Lending club case study

Swarnim Iyer

Supriya Sinha

Agenda

- Problem definition
  - Business understanding and objective

- Solution approach
  - Solution checklist

- Result Synthesis
  - Major insights, underlying assumptions and recommendations

- Appendix

Problem Definition

Basic understanding of the risk analytics problem in banking and financial services by breaking it down into the WHO, HOW, WHAT and WHYs.

The biggest loss incurred in the lending business is through **defaults from risky customers**

WHO is the decision maker?
Lending executive in a customer finance company responsible to provide loans to urban customers ensuring risk mitigation is done for loan provisions to potential defaults and potential applicants.

HOW are the decisions taken?
Once a loan application is generated there are 2 types of decisions taken:

1. **Acceptance of loan (revenue improvement)** for the potential customers who are likely to pay the loan back on various rates and tenures as per the profile of the customer.

2. **Rejection of loan (loss reduction)** for the potential defaulter who may not pay the loan back.

WHAT are the pain points?
Credit loss i.e. non-repayment of the loan amount by the risky customer is the biggest cost incurred by the finance company.

WHY do these pain points exist?
Lack of data-based drivers to predictively determine the risky customers.

Objective

Identifying the data-based driving factors for loan defaults which can proactively assist in portfolio and risk assessment for a risky customer.

Solution Approach

The approach to the driver analytics problem highlighting the methodology to solve the problem with required assumptions

Our solution approach is broadly divided into **three sections**

1 **PROBLEM DEFINITION AND BUSINESS UNDERSTANDING**
Defining the objective of the ask and performing external research to get up to speed with the business

2 **DATA EXPLORATION AND UNDERSTANDING**
Post deciding the objective, data exploration was taken up which comprised of following steps:

DATA PREPROCESSING AND CLEANING
Cleaning the data and changing it into the required format for analysis. This included the format changes required within the data along with removal of the outliers based on our problem understanding.

INITIAL DATA EXPLORATION
Performing univariate, segmented univariate and bivariate analysis on the cleaned data to achieve a directionality in performing deep dive analysis within the data.

DEEP DIVE SYNTHESIS
Based on the learnings from the initial EDA specific findings were targeted and explored in-depth by creating business driven, type driven and data driven derived metrics.

3 **RESULT SYNTHESIS**
Completion of a thorough EDA is followed by the insight and recommendation provision for the problem at hand

Result Synthesis

The summarization of the data exploration exercise presented with the underlying assumptions and actionable recommendations.

# There are **5 major insights** from the data exploration exercise

| Insights	Recommendation
The last payment value's ratio the installment amount is comparatively lower for the defaulting customers.	Identify the customers whose last payment to installment amount ratio is very low (<5), they are highly likely to default.
For the defaulting customers, the interest rate distribution is on higher end compared to the existing customer.	Interest rate can be utilized as a data marker to identify the at-risk customers when loan provisioning is done.
Customers who are likely going to default prefer loans with higher term compared to non defaulting customers.	Identify the customers who are preferring higher loan terms and flag them to have higher propensity to be at risk
Customers getting loans with lower grades (i.e. B or lower) are likely going to default compared to higher grade ones.	Flag the customers who are getting lower grade loans and keep a watch over these customer on a regular basis for defaulting.
The customers who are highly unlikely to default have very rounded loan amounts (i.e. in the multiples of 1000s) and comparatively of bigger values as well.	Keep a note of the customer who are applying such loans of non rounded denominations and of smaller values

