

upGrad

#LifeKoKaroLift



Credit EDA Case Study



Module : Credit EDA Case Study

Instructor: Dr. Darshan
Ingle



Introduction

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Business Understanding

banks

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

1 yr

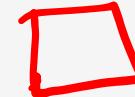
1 lakh - 4th day of every month

5th month → 7th day
=====

The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

- **The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
- **All other cases:** All other cases when the payment is paid on time.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

1. Approved: The Company has approved loan Application
2. Cancelled: The client cancelled the application sometime during approval. Either the client ~~changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.~~
3. Refused: The company had ~~rejected the loan (because the client does not meet their requirements etc.).~~  
4. Unused offer: Loan has been ~~cancelled by the client but on different stages of the process.~~

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Business Objectives

This case study aims to identify patterns which indicate if a client has difficulty paying their installments 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. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

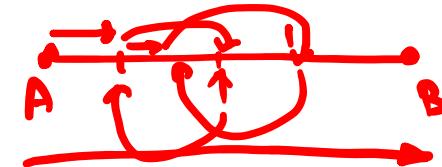
In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

To develop your understanding of the domain, you are advised to independently research a little about risk analytics - understanding the types of variables and their significance should be enough).

4.

- Keep in mind “There is no correct or incorrect solution”.
- Every approach is correct if you are able to answer all the questions asked.
- The main objective of this case study is to learn and implement EDA techniques. So don’t focus so much on columns and their descriptions.
- Remember you need to use plots and then understand the pattern. Then report your analysis in your notebook or PPT. No marks will be awarded if you have just plotted so many variables and have not explained the pattern.

I will tell u.
- It’s not possible to cover all the columns, so try to cover some of them. Based on the plots you get, try to identify the important variables.



Data Understanding

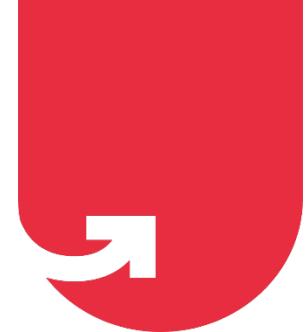
This dataset has 3 files as explained below:

- * 1. 'application_data.csv' contains all the information of the client at the time of application.
The data is about whether a client has payment difficulties.

- 2. 'previous_application.csv' contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

- 3. 'columns_description.csv' is data dictionary which describes the meaning of the variables.

Lets get started



Thank You!

Connect me: <https://www.linkedin.com/in/dr-darshan-ingle-corporate-trainer/>

