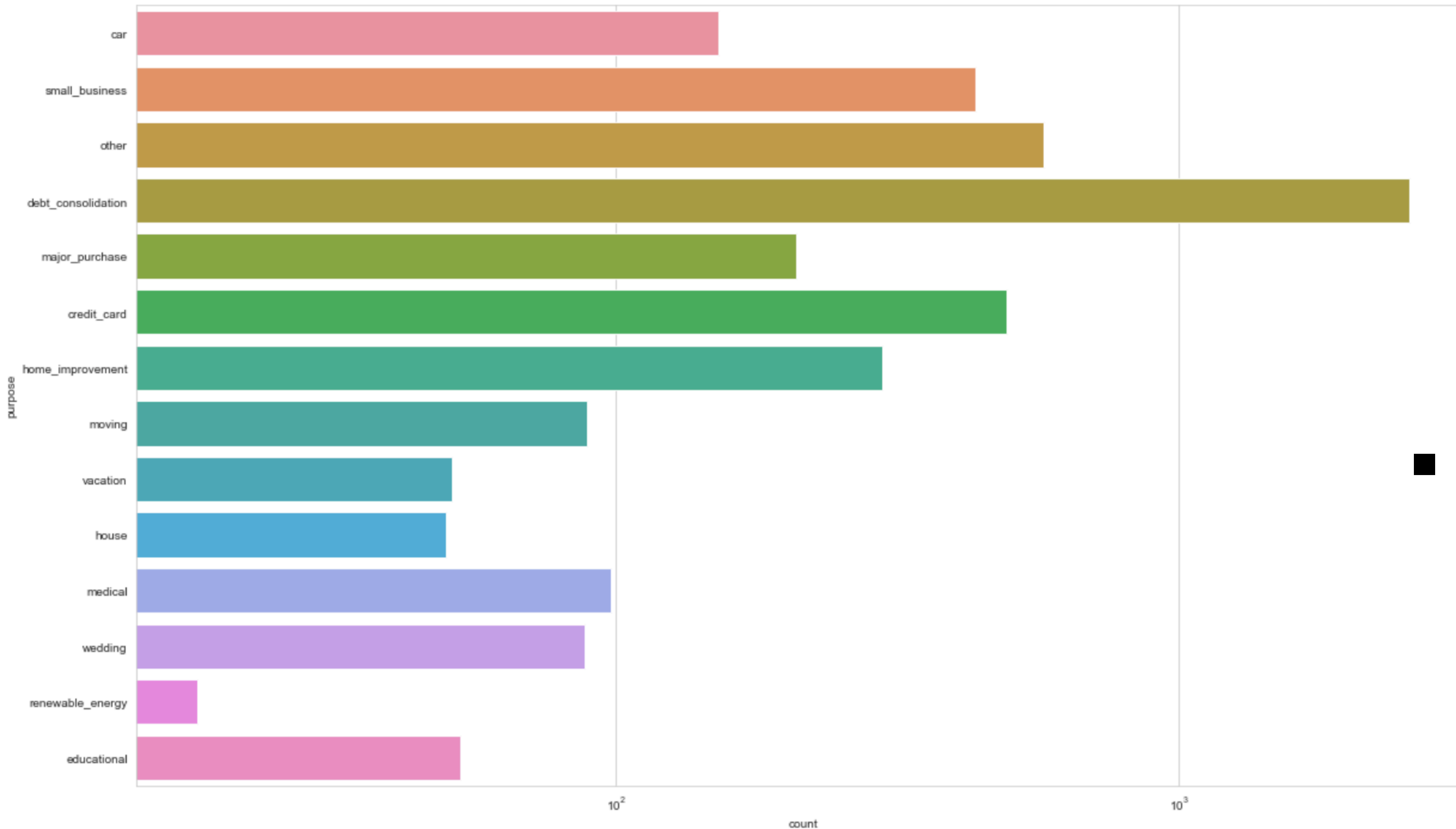
Impact of Verification status

Impact of Home ownership
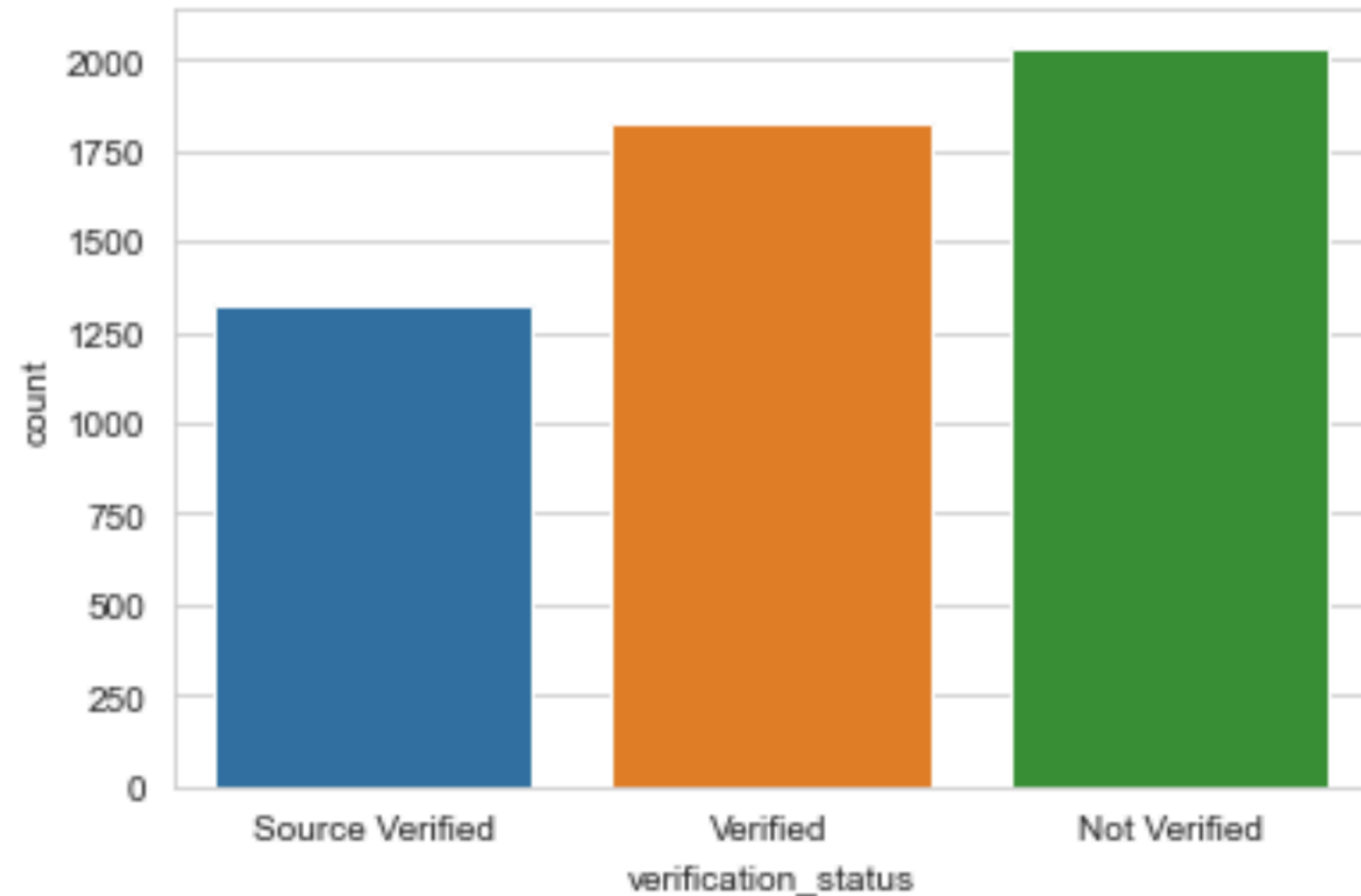
- More than 2000 of the defaulters were 'Not Verified'

