Lending Club Case Study

Program: M.Sc. in Machine Learning and A.I.

Course: Statistics Essentials

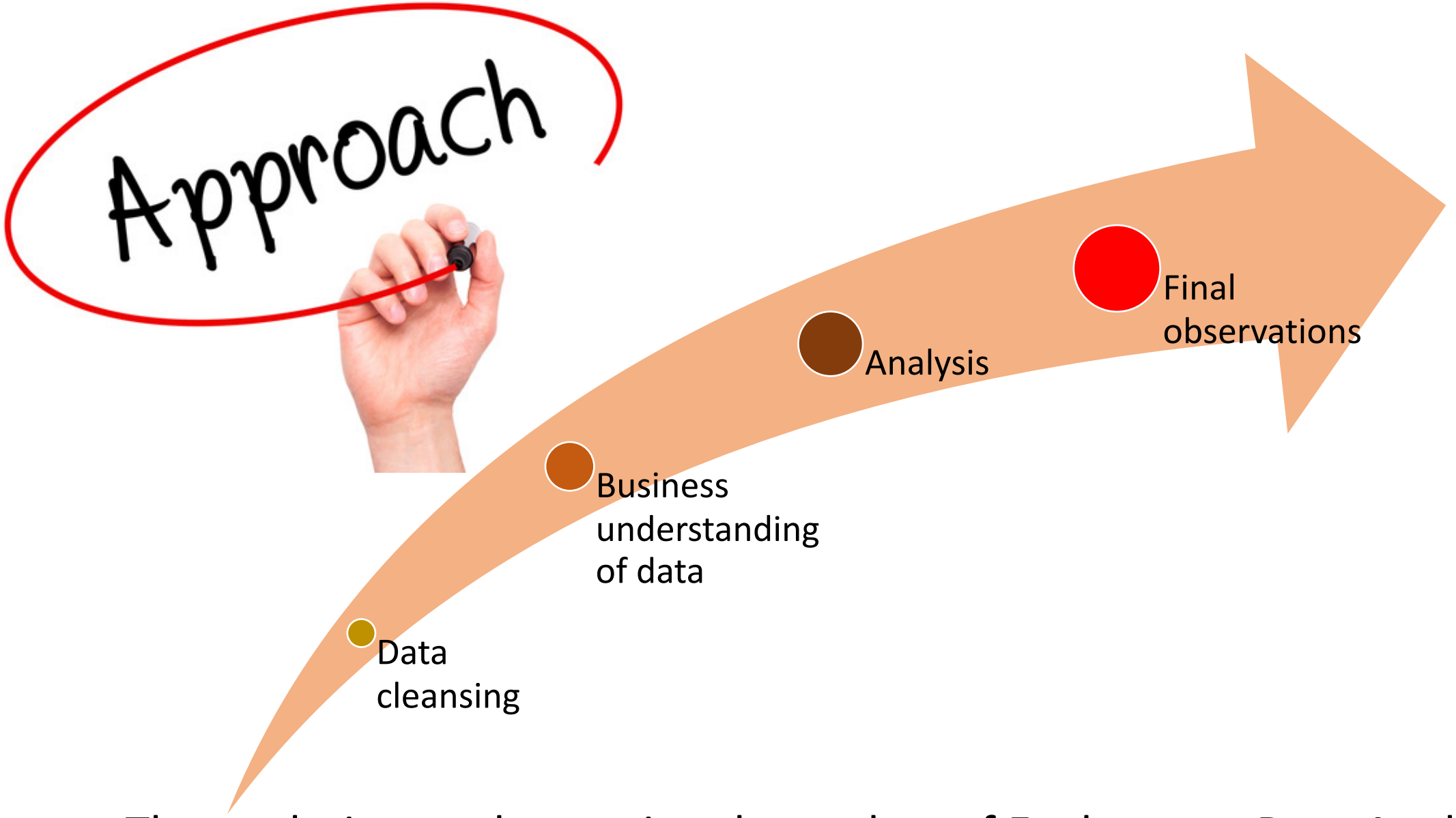
By: Zoltán Nánai

Problem statement

- „Lending Club” is a consumer finance company which specializes in lending various types of loans to urban customers
- The company wants to support its credit decision by understanding the driving factors of credit default
- The goal of the current analysis is to uncover those customer and loan attributes that can serve as an indicator for future credit loss and that are already at the time of the credit decision

Data

- **Loan.csv** that serves as the basis of the analysis contains
  - 39717 credit records
  - 111 attributes
- The file contains information about loans approved between 2007 and 2011
- Around 15% of the loans are defaulted



The analysis was done using the toolset of Exploratory Data Analysis (no ML or other modeling activity was done)

Results of data cleaning

Half of the columns were deleted:

- 56 columns were totally empty or had only handful of datapoints and have been deleted
- 14 further columns were deleted as they contained unique categorical unordered values for each datapoint or contained the very same value for every single datapoint
- Minor missing data problems were solved
- Data type transformation and / or derived variable creation was done during the analysis

Business understanding of Data

The attributes were grouped into 3 categories manually:

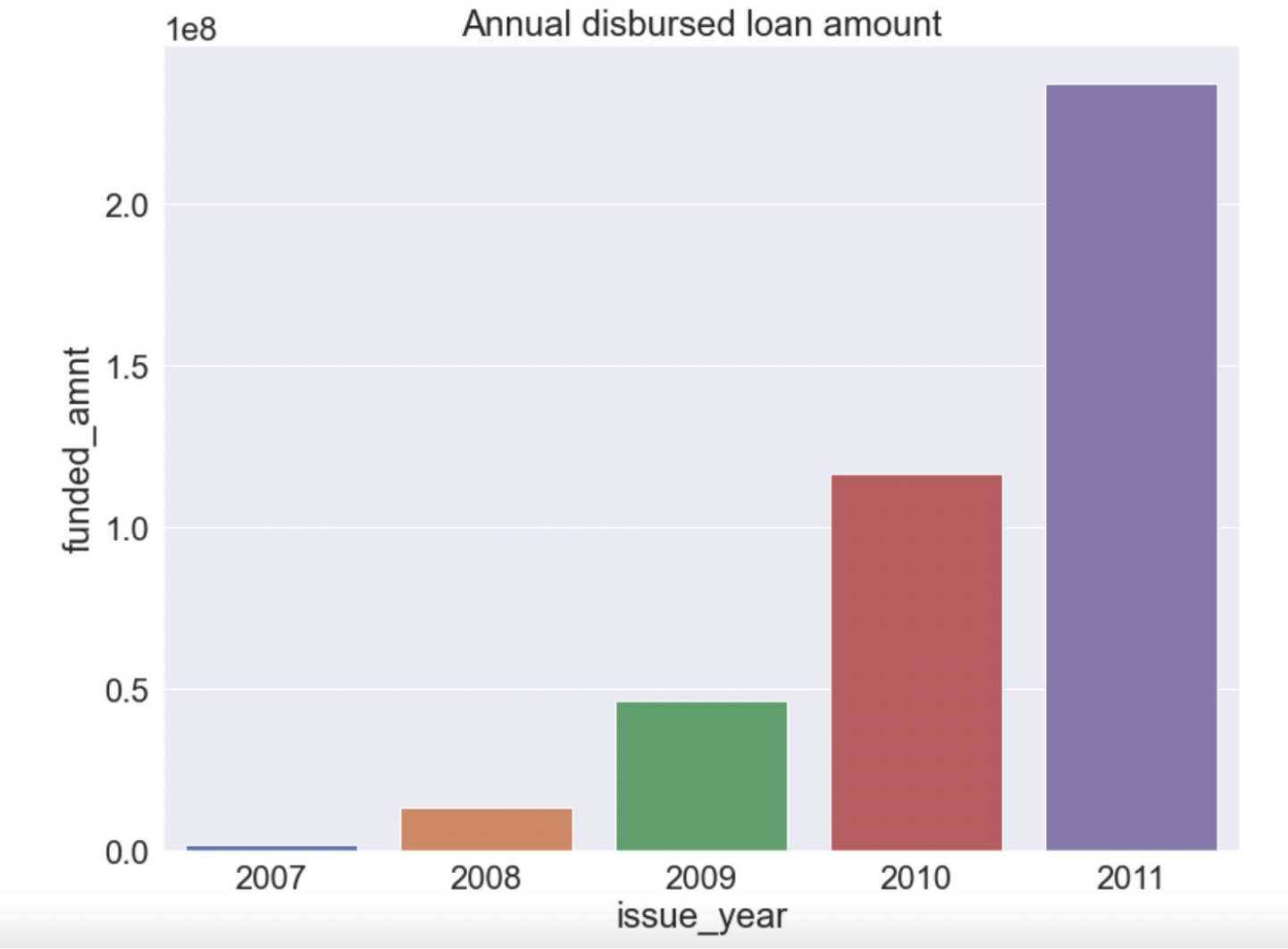
- ’Pre-Decision’ data: Information available before the credit decision (Demographic information plus loan application information)
- Behaviour information available for approved customers
- Leaking information available for defaulted customers only

Business understanding of Data II.