Appendix

Enhancing the insights

The last payment value's ratio the installment amount is comparatively lower for the defaulting customers
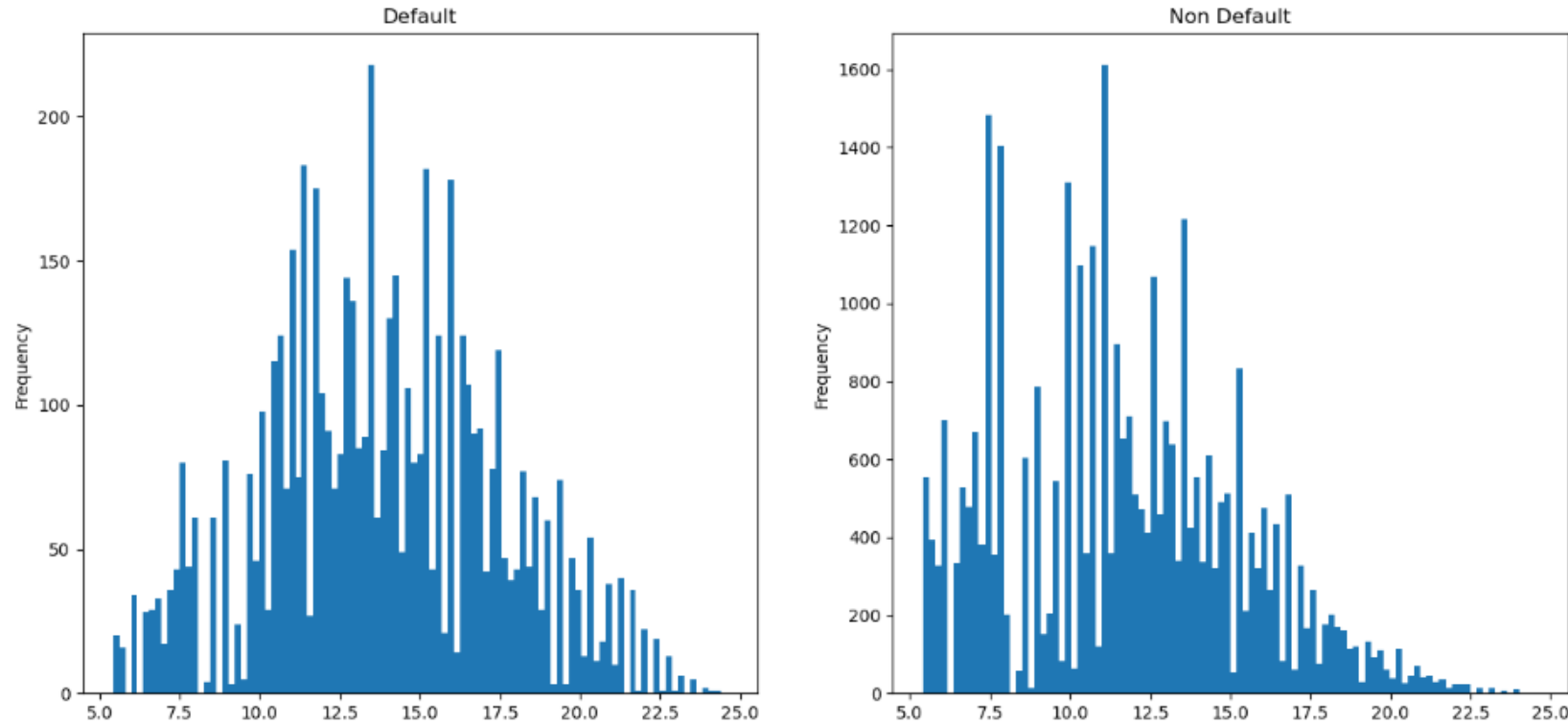