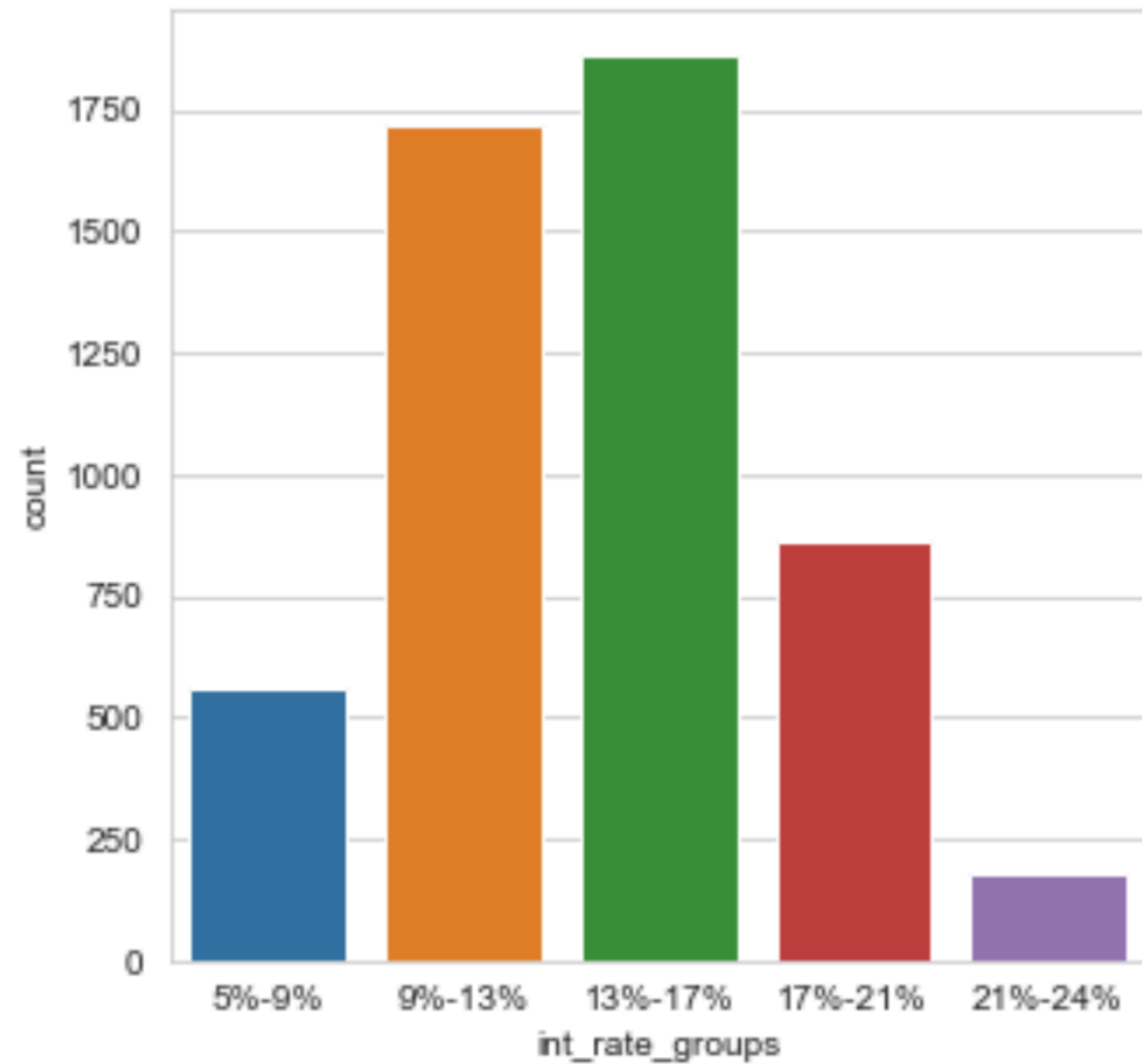- However, even a large no. of those verified were also defaulting
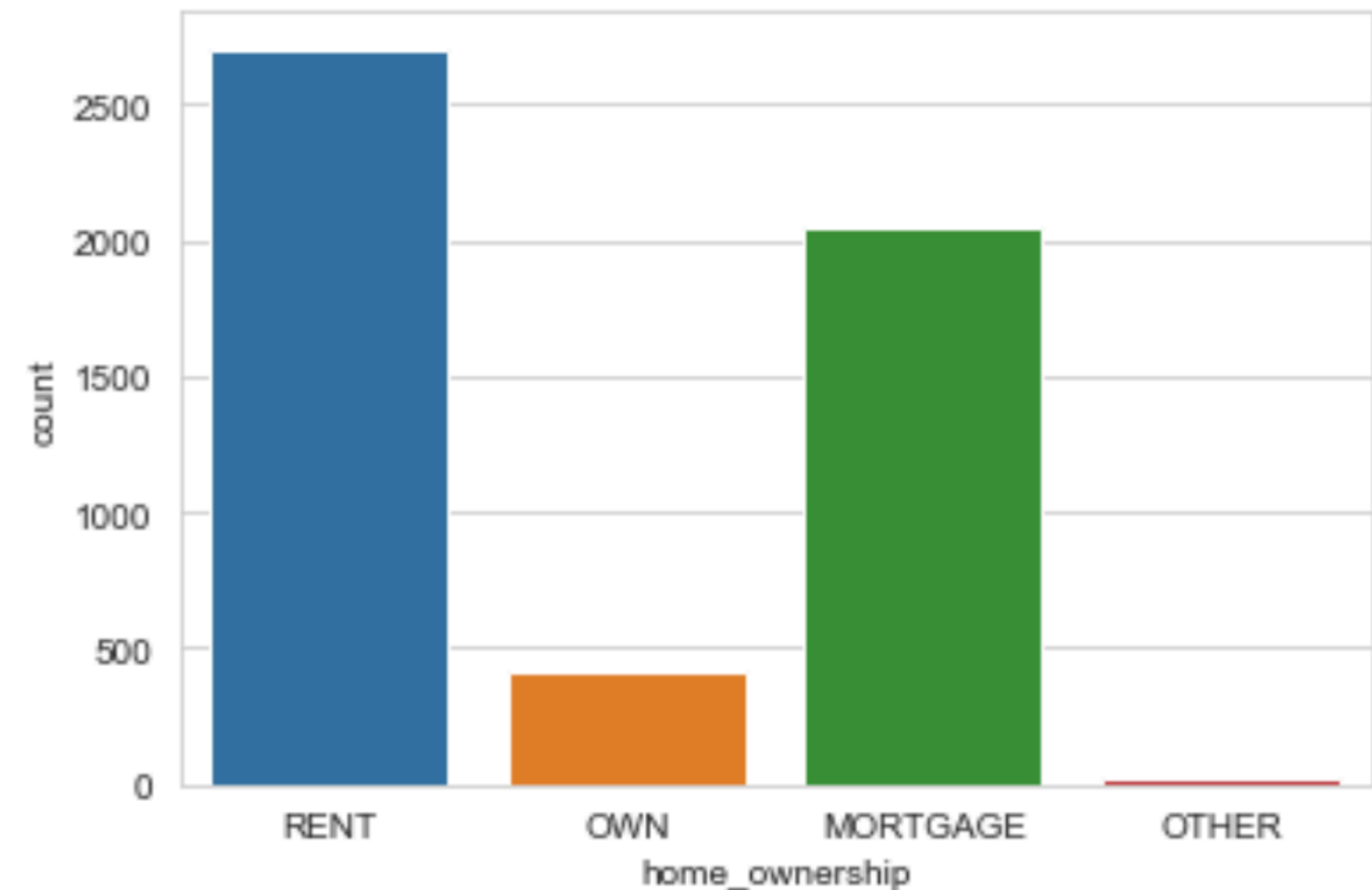
- Less than 500 of the defaulter accounts belonged to homeowners

