Credit EDA Case Study

By

Pavan Manohar Deshpande

Business Understanding:

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. 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:

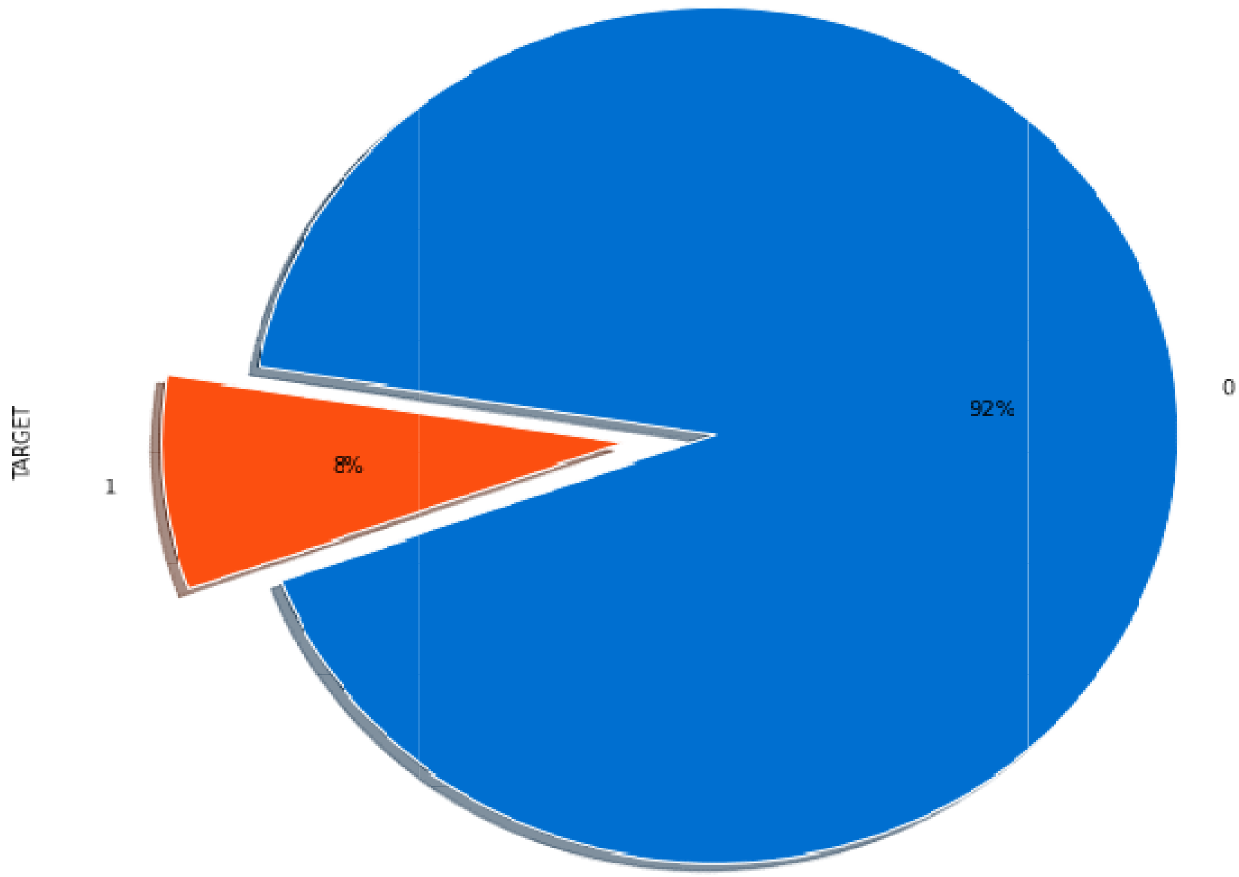
If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

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 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. 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.

