



Credit EDA Case Study

SIVAGIRI.N

Introduction

This assignment aims to give you an idea of applying EDA in a real business scenario. In this assignment, 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 minimize the risk of losing money while lending to customers.

Business Understanding -1

- 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.

Business Understanding -2

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, we will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

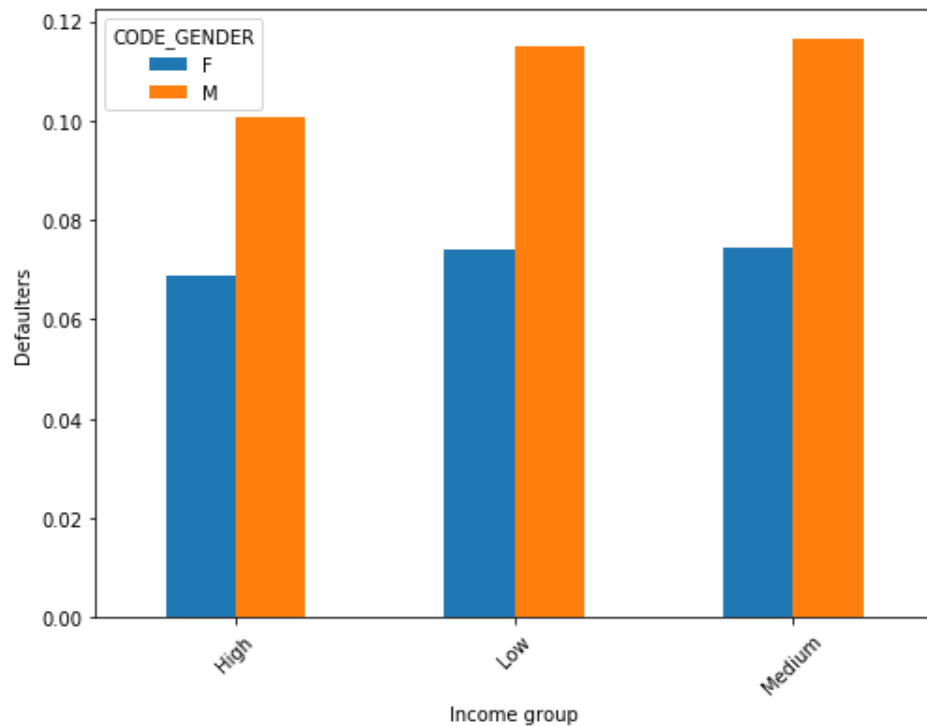
Business Objectives

- 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 utilize 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).

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 a data dictionary which describes the meaning of the variables.

Current applications



Income groups & age groups

Observations:-

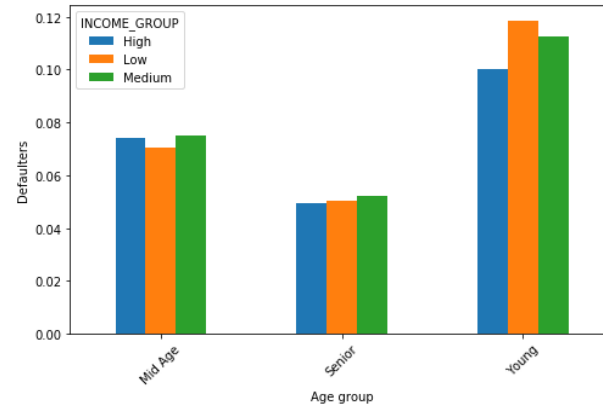
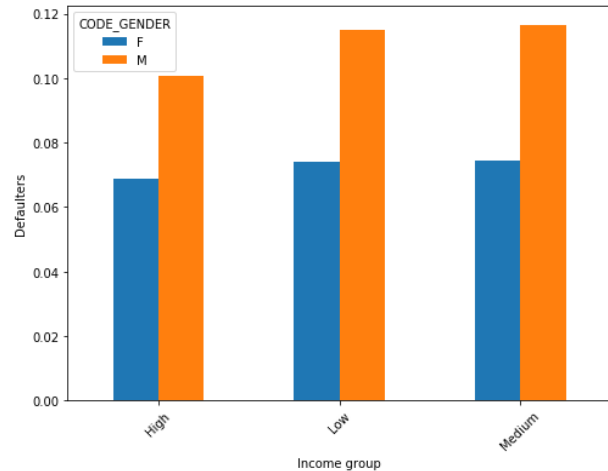
1. High income groups are less defaulter than comparatively lower income groups.
2. Mid age and senior people with all income groups are less defaulted.

Recommendation:-

1. Safer to grant loan for mid age and senior citizen clients with higher income.
2. Risky to grant loans for young people with low income groups.