- More than 2500 of the defaulters were renters

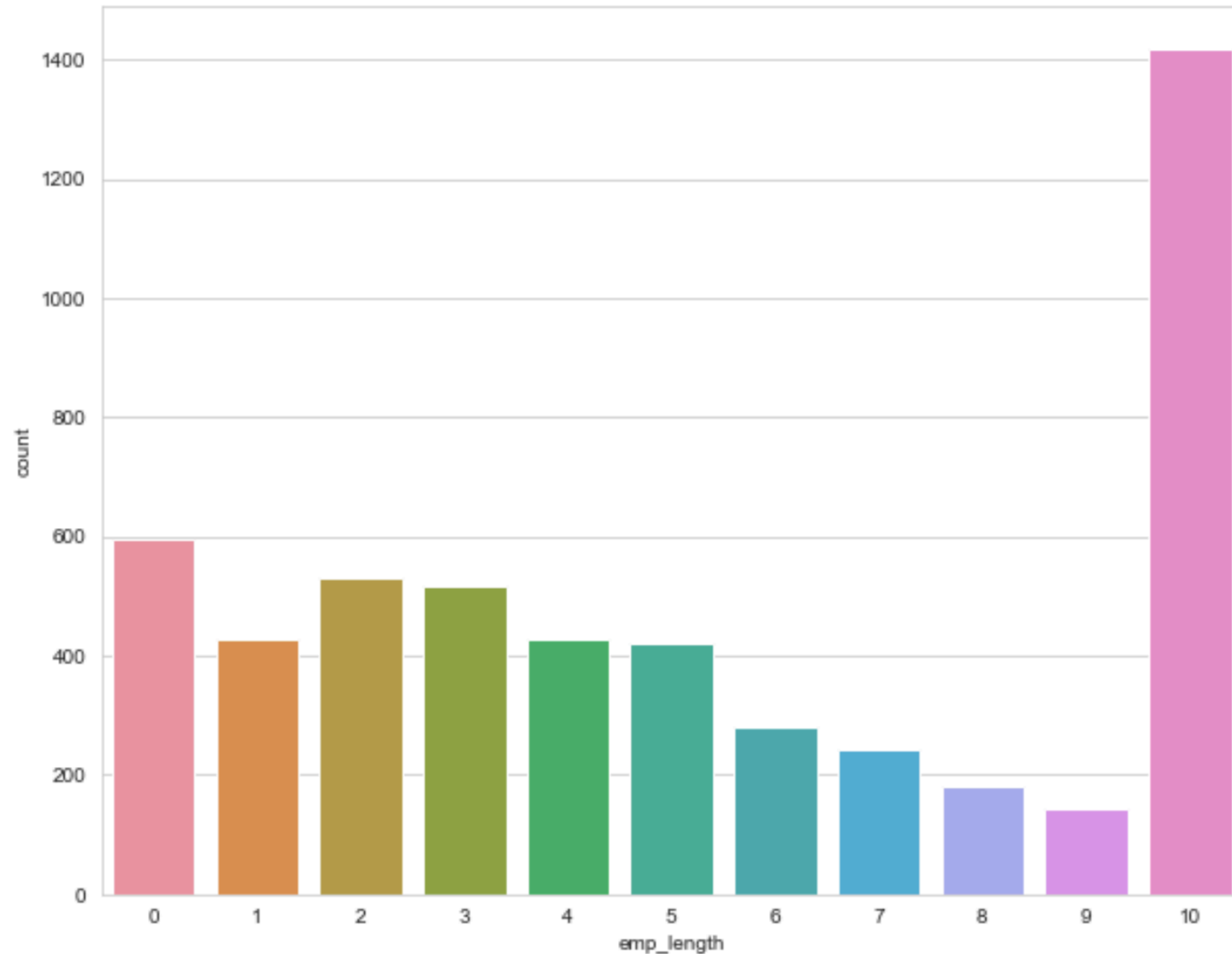# Impact of Purpose of loan specified by Defaulters

- A huge no of defaulters specified their 'Purpose' as Debt Consolidation.
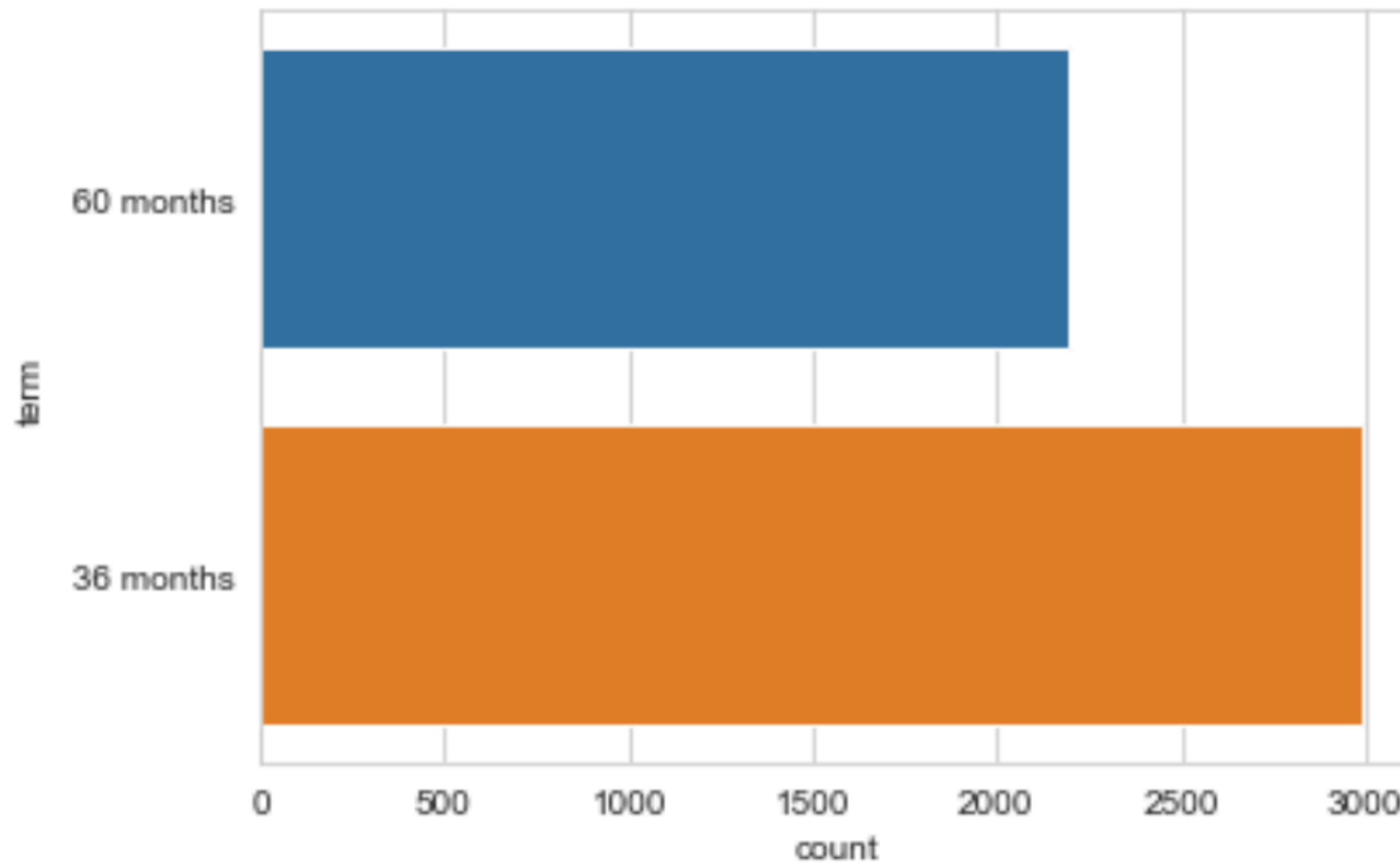
Impact of Interest rate