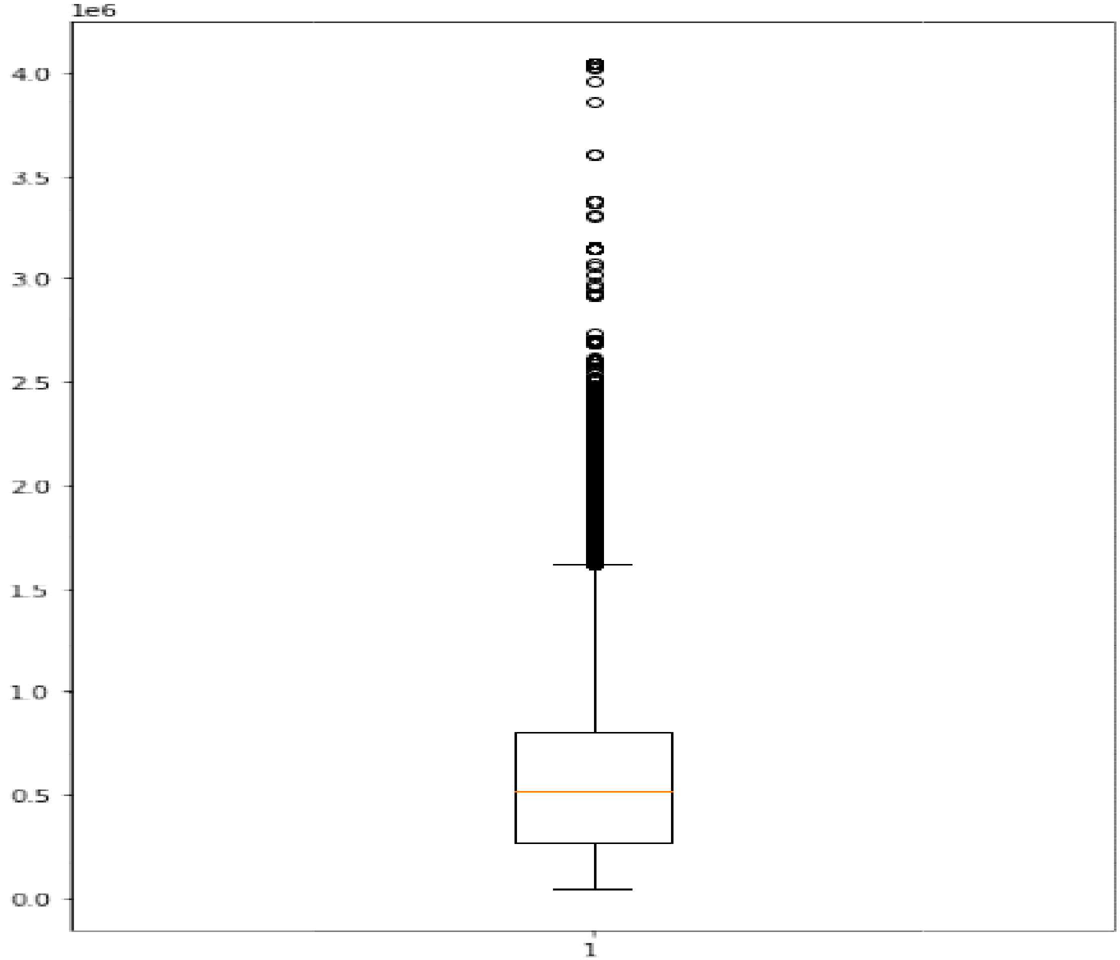
PROPRTION OF APPLICANTS FACING DIFFICULTY IN PAYMENT