Current applications

Income groups & gender vs Income groups & age groups



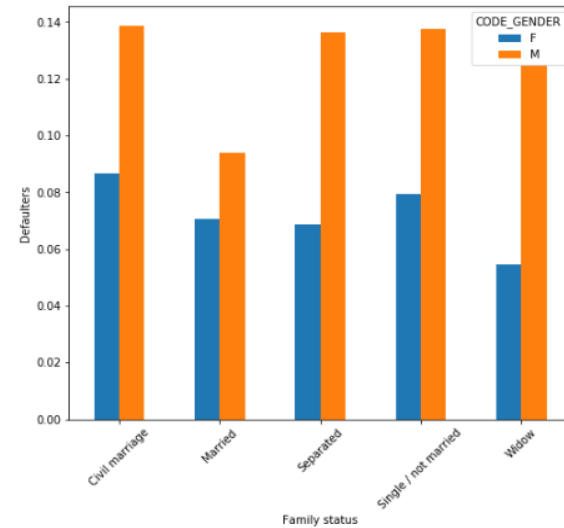
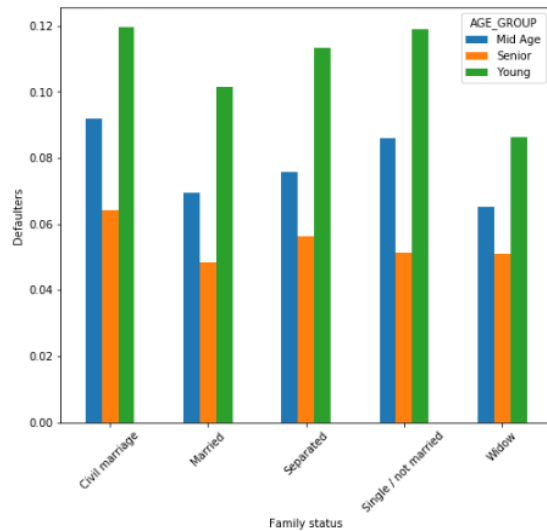
Observations:-

1. Across all income groups clients with medium amount credited is highly defaulted followed by low and high credit amount.
2. Young clients with medium and low amount credited are most likely defaulted.

Recommendations:-

1. Recommended to grant slightly higher amount of loan to all income groups.
2. It is very risky to grant medium and low amount of loan to young clients.

Family status & age group vs Family status & gender



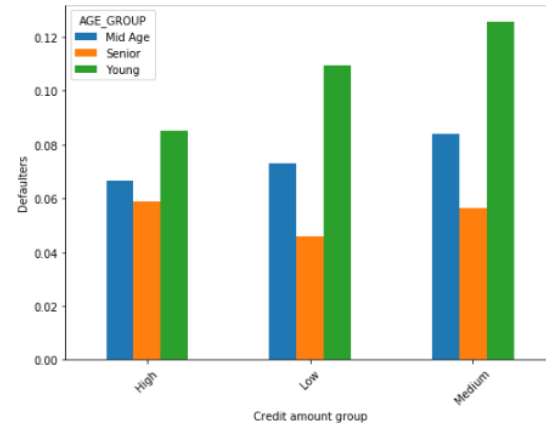
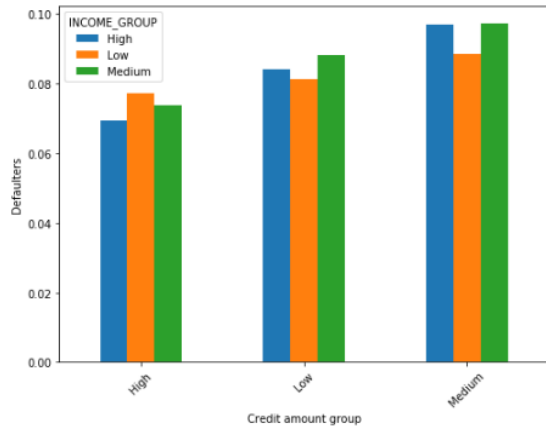
Observations:-

1. Senior people irrespective of family status are less likely to be defaulted.
2. Young people are more likely to be defaulted in all family status.
3. Males are more like to be defaulted than females.

Recommendations:-

1. Better to grant loan for senior citizen of all family status.
2. It is risky to grant loan for single, separated and civil marriage young men.