- Most of the defaulters had interest rates in the range 13%-17%

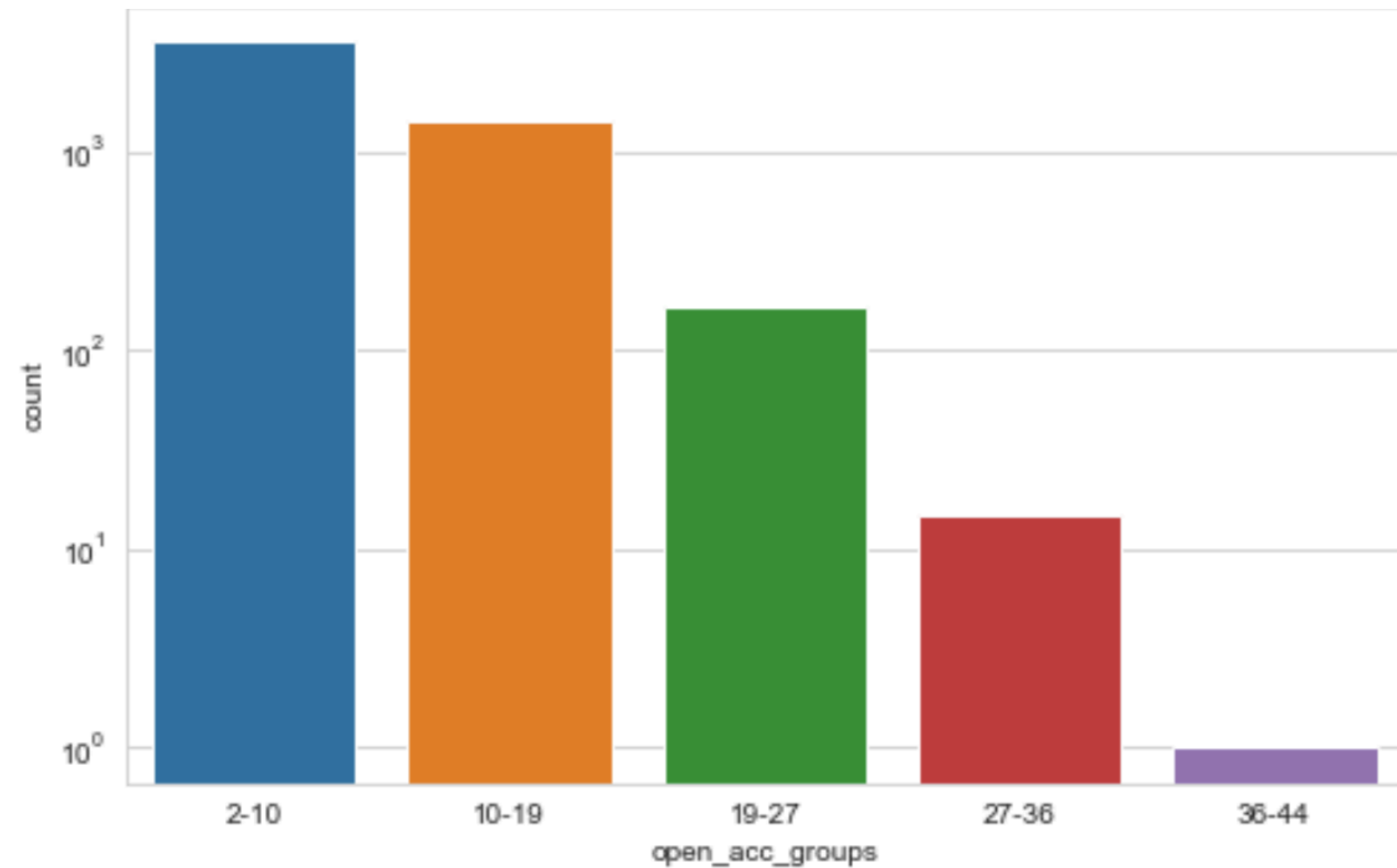- Least defaulters in the range 21%-24%

Impact of Length of employment

■ A huge no of defaulters had 10 years+ service
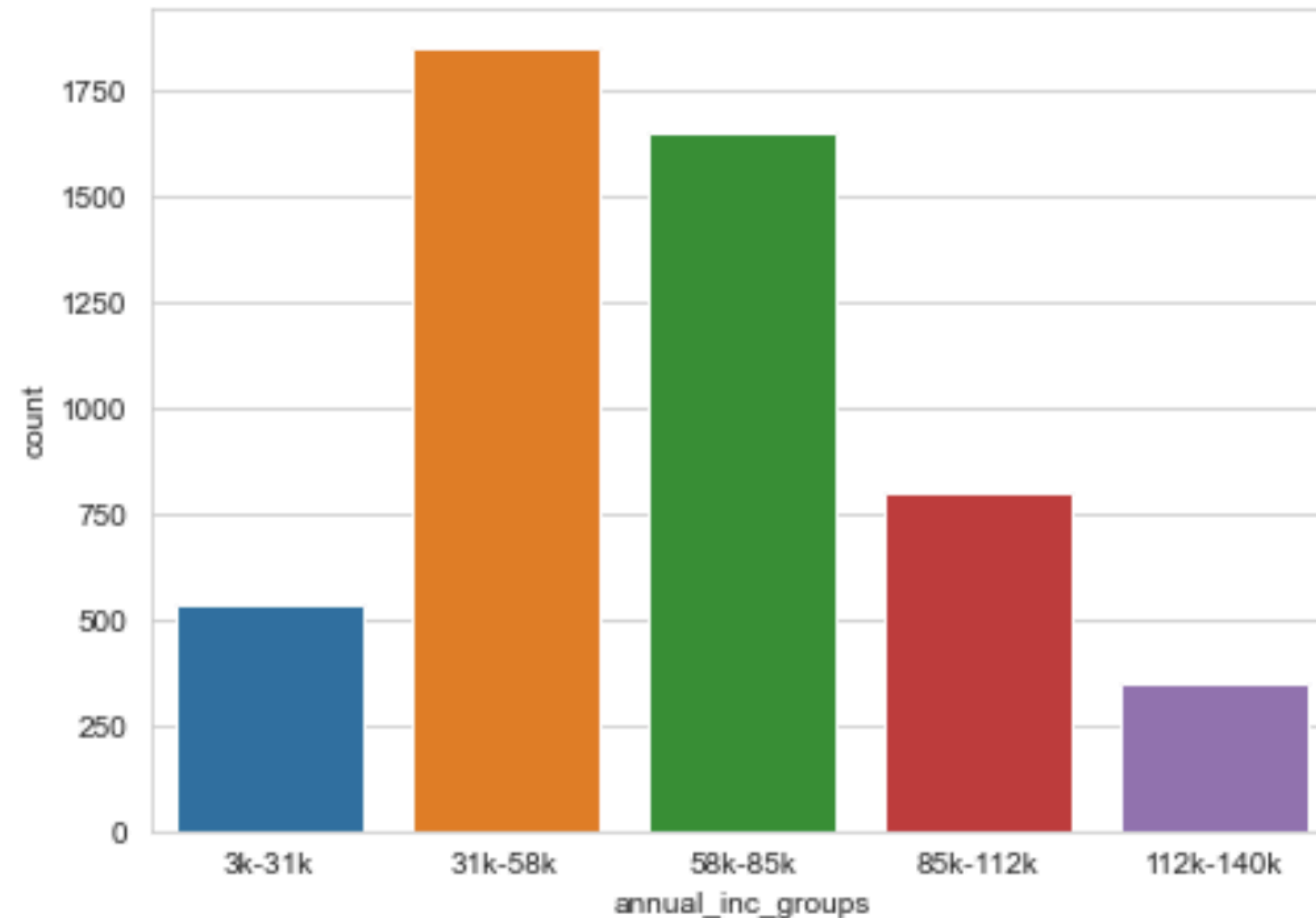
Impact of Loan term