We have application data of 307511 customers

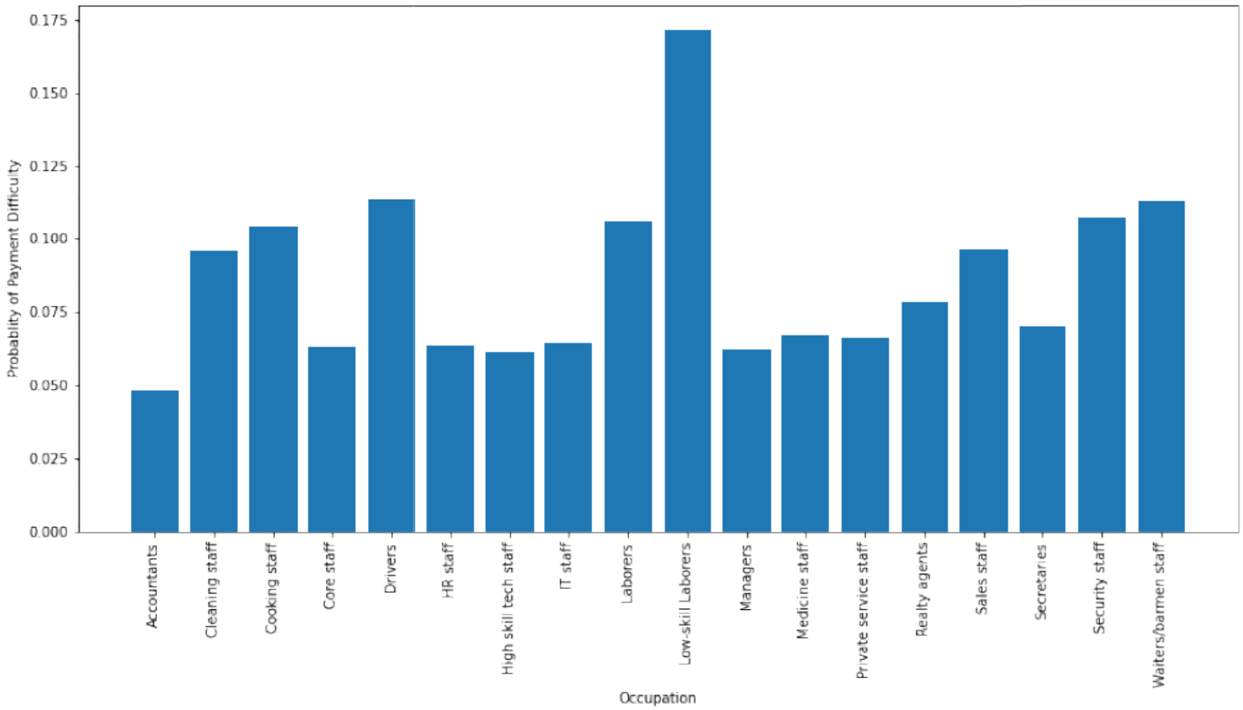
From the 307511 customers, 8% of the customers are facing difficulty in payment. In this case study we will try to segregate population who has difficulty more than 8% percent and less than 8 %. We will try to understand diving factors for loan default. Since we will be analyzing categorical data we will predominantly use Bar Charts and Heat Maps to draw inferences.