The analysis focused on the information set available for ‚new-to-bank' customers (first category) as the goal of the exercise was to find driving factors for credit default that can be used during the credit decision.

Lending Club business overview I.

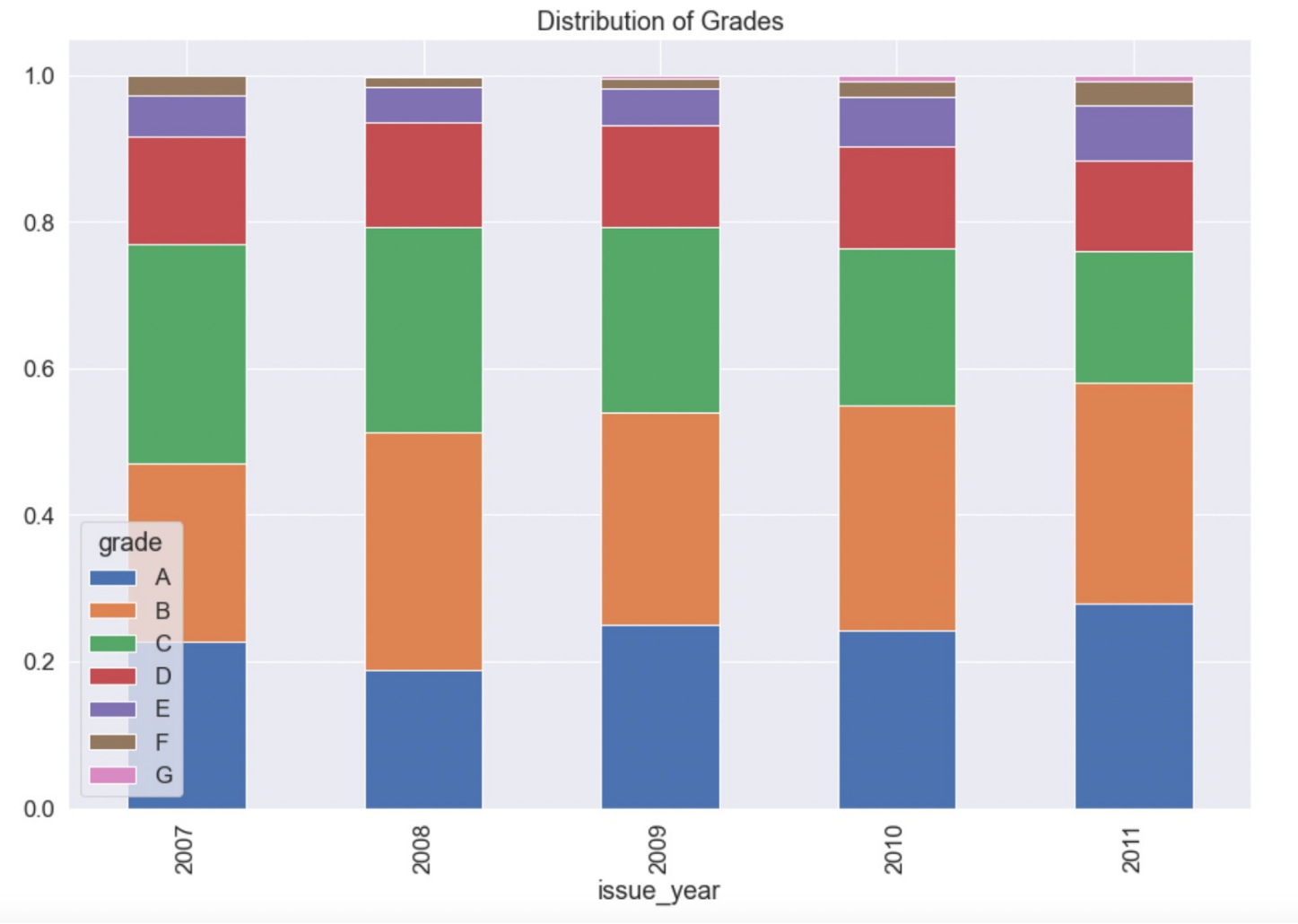
- The Lending club extended its lending activities almost exponentially
- From 2007 to 2011 it increased the annual disbursed loan amount from 2.1 million to 237.4 million



Lending Club business overview II.

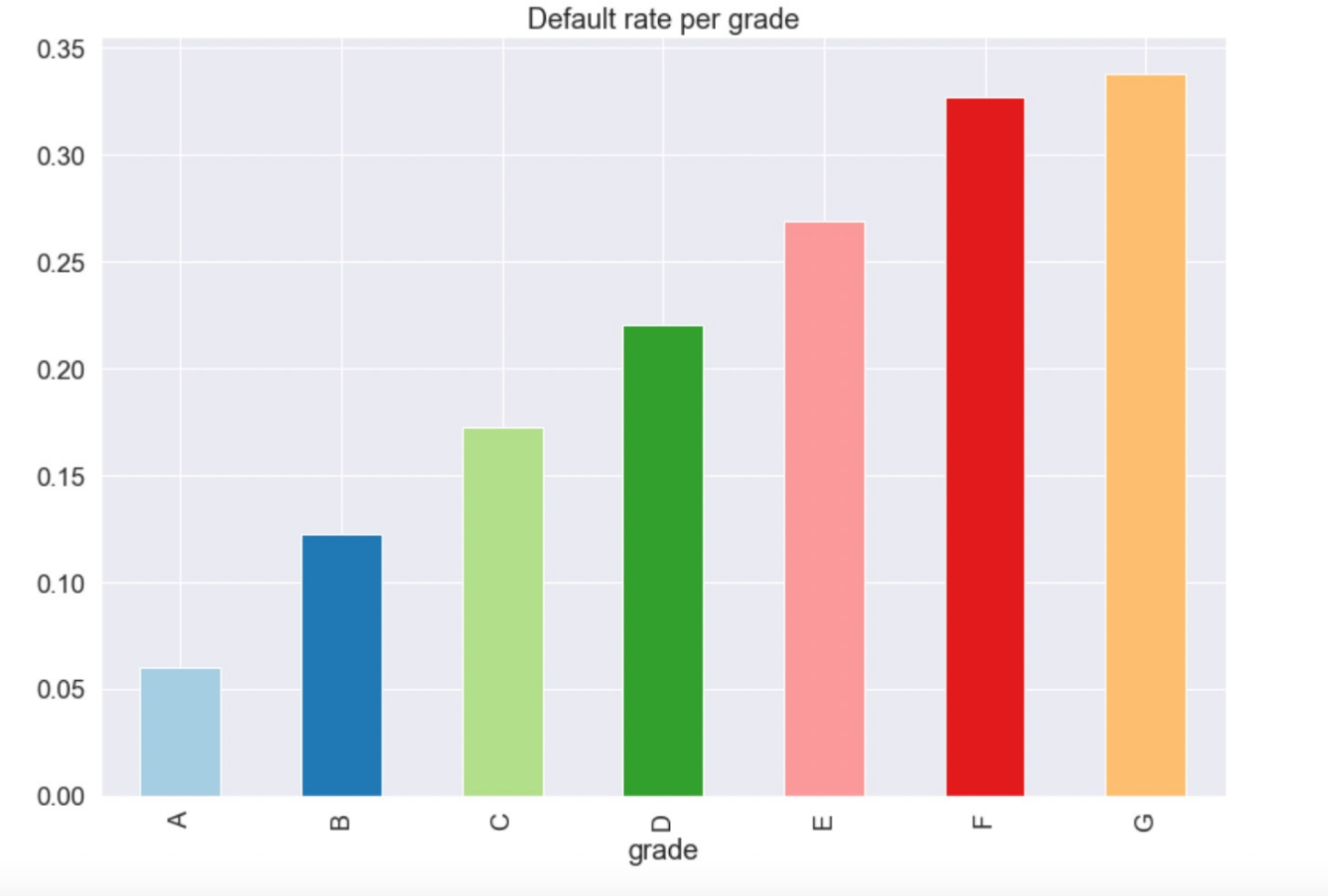
This expansion is impressive, did however come with a cost.