Credit amount group & income group vs Credit amount group & age group



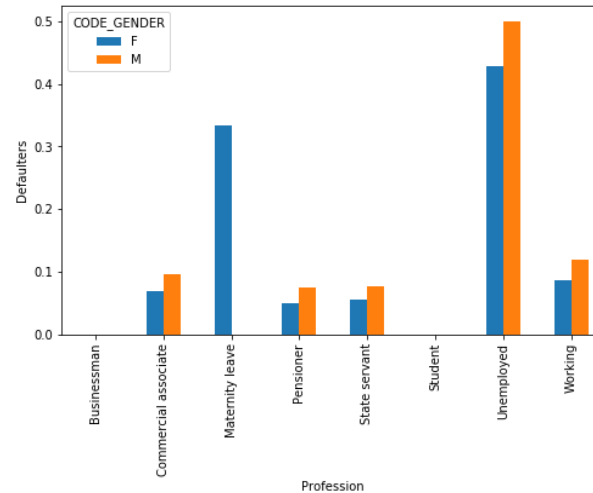
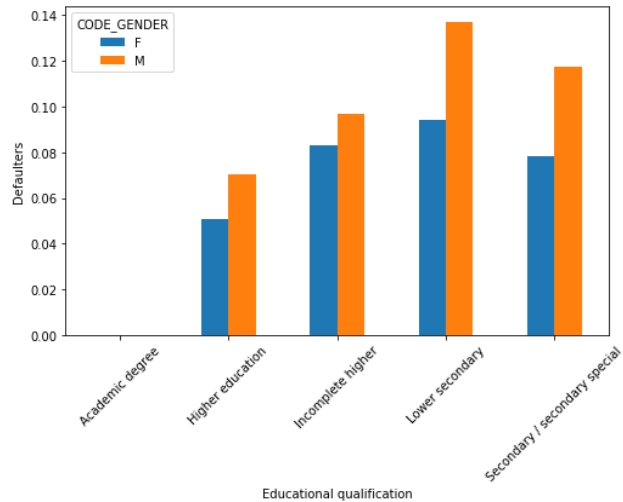
Observations:-

1. Across all income groups clients with medium amount credited is highly defaulted followed by low and high credit amount.
2. Young clients with medium and low amount credited are most likely defaulted.

Recommendations:-

1. Recommended to grant slightly higher amount of loan to all income groups.
2. It is very risky to grant medium and low amount of loan to young clients.

Educational qualification & gender vs Profession & gender



Observations:-

1. Higher educated people are less defaulted and lower secondary educated people are more.
2. Unemployed clients along with clients with maternity leave are heavily defaulted.

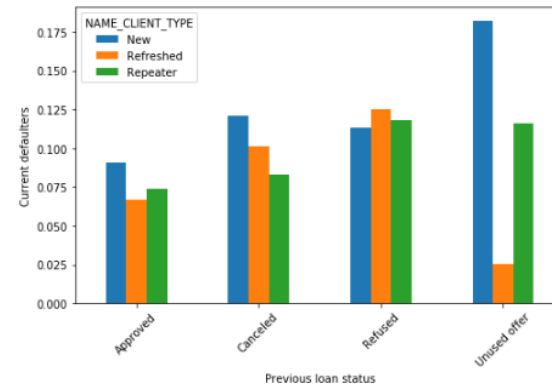
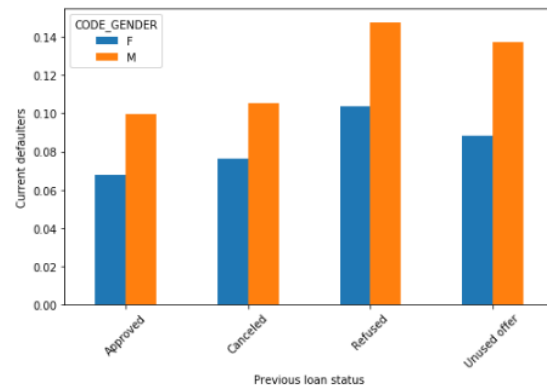
Recommendations:-

1. Safe to grant loans to higher educated clients across all profession except unemployed and women with maternity.

Loan application status relations

CURRENT DATA AND PREVIOUS DATA

Previous loan status & gender vs Previous loan status & client type



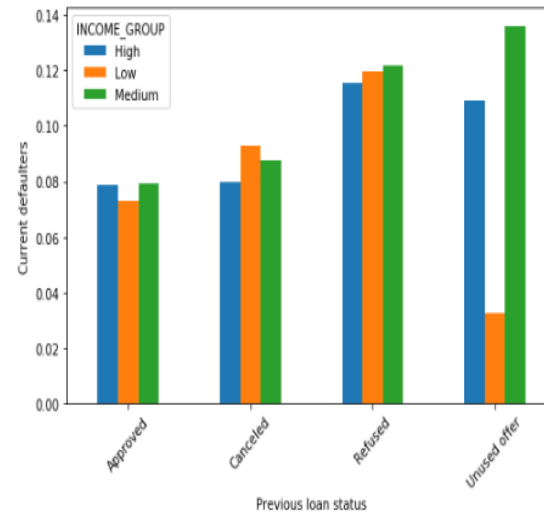
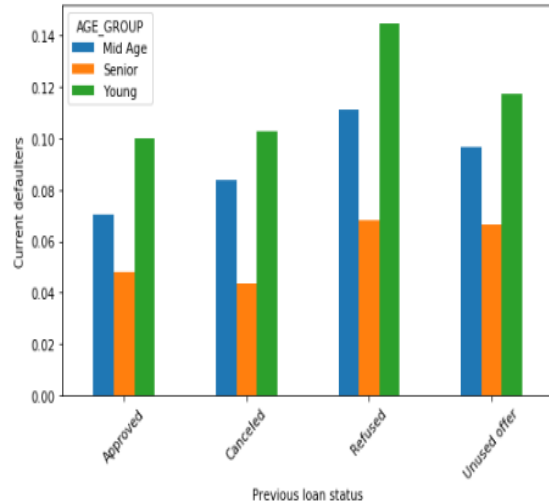
Observations:-