TICKET SIZE OF LOAN



The median ticket size of loan is 500000

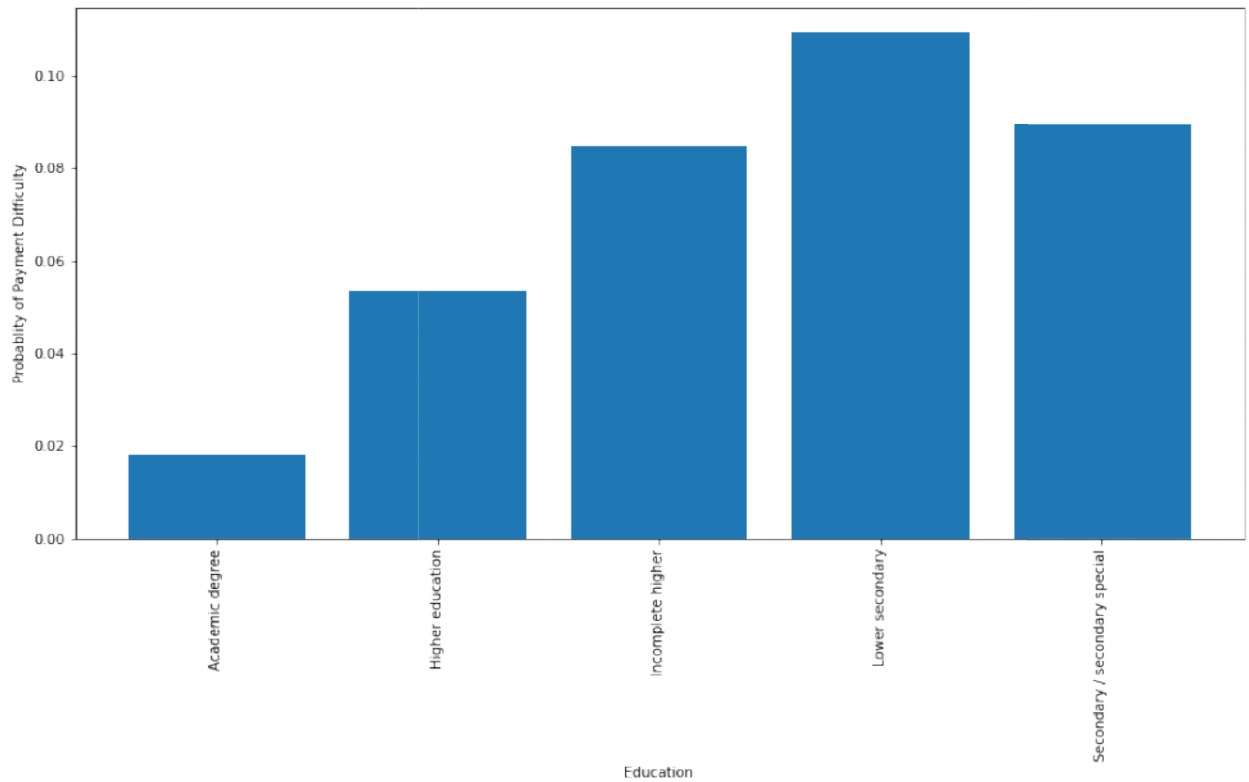
Ananlysis of relationship between occupation type and likelyhood of payment



Accountants, Core staff, HR staff, High skill tech staff, IT staff, Managers, Medicine staff, Private service staff, Realty