- There were fewer defaulters with longer loan term of 60 months
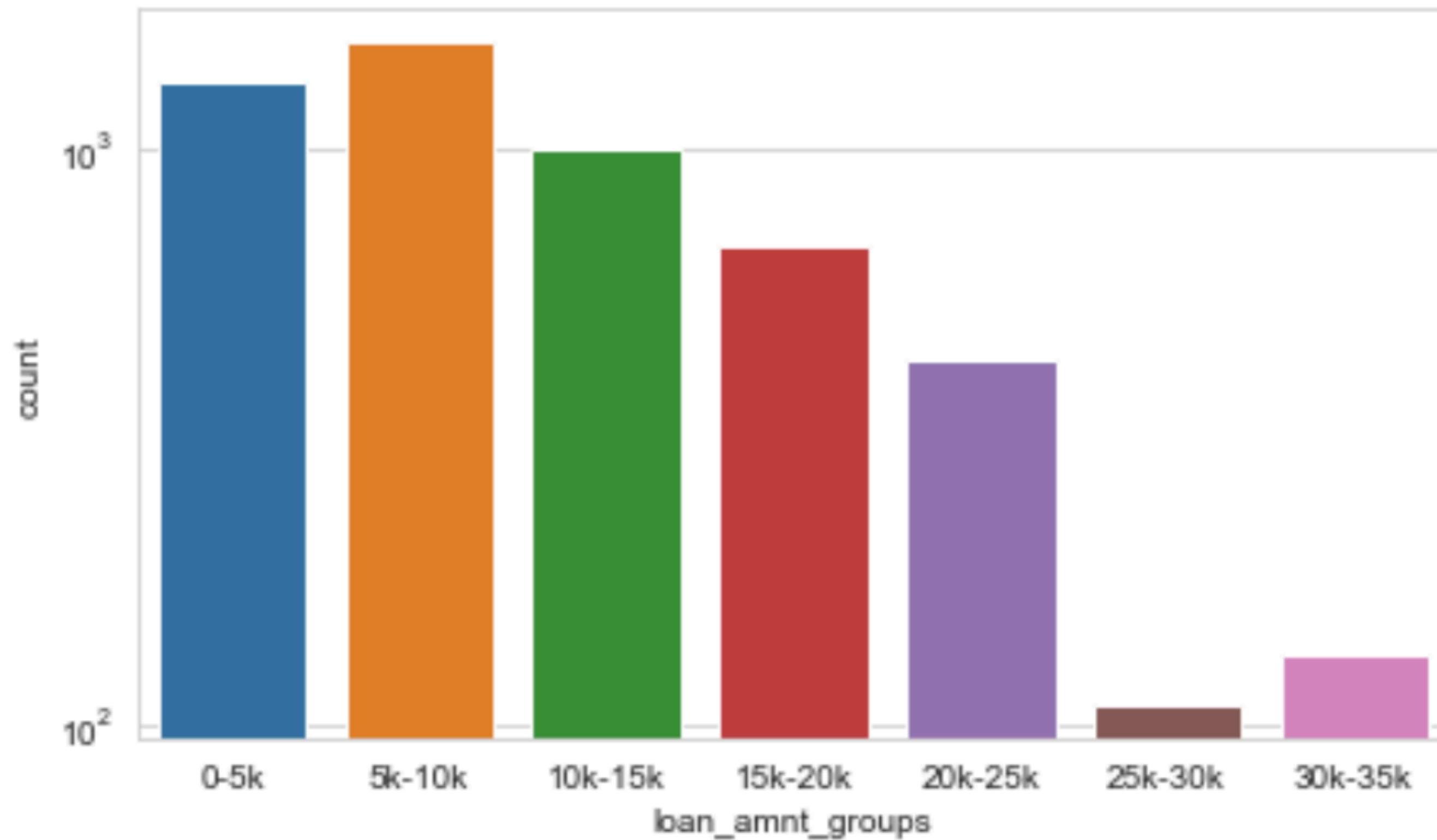
Impact of No of Open Accounts

- There were fewer defaulters with no of open accounts in the range 36-44
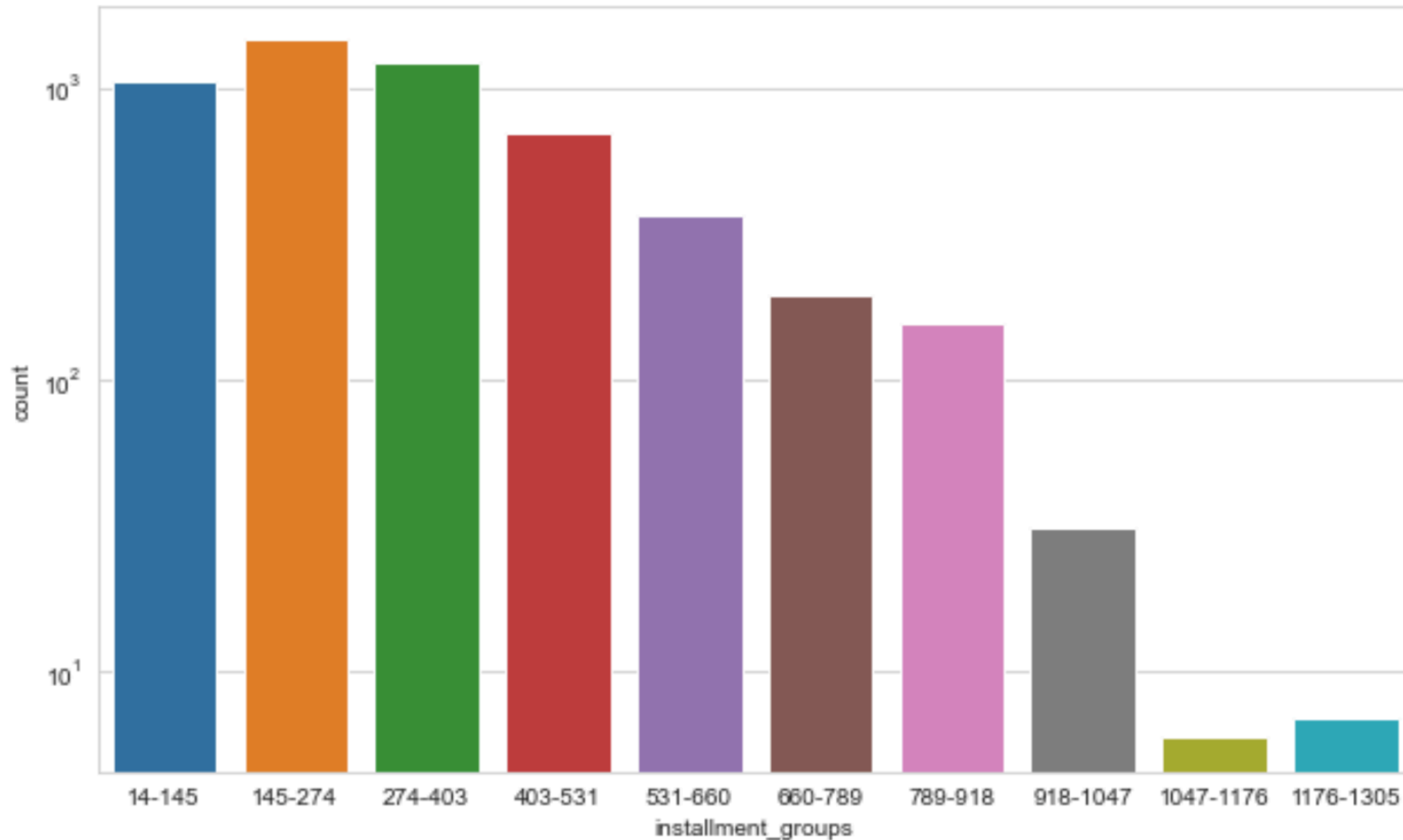
Impact of Income