It might seem that the portfolio is under control as the distribution of the grades is unchanged over the years (with a small increase on the edges (best and worst categories) and decrease in the mid range)...



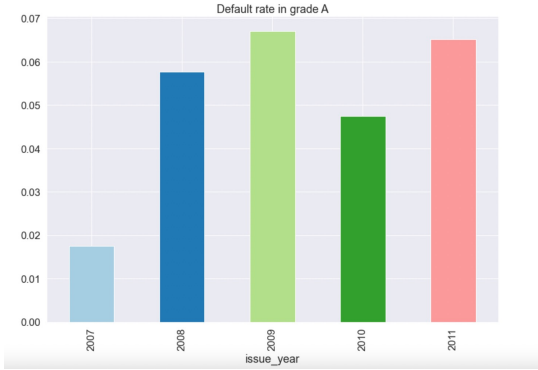
Lending Club business overview lll.

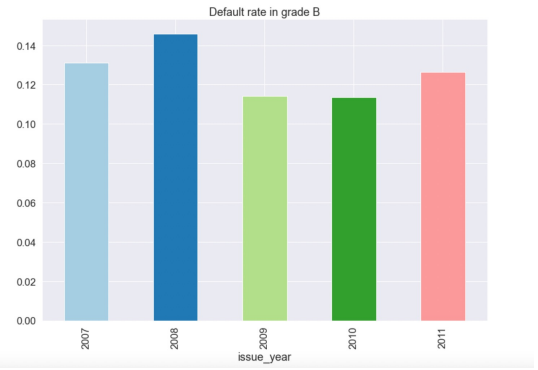
...and it might seem that the grades represent the risk categories very well...

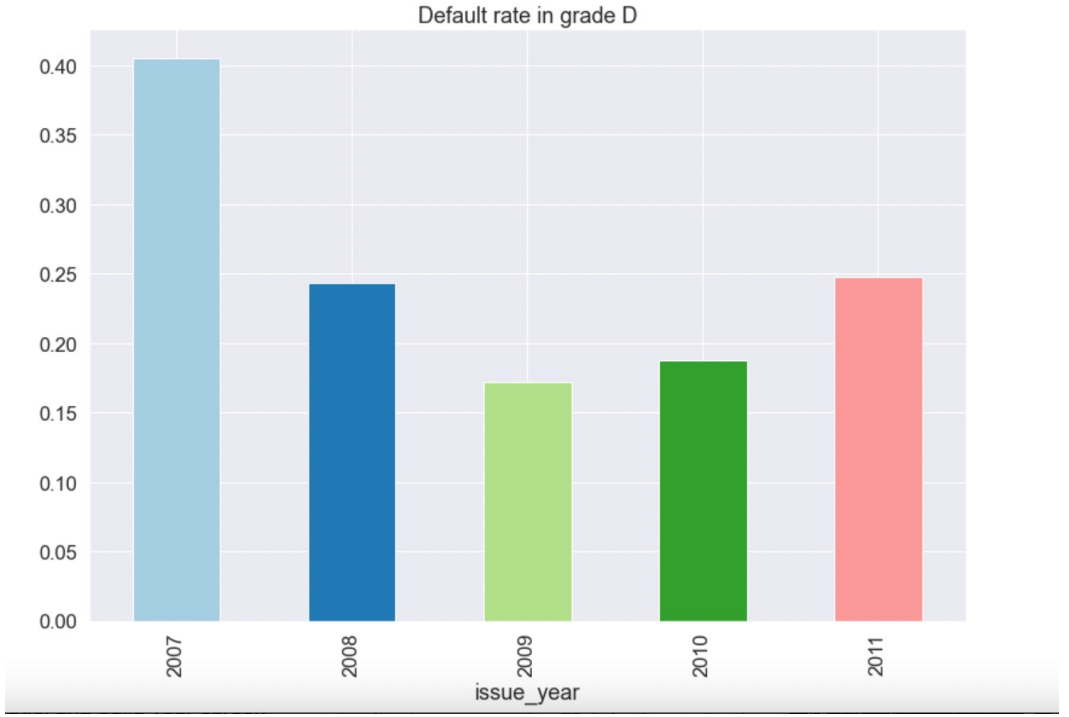


Lending Club business overview IV.

... but in fact the grading system stared losing its predictive power somewhere around 2009 and the default rate started to increase in all grades – with different pace.