1. Previously refused and unused offer applications were more defaulted in male.
2. New clients with previously unused offer are more defaulted.

Recommendations:-

1. It is recommended to provide loans to previously approved females.
2. There is a risk to grant loans for clients, whose applications were refused or unused previously.

Age group & previous loan status vs Income group & previous loan status



Observations:-

- 1.Young people, who were previously refused are mostly defaulted.
- 2.The senior citizens are less defaulted irrespective of their previous loan status.
- 3.In all income groups previously refused applicants are more defaulted.

Recommendations:-

- 1.Safer to grant loans for senior citizen.
- 2.Lesser risk to grant loans for approved applicants to all income groups.

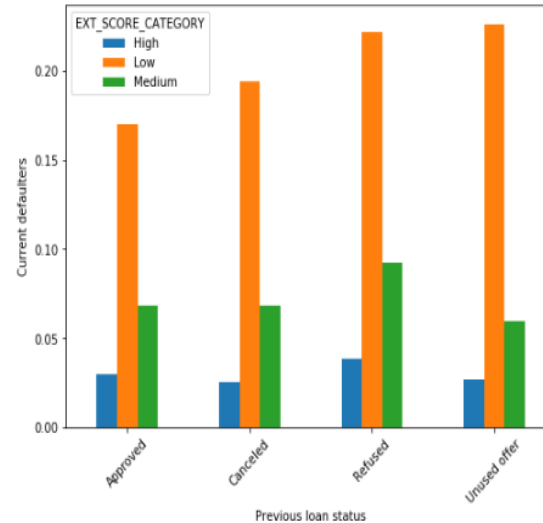
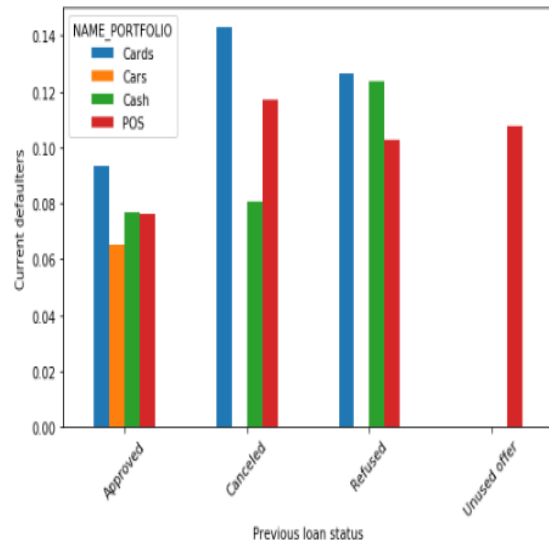
Portfolio & previous loan status vs External source score & previous loan status

Observations:-

- 1.The previous applications for portfolio Cards and POS are mostly defaulted .
- 2.Previously refused applications for Cash are also defaulted in higher rate.
- 3.Low external source scorer are highly defaulted irrespective of their previous loan status.

Recommendations:-

- 1.It is safer to grant loans for any portfolio for previously approved applicants.
- 2.It is high risk to grant loans for applicants, who have poor external source score specially whose loan were previously refused, unused or cancel.



Final note

Highly recommended groups:-

- 1.Approved clients in their previous applications.
- 2.Highly educated clients with higher income.
- 3.Clients with higher external source score.
- 4.Senior citizens in all categories.
- 5.Married clients compared to other family status.
- 6.Females are comparatively favourable than male.

High risk groups:-

- 1.Previously refused, cancelled or unused offer clients.
- 2.Low income groups with previously refused status.
- 3.Unemployed clients.
- 4.Poor external source scorer.
- 5.Young clients are comparatively riskier than mid age clients and senior citizens.
- 6.Lower secondary and secondary educated clients.

Thank you !

- Sivagiri.N

Your best quote that reflects your approach... “It’s one small step for man, one giant leap for mankind.”

- NEIL ARMSTRONG