- Largest no of defaulters were in the income range of 31k-58k

- Lowest no of defaulters in the income range of 112k-140k

- Even the Income range of 3k-31k had the second lowest defaults

Impact of Loan Amount

- Largest no of defaulters were in the loan range of 5k-10k

- Lowest no of defaulters in the loan range of 25k-30k

Impact of Instalment Amount

- Largest no of defaulters were in the instalment range of 145-274

- Lowest no of defaulters in the instalment range of 1047-1176
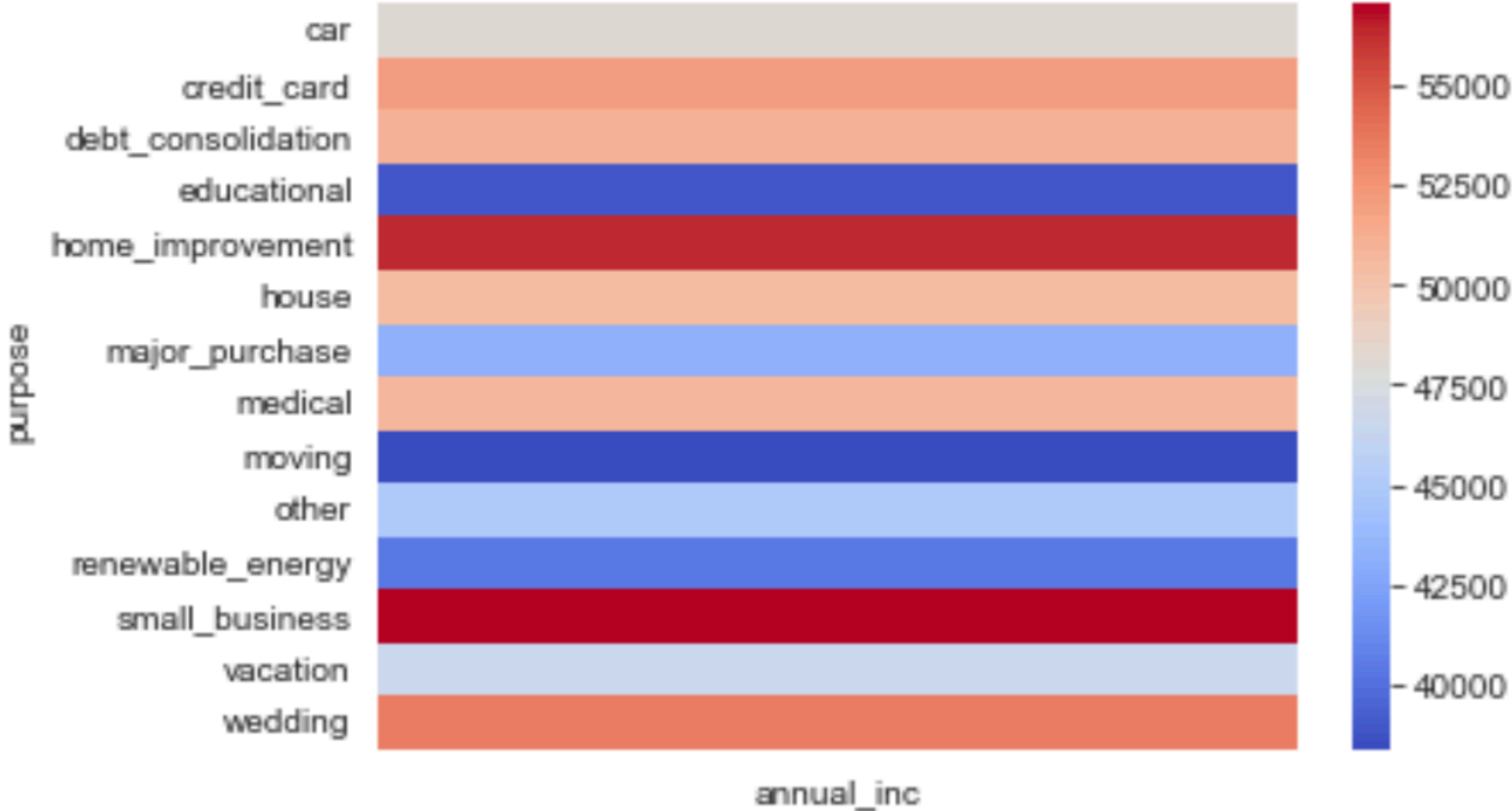
Purpose vis-a-vis Median Loan Amounts of Defaulter