agents, Secretaries are more likely to repay

Cleaning staff, Cooking staff, Drivers, Laborers, Low-skill Laborers, Sales staff, Security staff, Waiters/barmen staff are less likely to repay

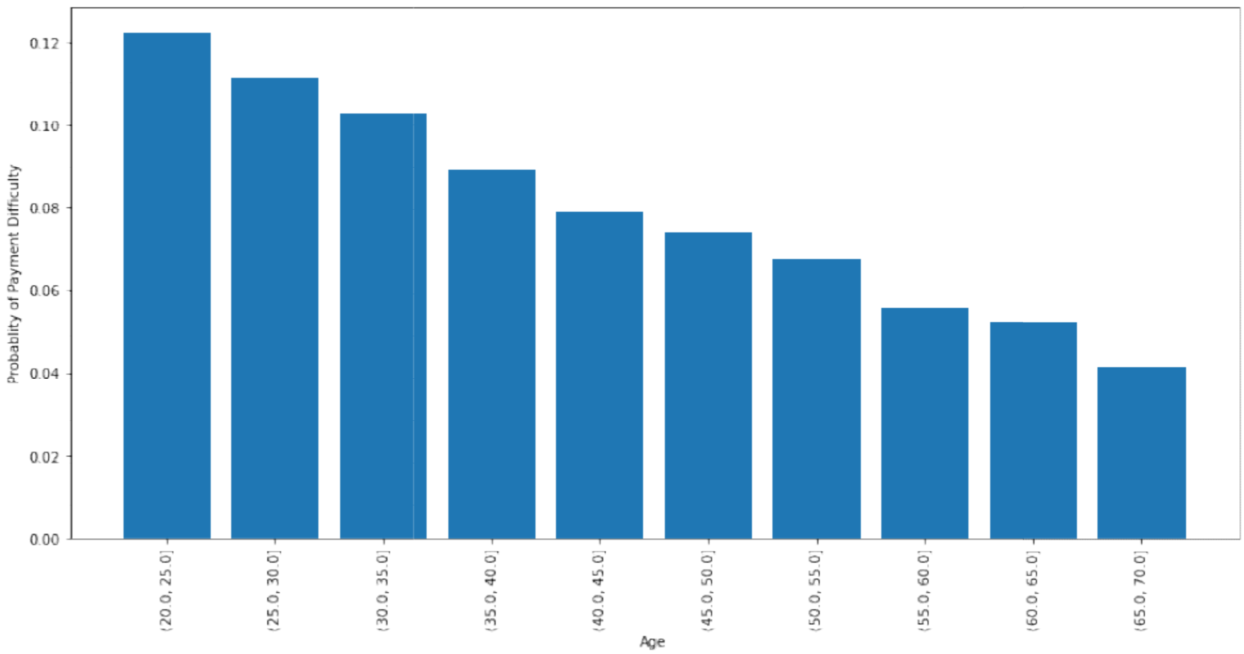
Ananlysis of relationship between education and likelyhood of payment