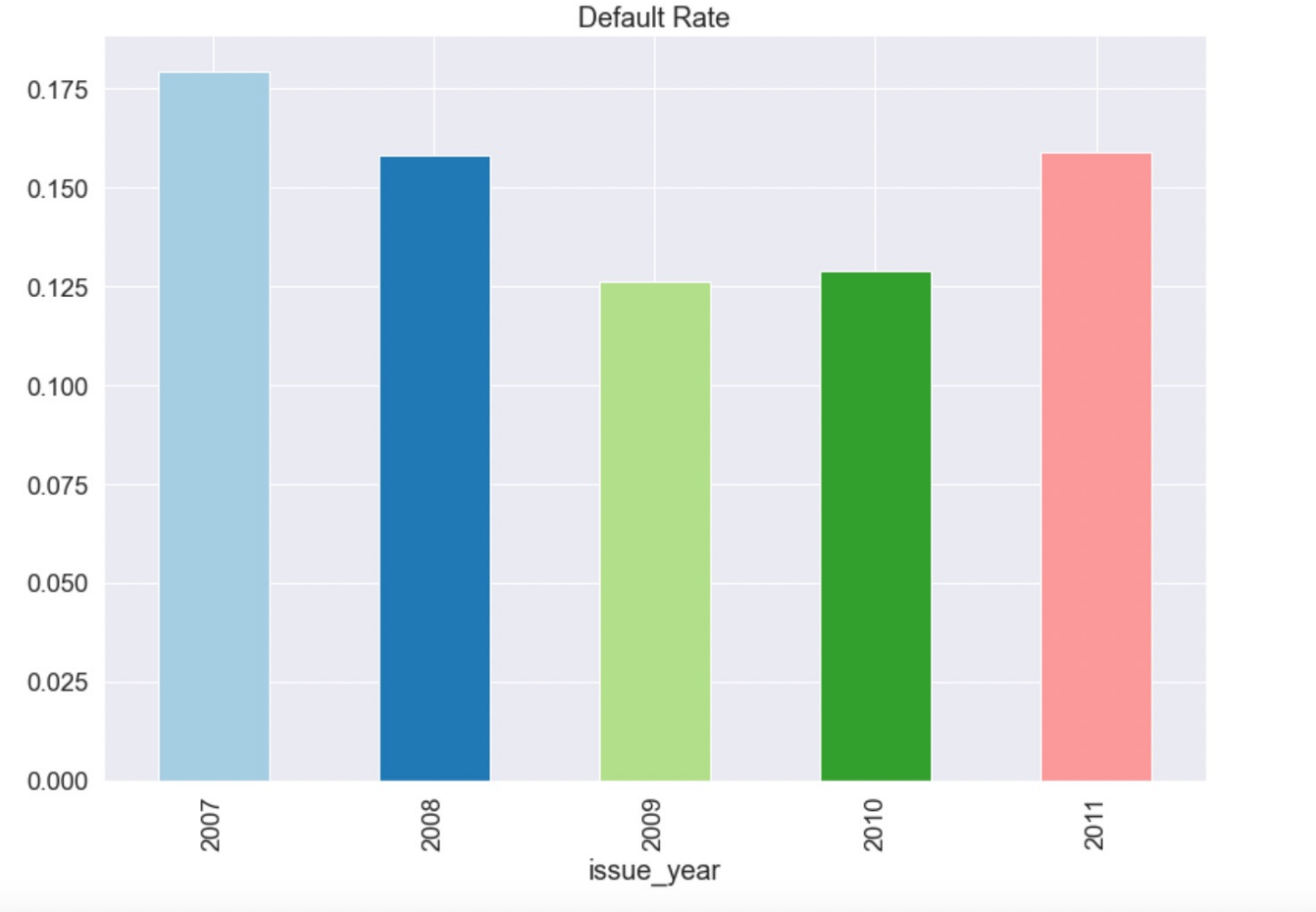




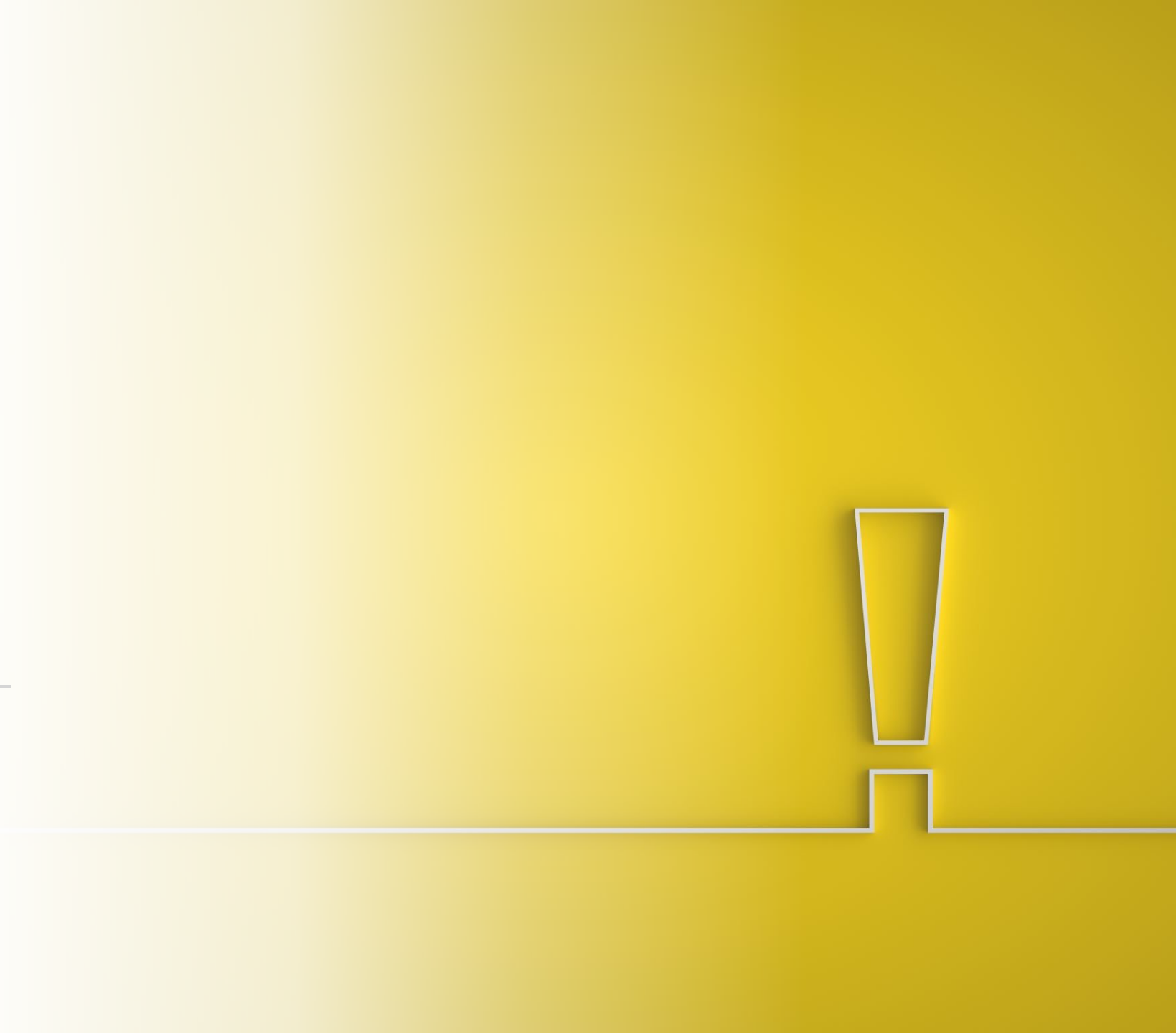
Lending Club business overview V.

...and as a result of the expansion and changes in the power of grading system the portfolio quality – after a continuous improvement for three years – started to decline