Default
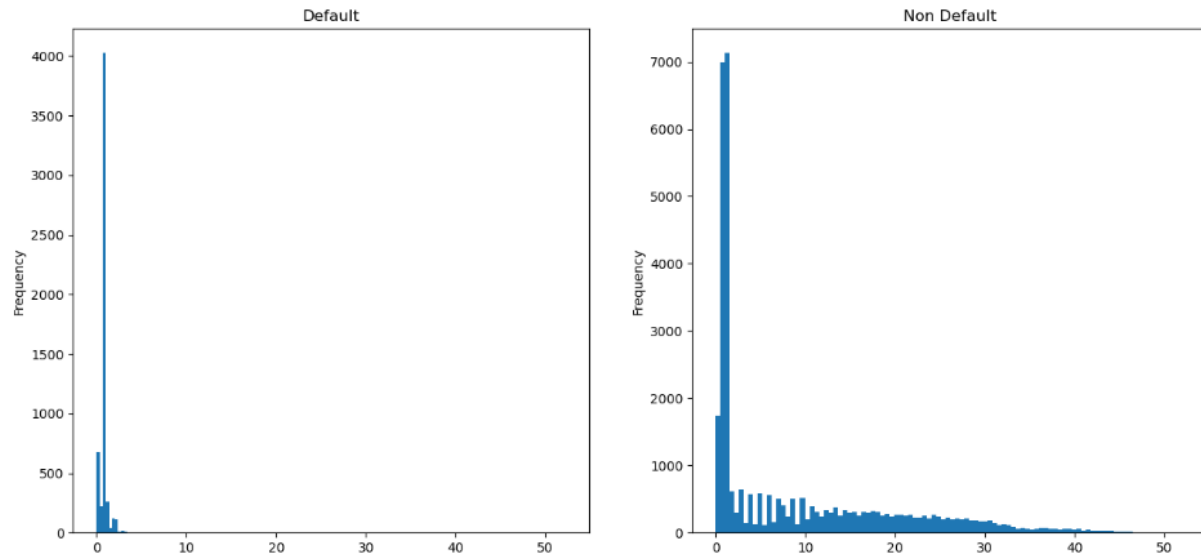
Non Default

<- Frequency distribution of the default vs non default for the PVI ratio

Box plot representation of the default vs non default for the PVI ratio ->
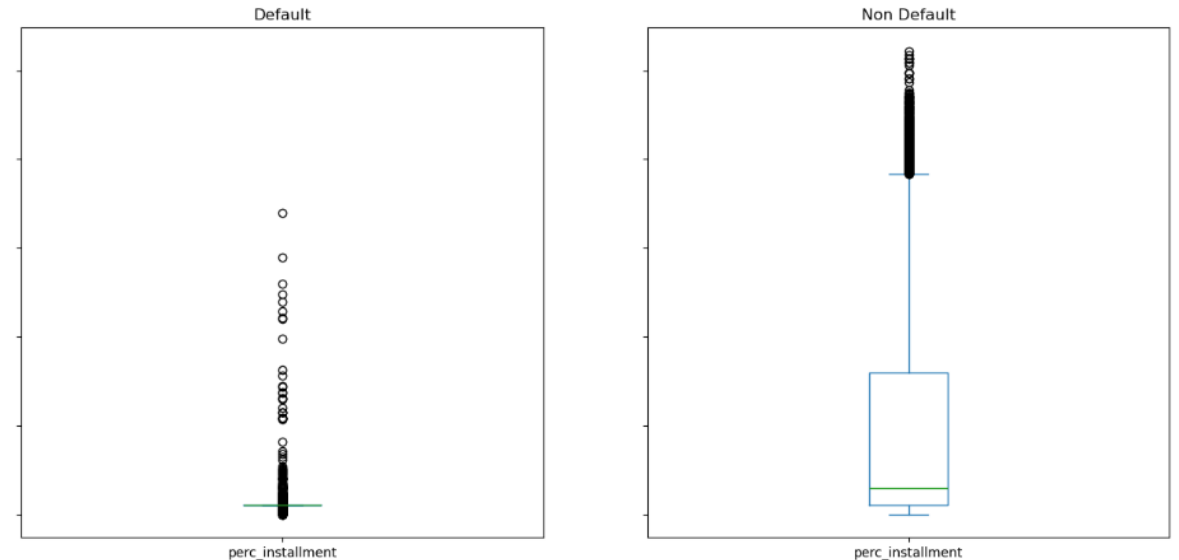
Default
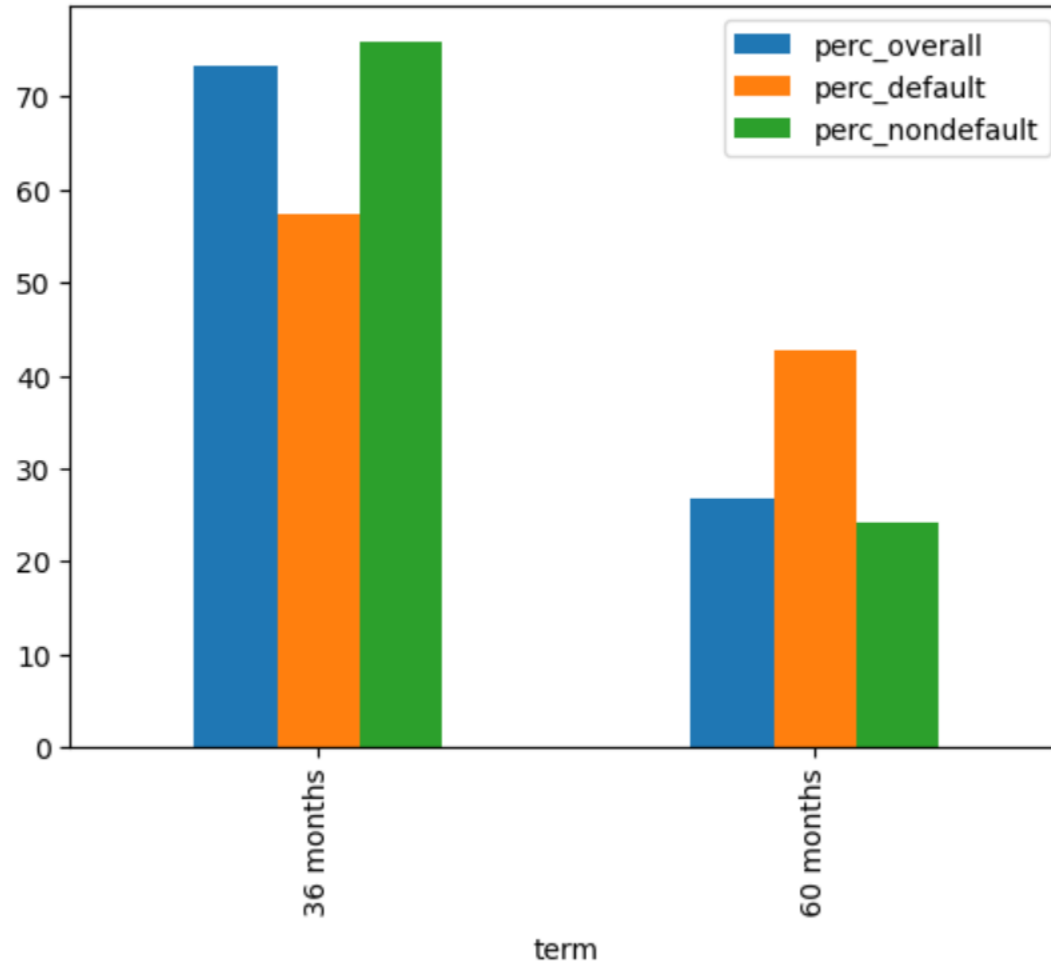
perc_installment

Non Default

perc_installment

For the defaulting customers, the interest rate distribution is on higher end compared to the existing customer


Frequency distribution of the default vs non default for the interest rate

Customers who are likely going to default prefer loans with higher term compared to non defaulting customers.