purpose	loan_amnt
car	6000
credit_card	11100
debt_consolidation	12000
educational	6000
home_improvement	9700
house	9675
major_purchase	7000
medical	6000
moving	4500
other	5375
renewable_energy	4100
small_business	12000
vacation	5000
wedding	8400
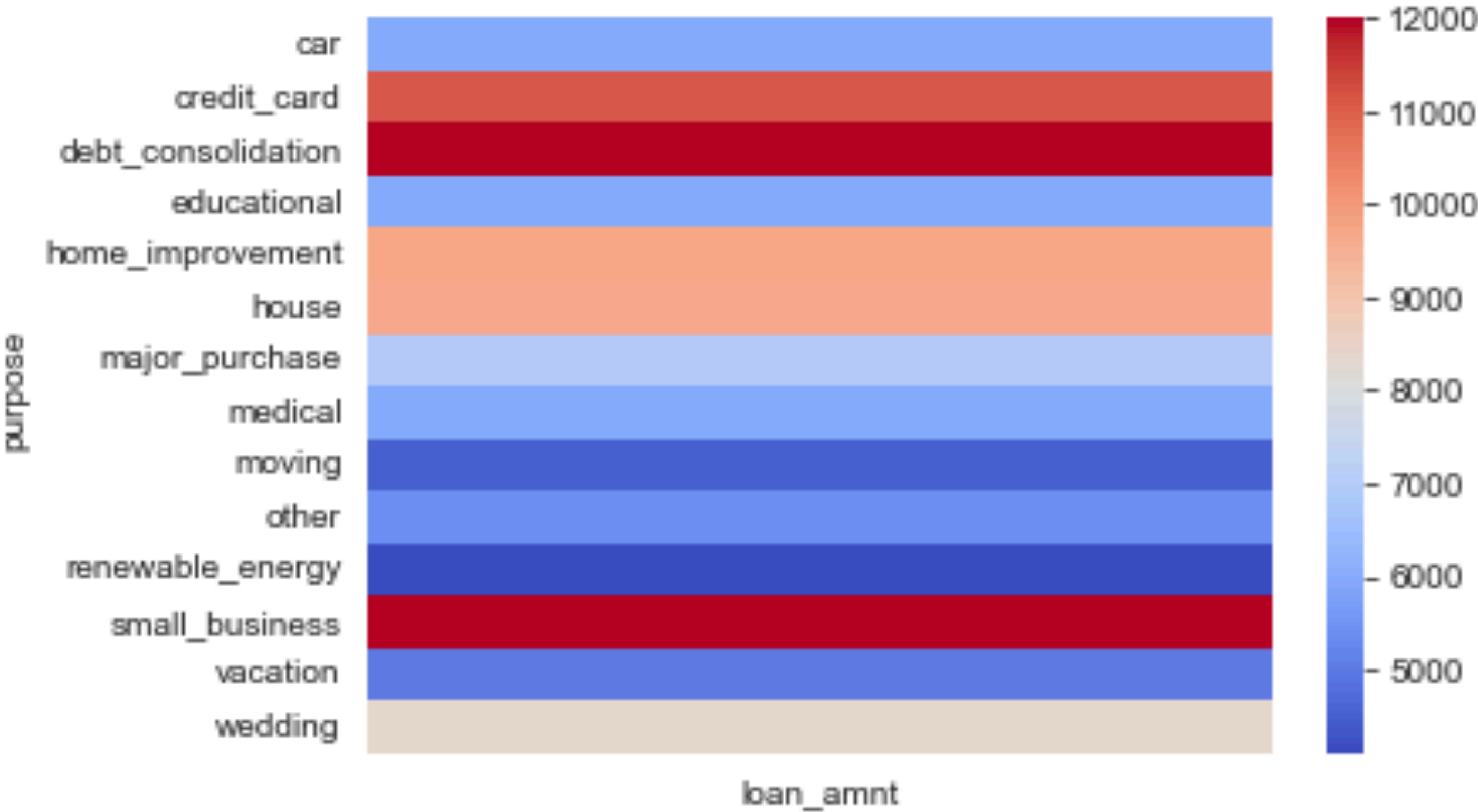
- Largest median loan amount of defaulters taken for purposes like debt consolidation and small businesses

- Smallest for Renewable energy

Purpose vis-a-vis Median Annual Income of Defaulter

| purpose	annual_inc
car	48000.0
credit_card	52000.0
debt_consolidation	51000.0
educational	39000.0
home_improvement	56300.0
house	50500.0
major_purchase	43200.0
medical	50800.0
moving	38400.0
other	45000.0
renewable_energy	40500.0
small_business	57000.0
vacation	46500.0
wedding	53500.0

- Largest median annual income of defaulters taken for purposes like home improvement and small businesses