So indeed, there is a need for improvement of the credit approval process



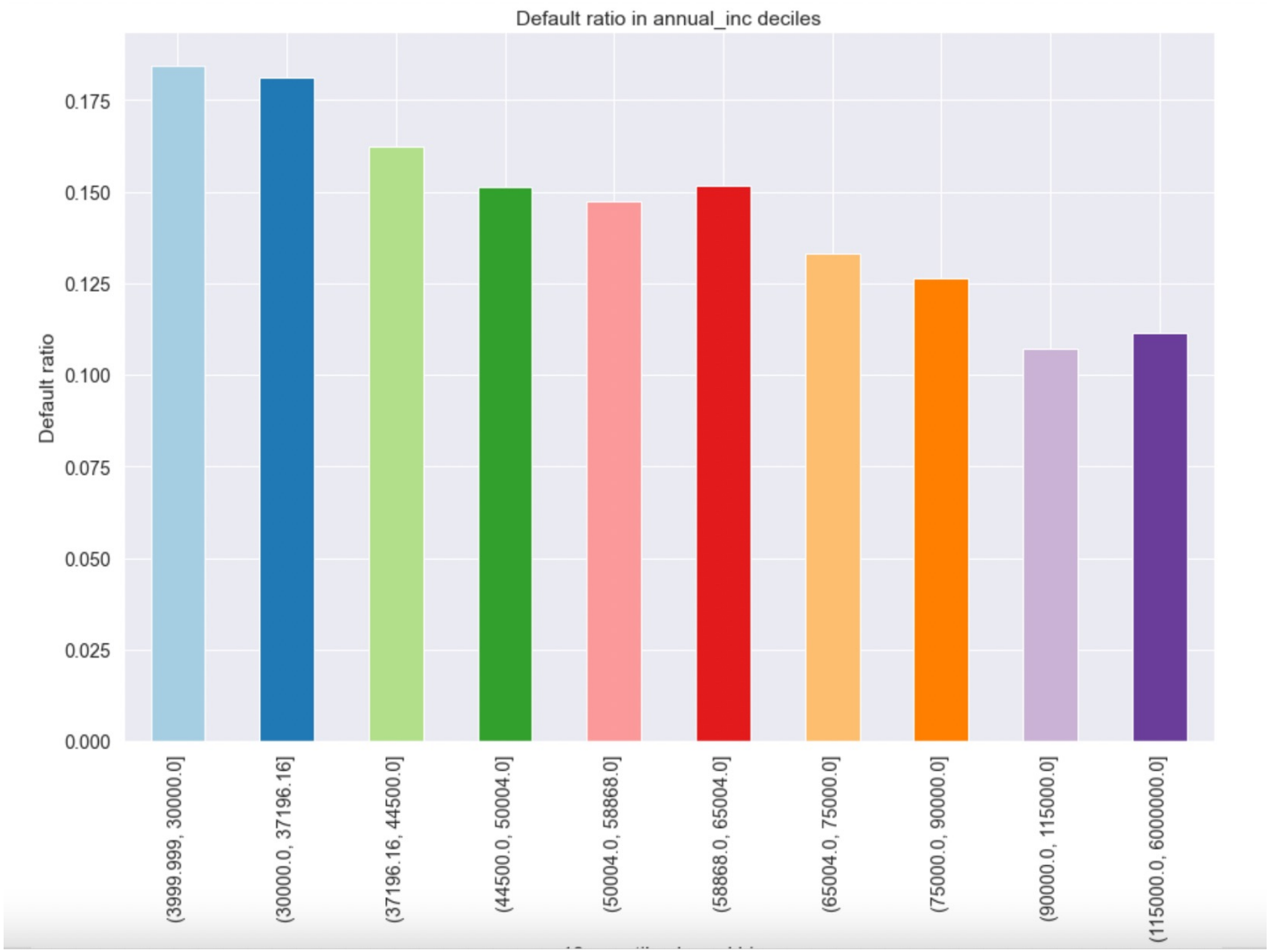
Identified potential predictors



Customer’s financials

Annual income

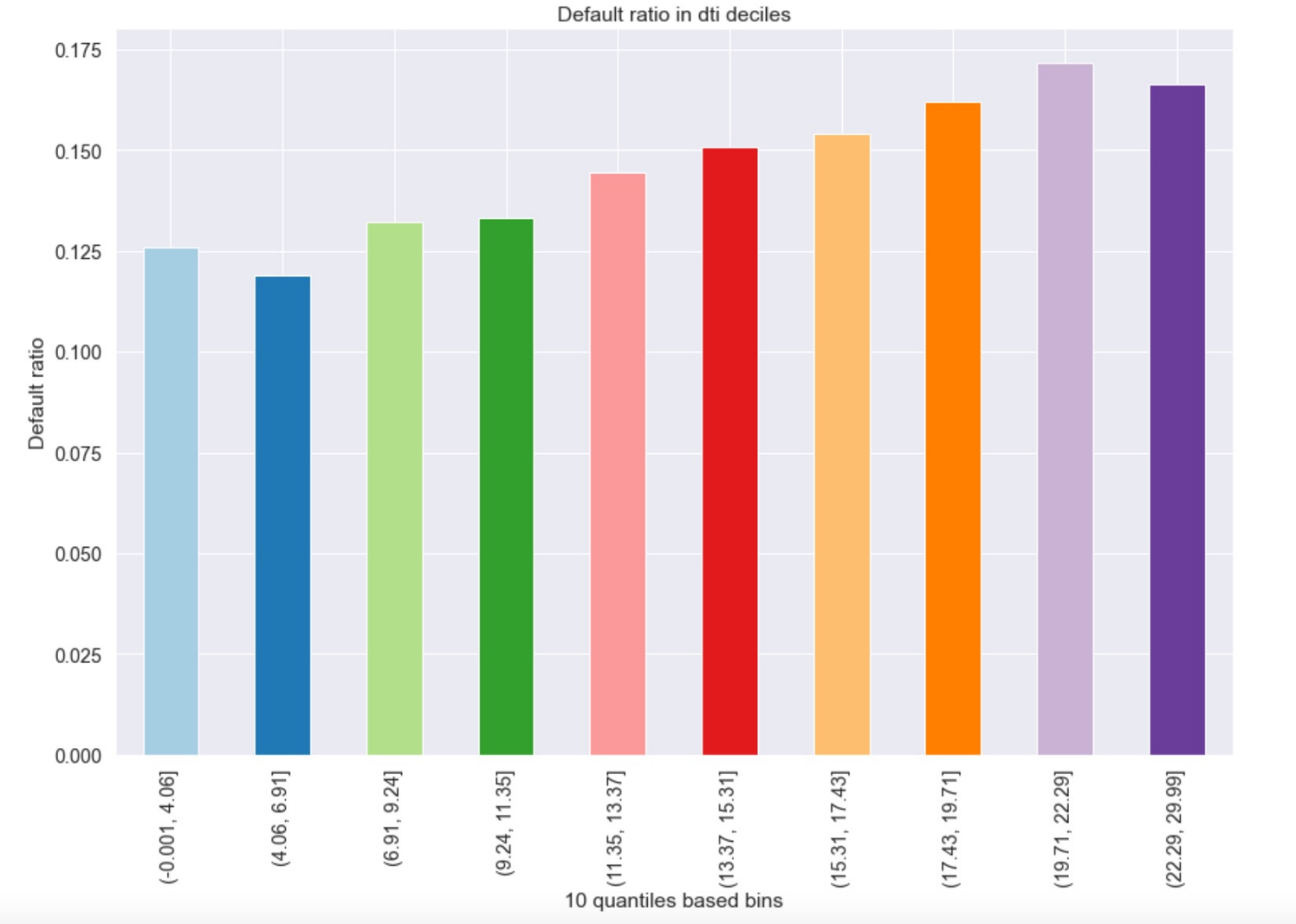
- The default rate decreases as the annual income increases.
- In the lowest earner decile, the default rate is over 17.5% while among the top earners is around 11%



Debt to Income ratio

DTI: The debt-to-income ratio shows what percentage of income is encumbered with debt

- The higher the DTI the higher the default ratio



Financials summary

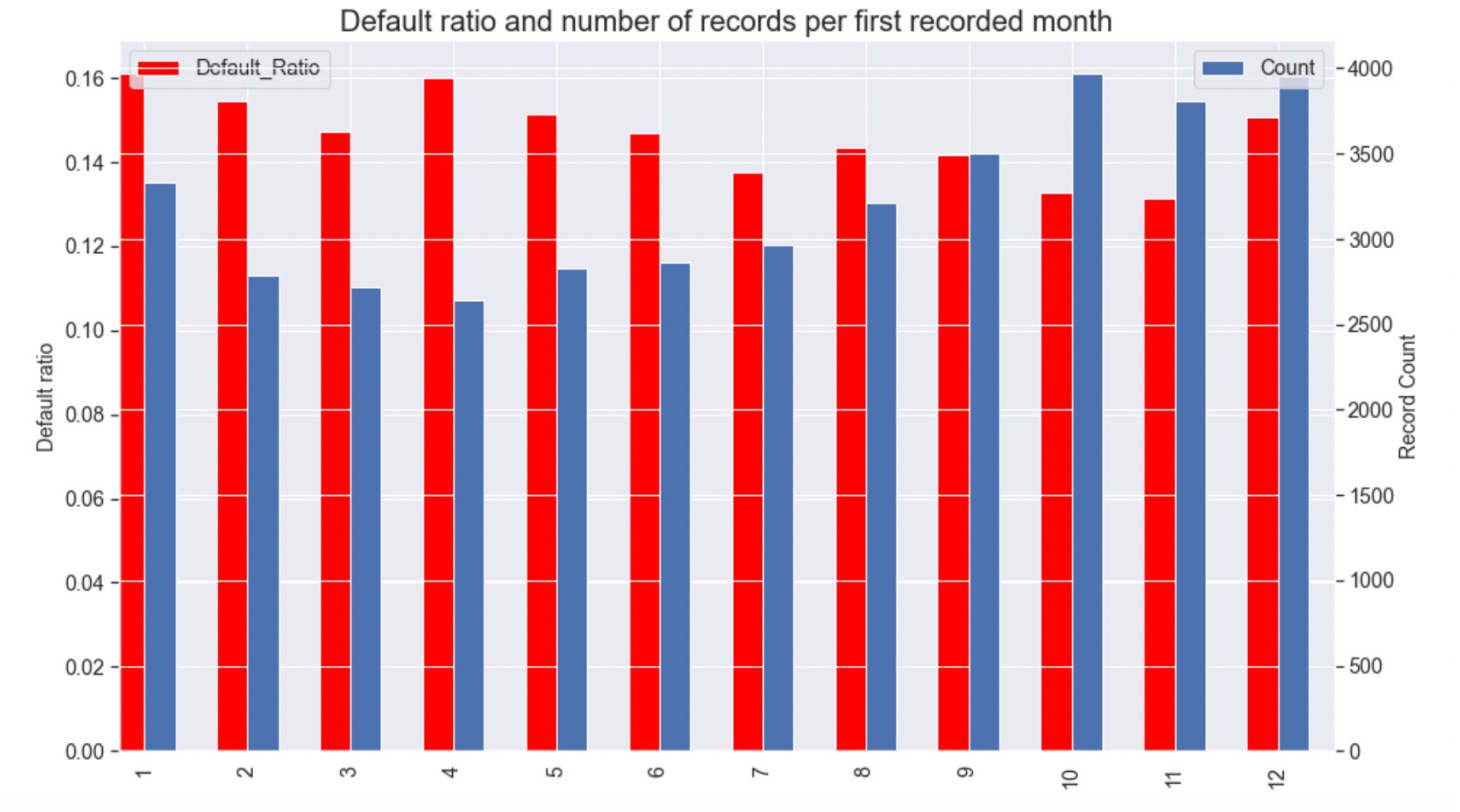
The correlation between the **annual income** and the **debt to income ratio** -0.12153. This is a relatively weak correlation meaning that the two factors can be taken into account separately.