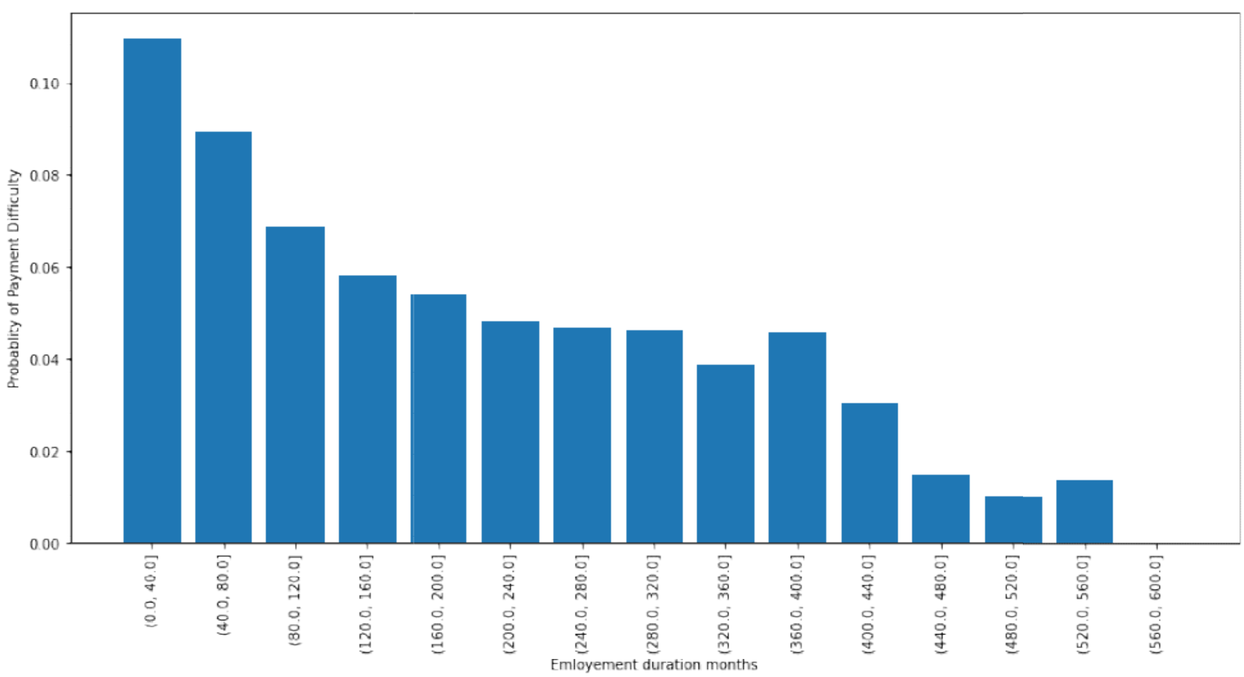


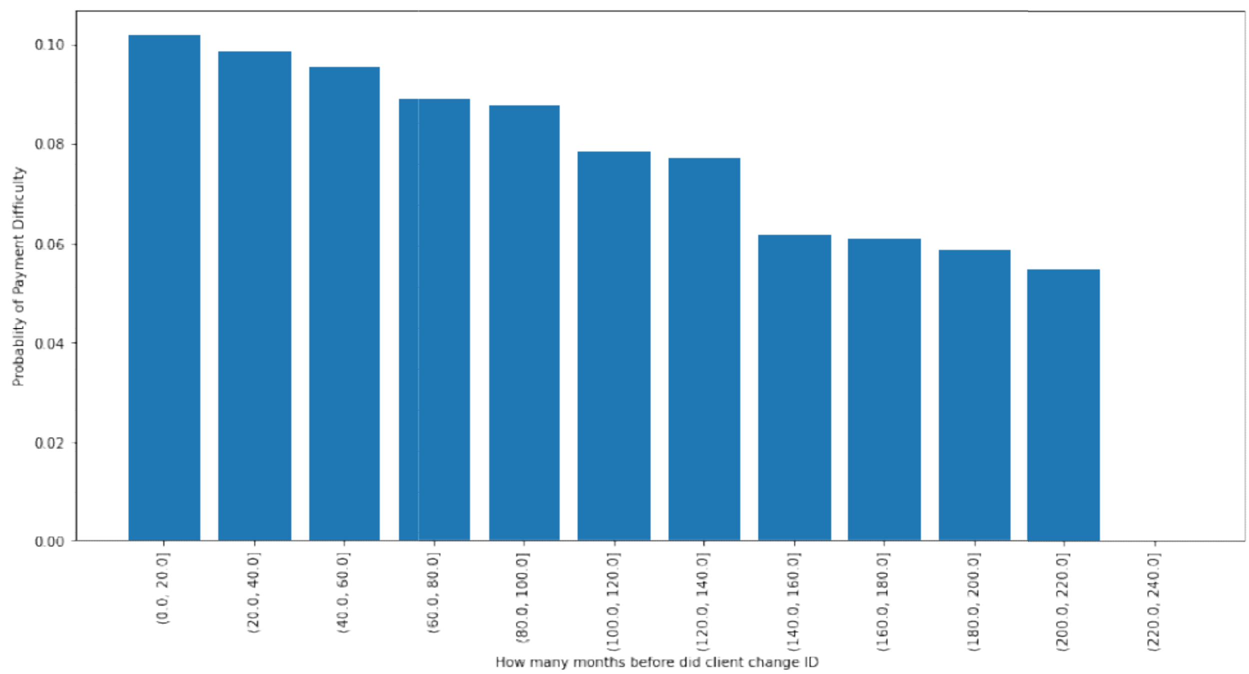
People having a academic degree or a higher education level are more likely to repay

People having incomplete higher or lesser education are less likely to repay

Ananlysis of relationship between age, employement duration, change in ID doc and likelyhood of payment