Term level split of the customer count with cohort segregations of overall, default and non default

Customers getting loans with lower grades are likely going to default compared to higher grade ones.
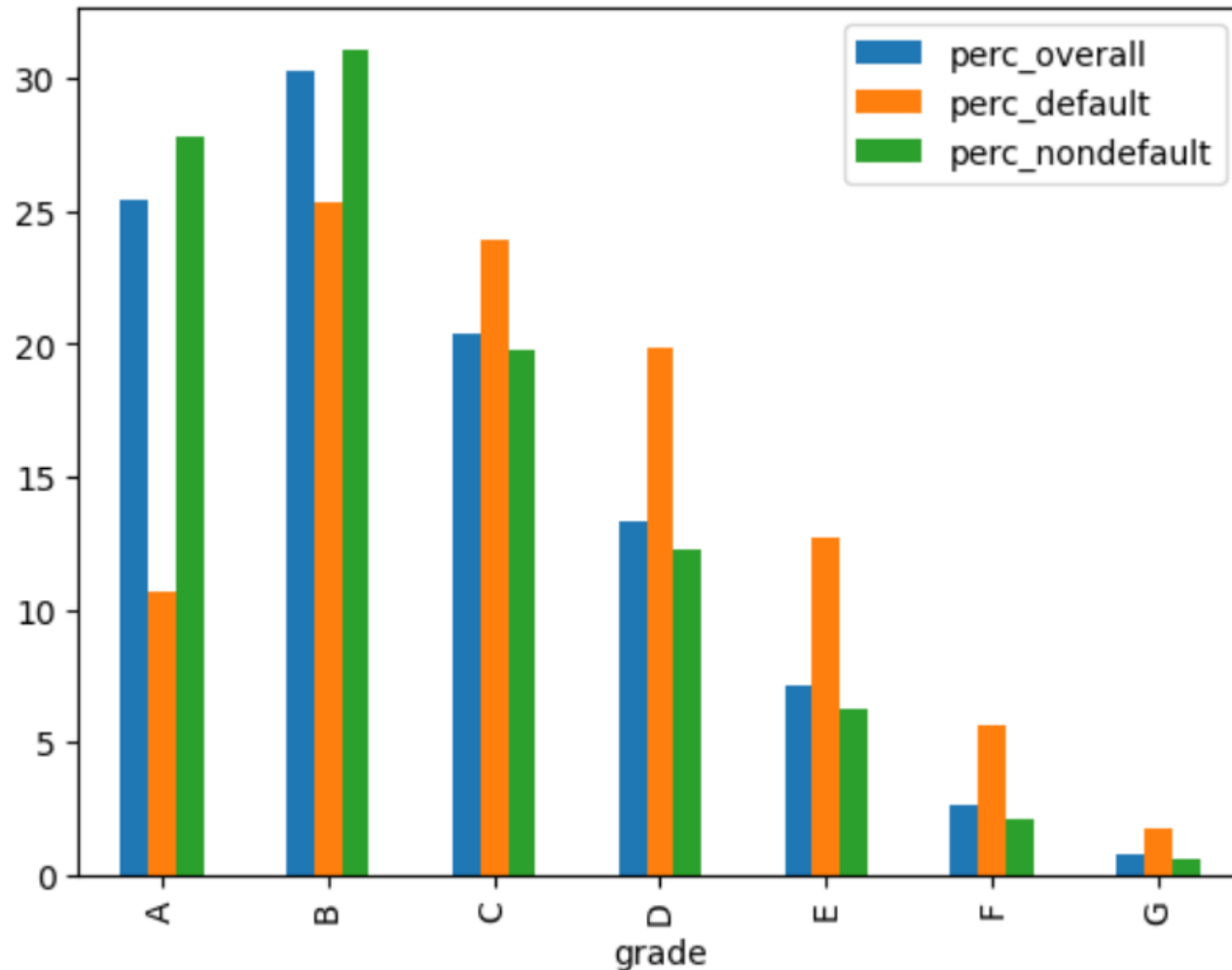
Grade level split of the customer count with cohort segregations of overall, default and non default

The customers who are highly unlikely to default have very rounded loan amounts and comparatively of bigger values as well.

flag	overall counts	default counts	non default counts
Regular	16.412941	4.497788	14.551893
Rounded-off	736.171429	102.685714	633.485714

Count of customers split across rounded off loan amounts for the three-customer cohort of overall, default and non default