Credit information

First Reported Credit Line Month

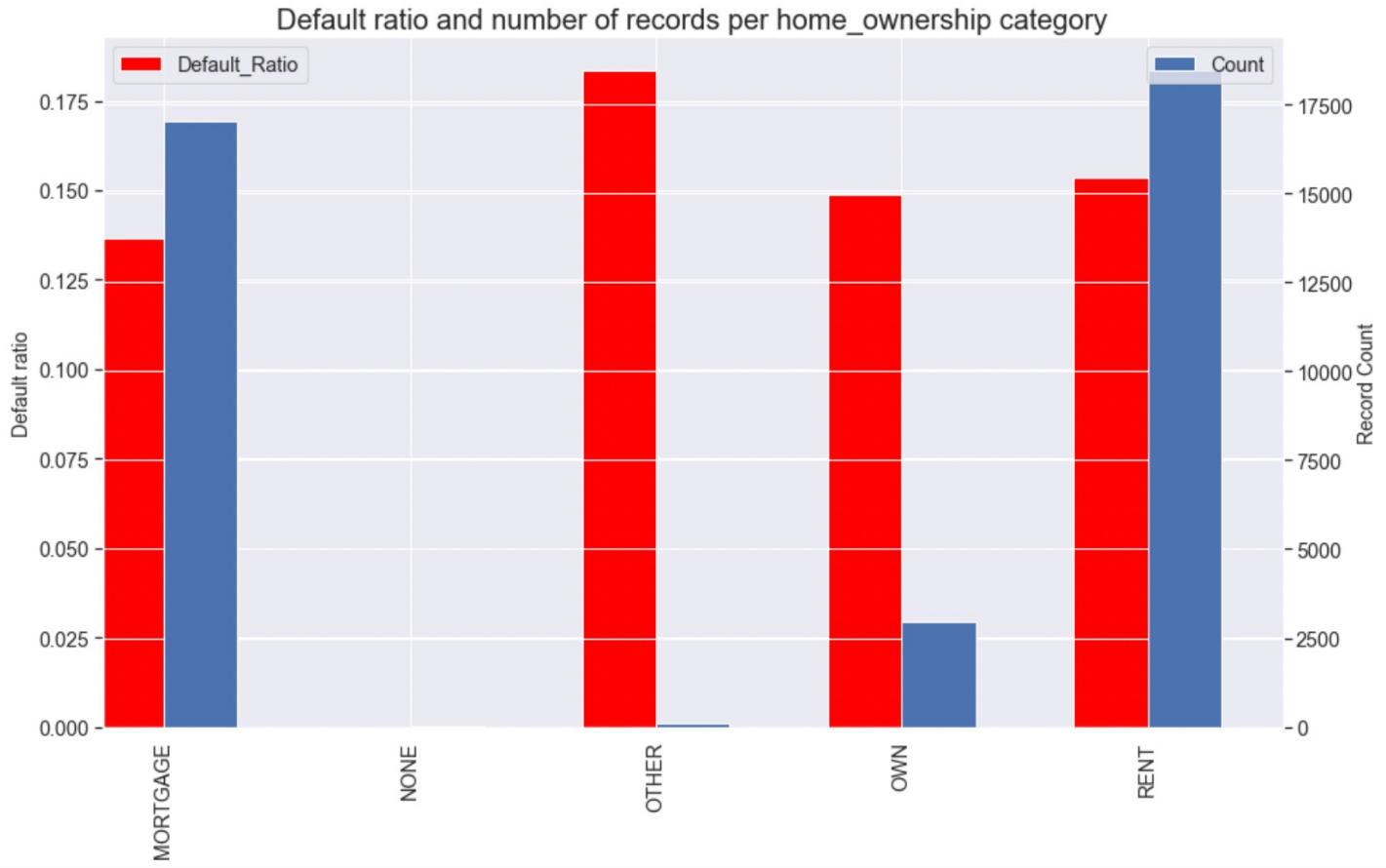
Not much of a predictor but an interesting feature. There is a clear trend in the number of cases in each month and in the default rates in each month. They however seem to be ‘shifted’.

It might indicate an operational problem where higher ‘load’ on the operation results in less precise decisions. Further analysis is needed