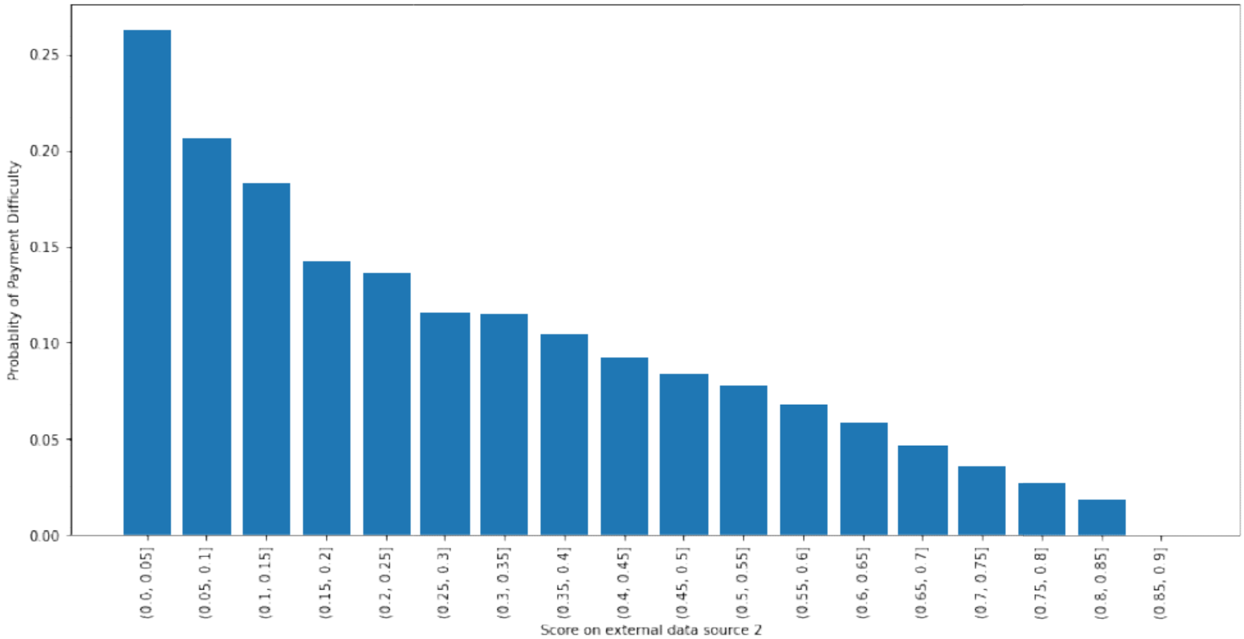


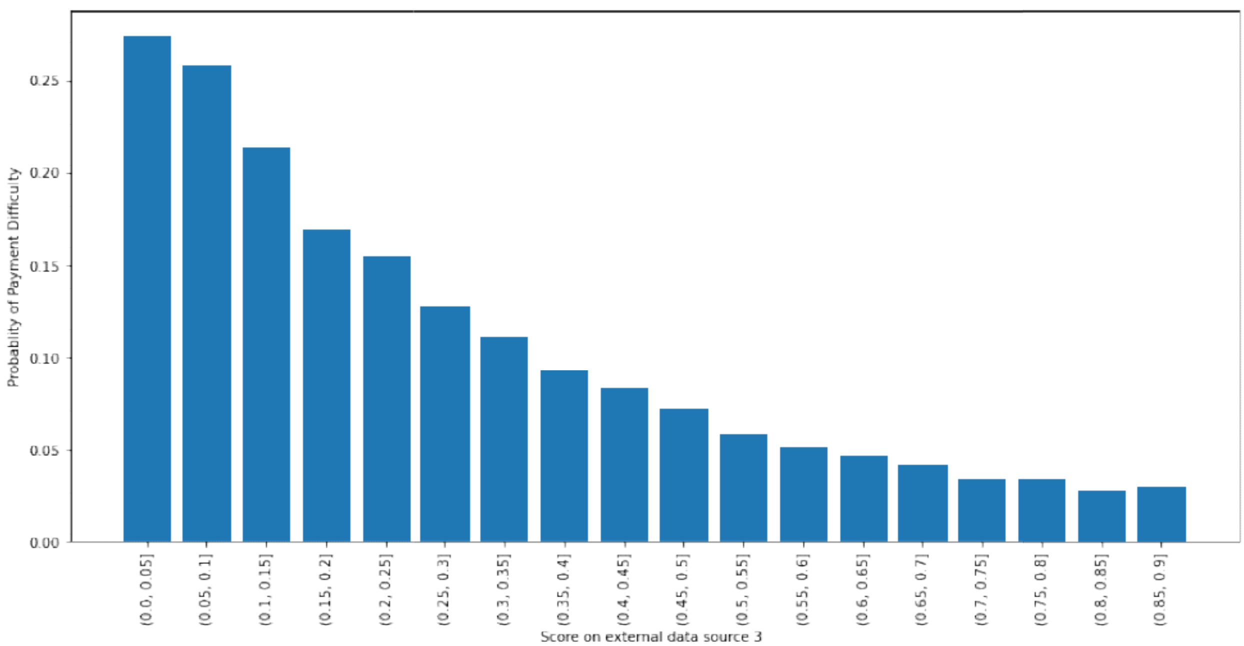




Older people ; people with higher prior employement duration ; people who have not changed their ID in last 100 months are more likely to repay. Since it is the older people who have relatively higher employement duration and who might not have changed ID (because younger people might have recently got an ID),we can generalise that older people are more likely to repay