- Smallest for Moving

Issue date vis-a-vis Count and Amount of Loans of Defaulter

■ Count and total sum of defaulted loans was maximum in December 2011
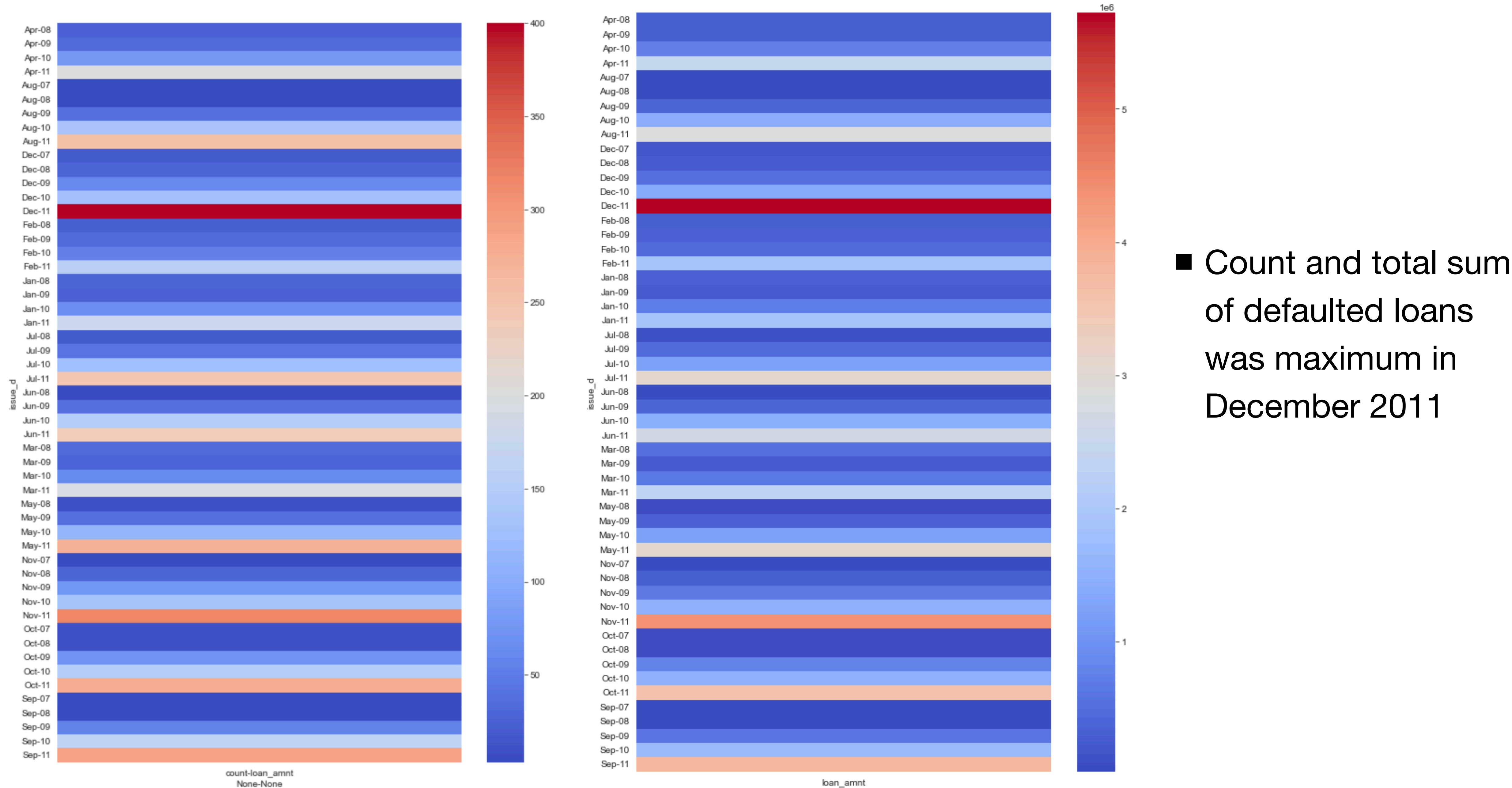
Loan grade vis-a-vis Count and Amount of Loans of Defaulter
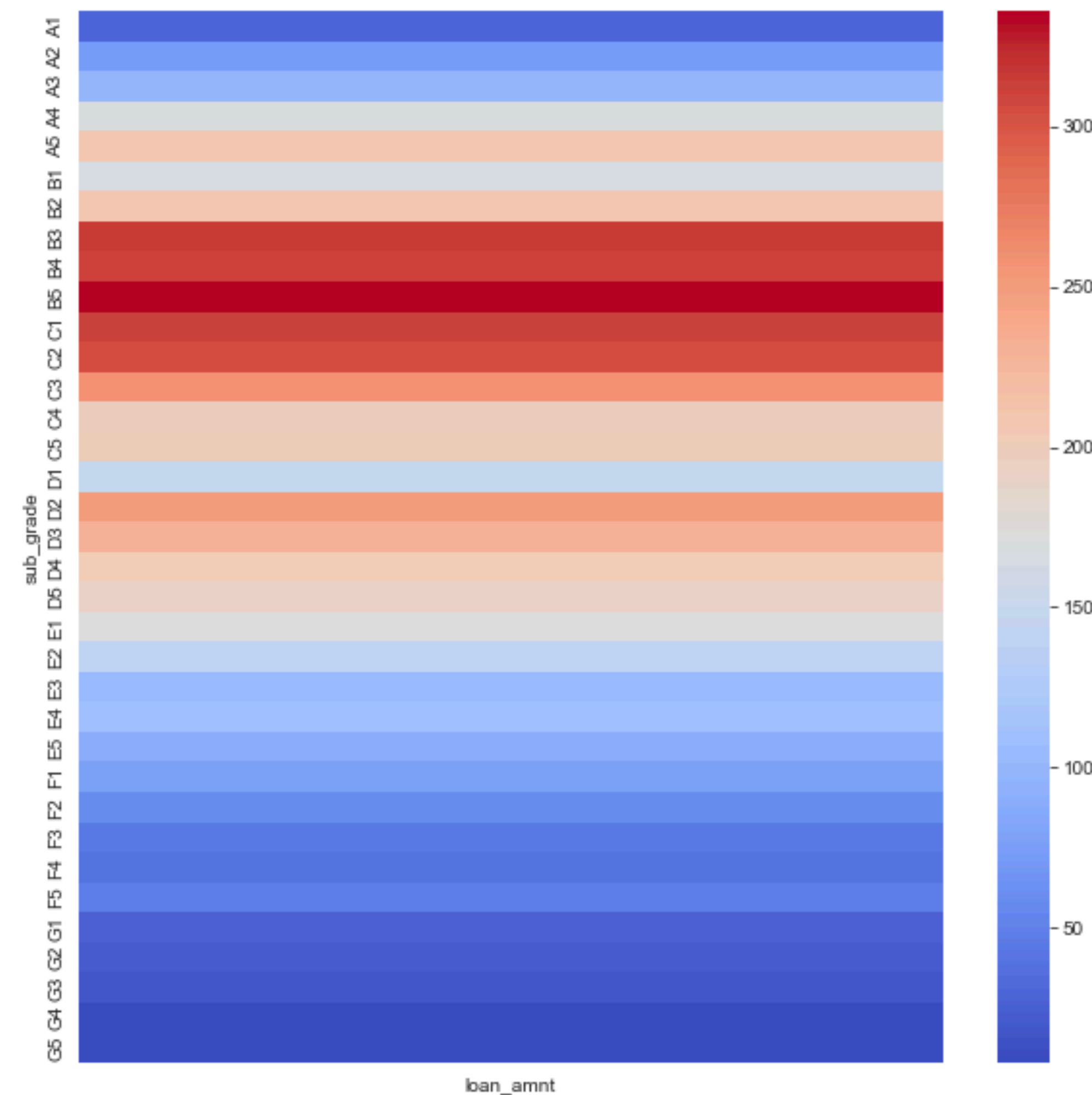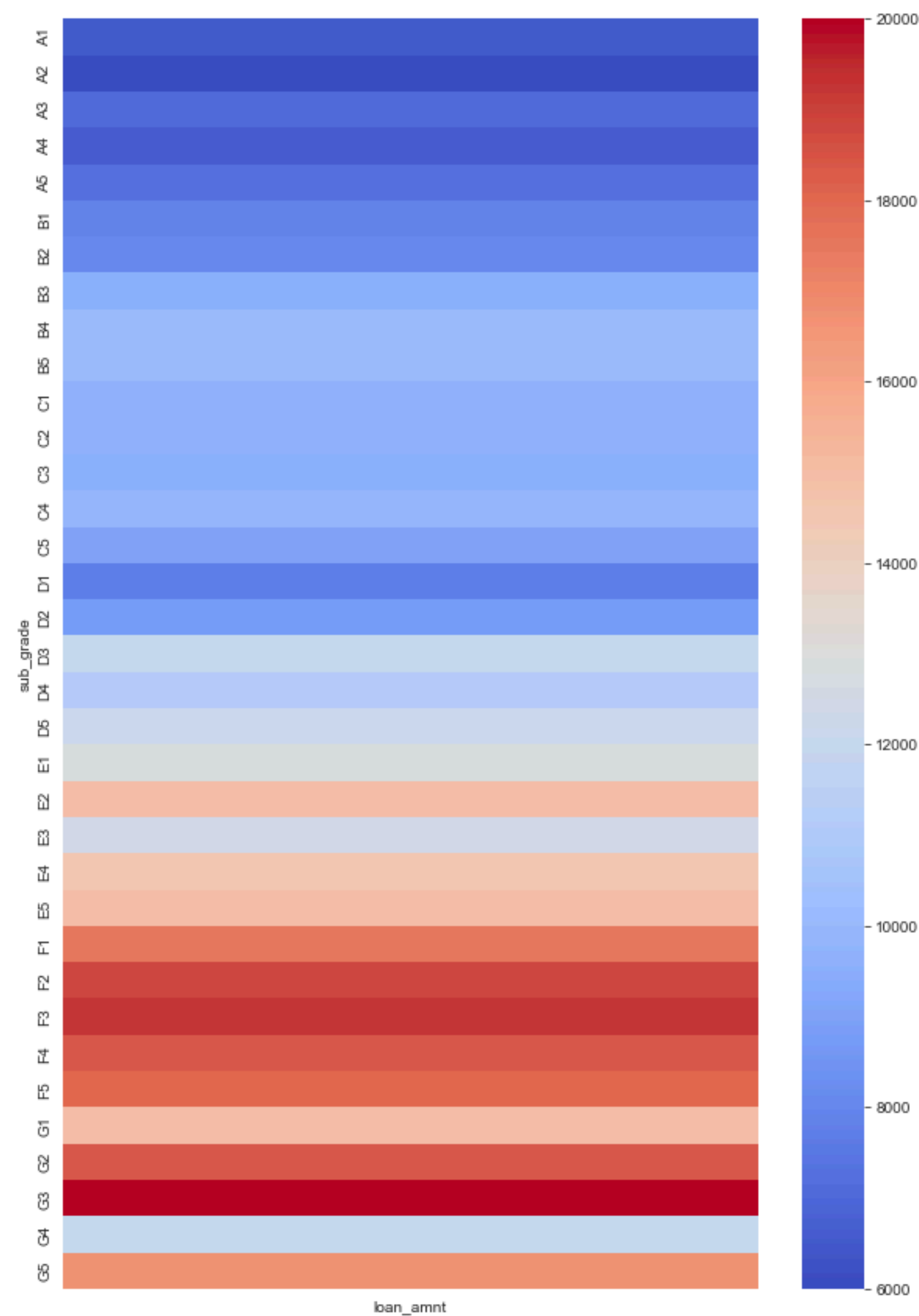
- Highest count in B4 and B5

- Total sum of defaulted loans was maximum in G3