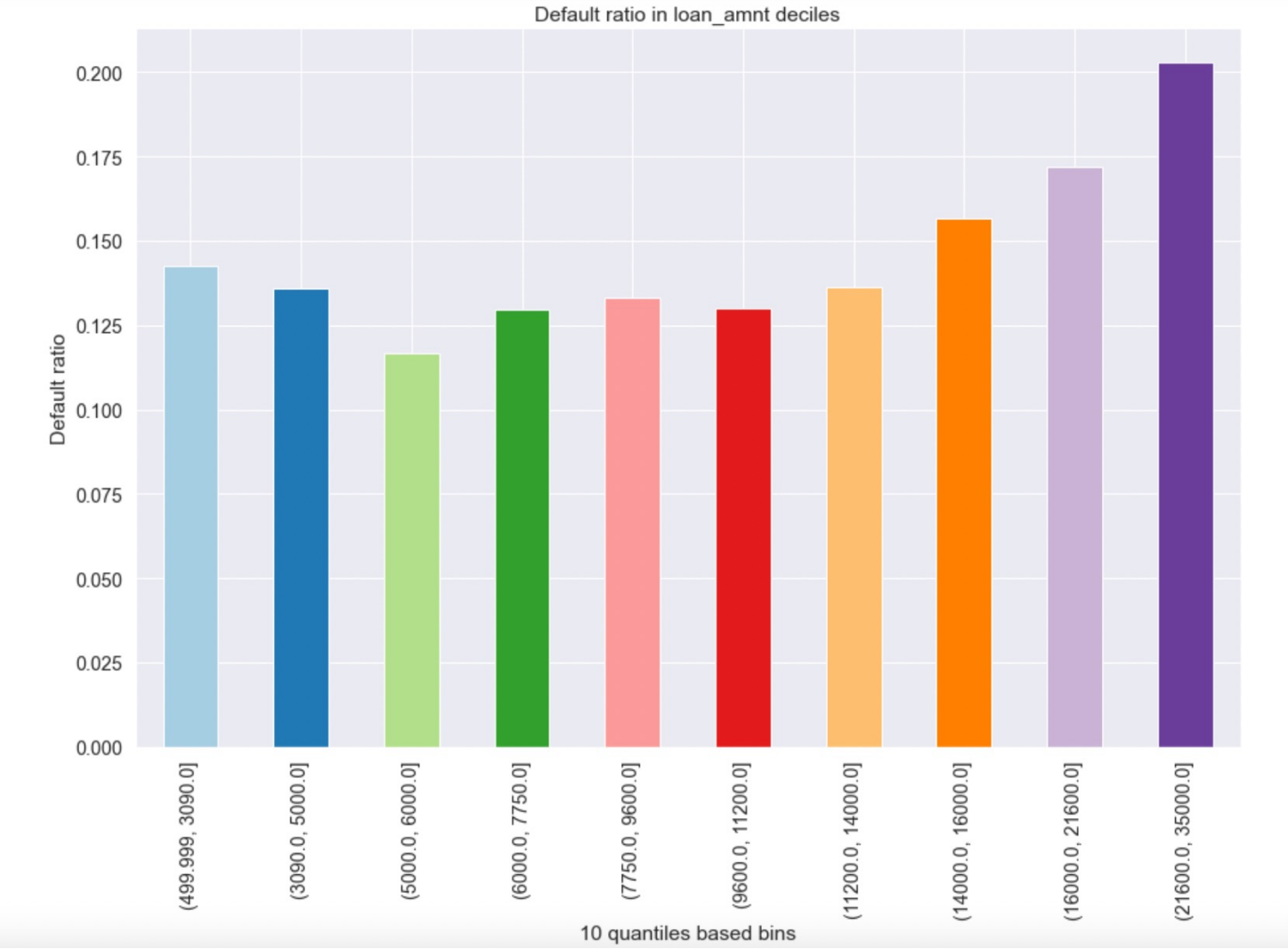
Home ownership

- In case of ‘mortgage’ the default rate is the lowest.
- It might indicate that the mortgage has indeed a deterrent effect



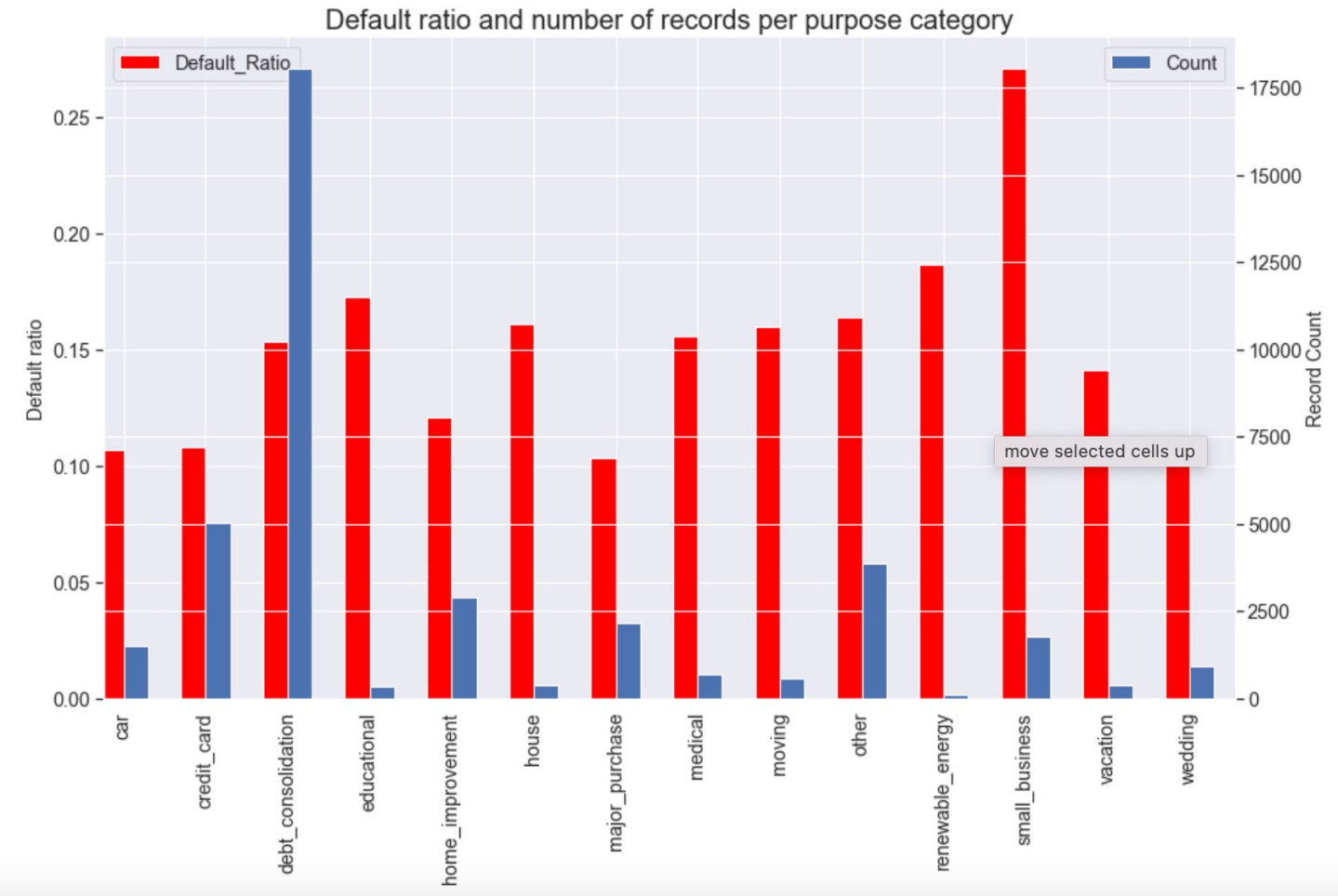
Loan amount requested

- There is a clear trend showing that the higher the loan amount the higher the default ratio
- Interestingly there is an increase in the defult ratio in the lowest loan amount decile



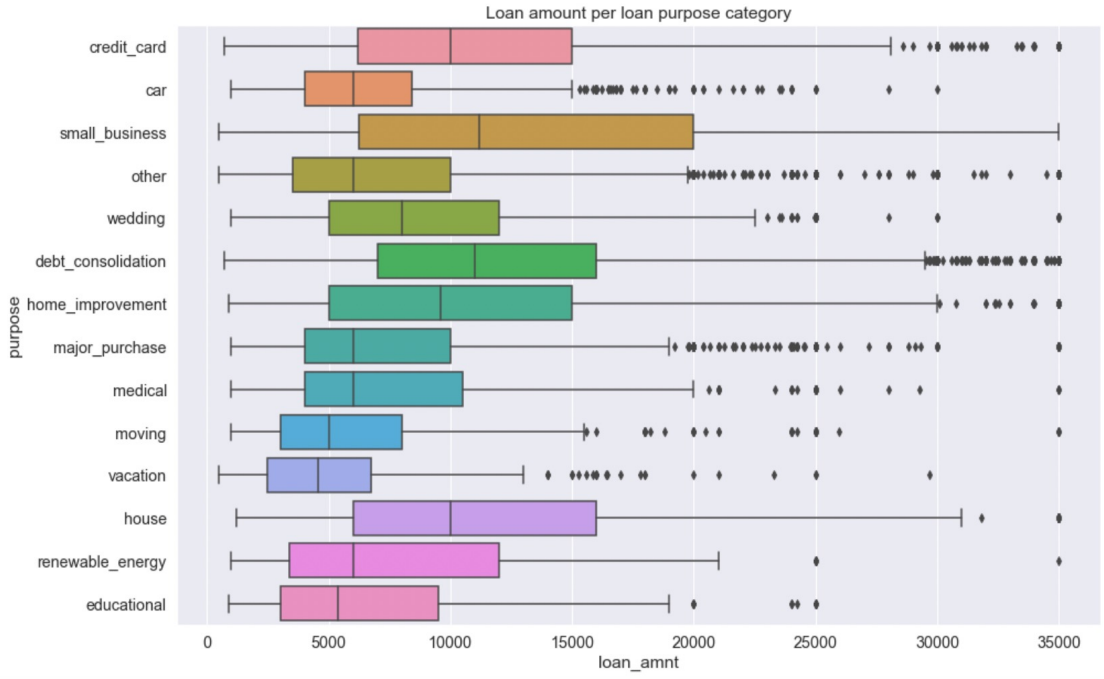
Loan purpose

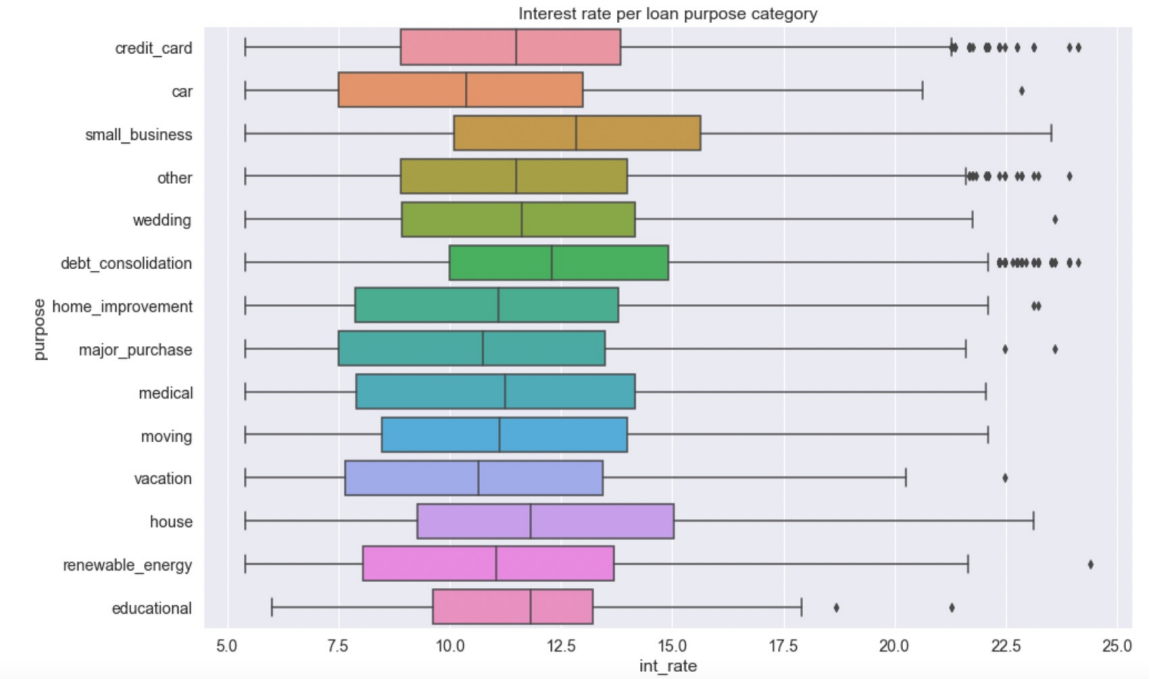
- Most of the loans are granted for consolidation
- The default rate in this category is top4
- Small business has the highest default rate. These deals might be small in number but the average loan size is the highest here (see next slide)



Loan purpose II.

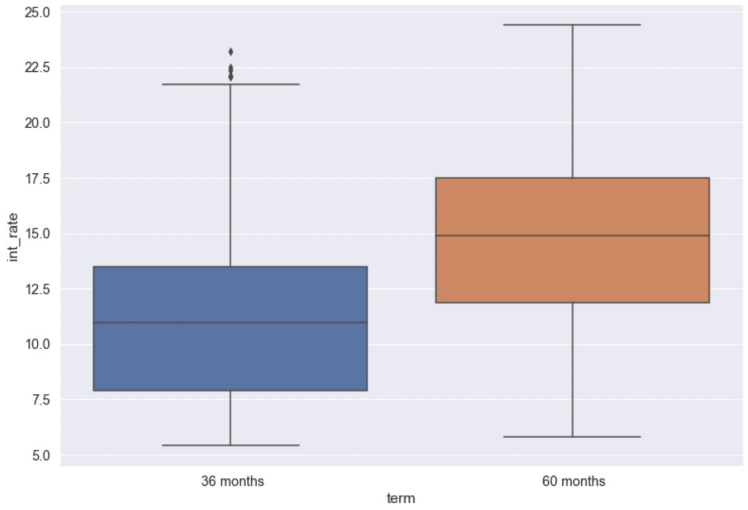
- The average interest rate is higher in the small business category that might cover the higher potential loss. But this requires further analysis.
- The debt consolidation has the 2nd highest interest rate but only the 4th highest default rate. It sounds like a strong business line
- Car loans are less likely to go default.

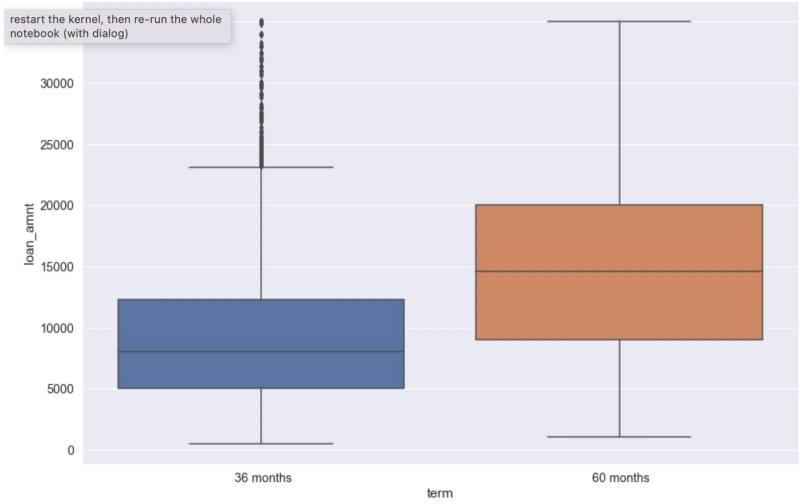


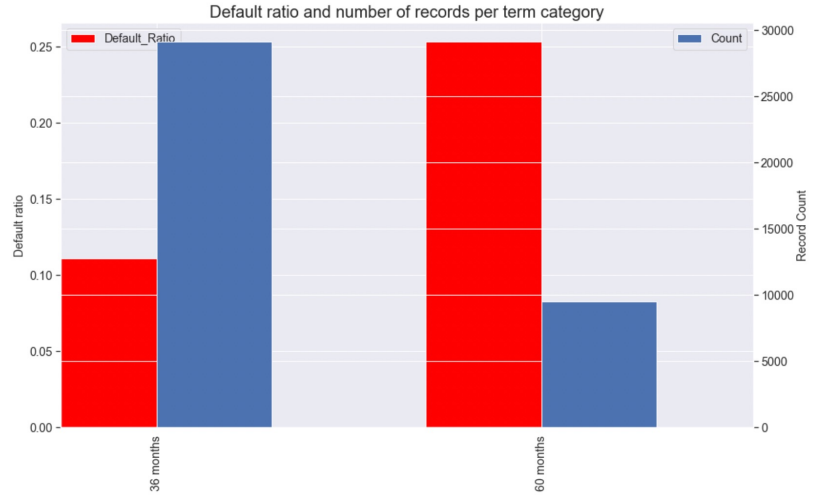


Term

- The longer-term deals have clearly higher default rates
- The higher risk might be covered however by the higher interest rates







Conclusion I.

- Customer financials: The **'Annual income'** of the customer is in negative correlation with the default ratio meaning the higher the income bucket, the lower the default ratio is. The **Debt To Income** ratio is in positive correlation with the default ratio meaning the higher the DTI the higher the default ratio is. These observations are matching the intuitions.
- Loan information: The **Loan Amount** is in positive correlation with the default ratio The **Purpose** of the loan can be a good indication. SME loans tend to have higher default ratio. The **Term** of the loan is a good indication as well. The longer (60 months) category has higher default ratio.
- Other factors: Loans with **Mortgage** are less likely to go default. The same applies for **Car** loans. It indicates that people who have something to lose (Car or Home) tend to be more careful with decision on loan payments.

Conclusion II.

- With the impressive expansion of the lending activities the portfolio quality seem to be declined.
- Higher loads in certain months seem to have a negative impact on quality
- The some of the tools – supporting the credit decision – seem to lose some predicting power.

Based on the above a complete review is recommended on risk appetite, pricing, sectors / purposes financed, internal operations and decision rules. The previously mentioned attributes might serve as a good foundation for setting up new/additional decision rules

Thank you for your attention!