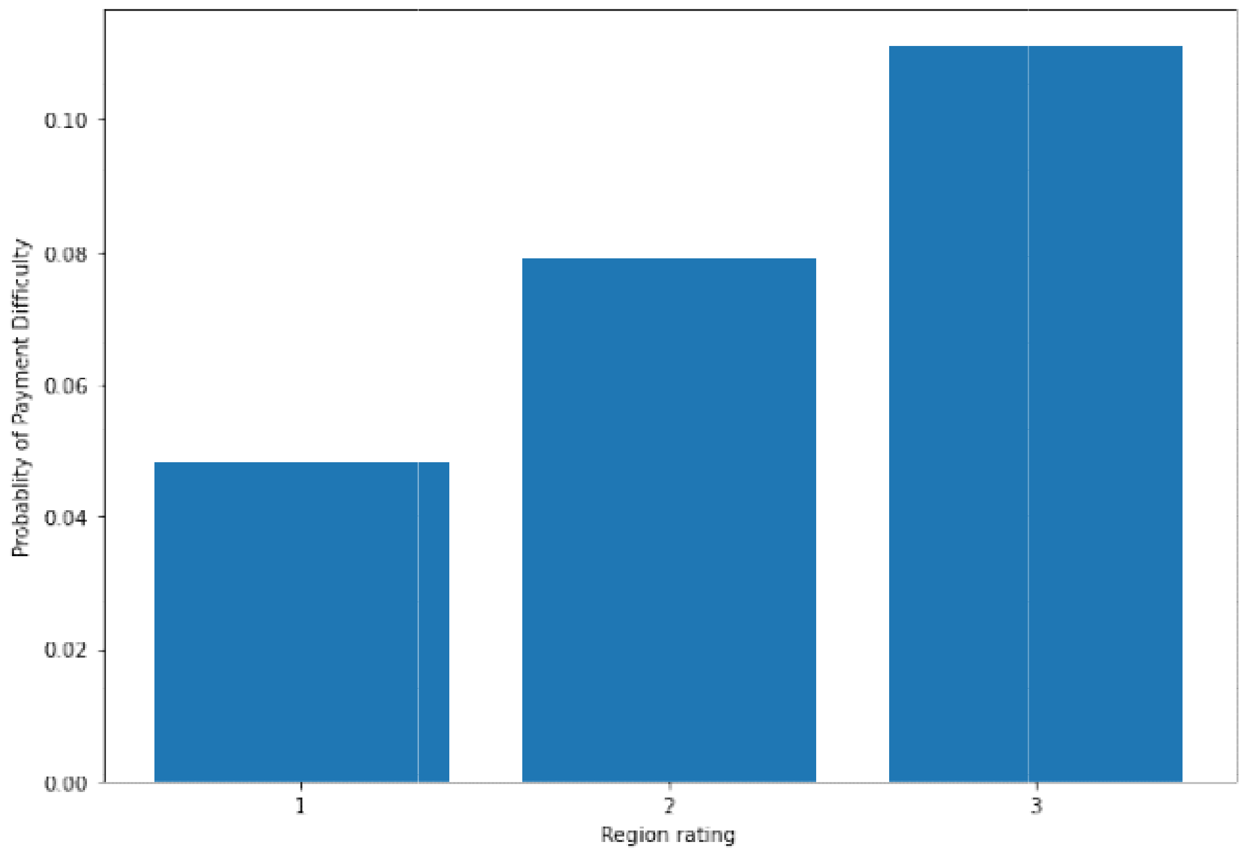
Score from external data source and likelihood of payment





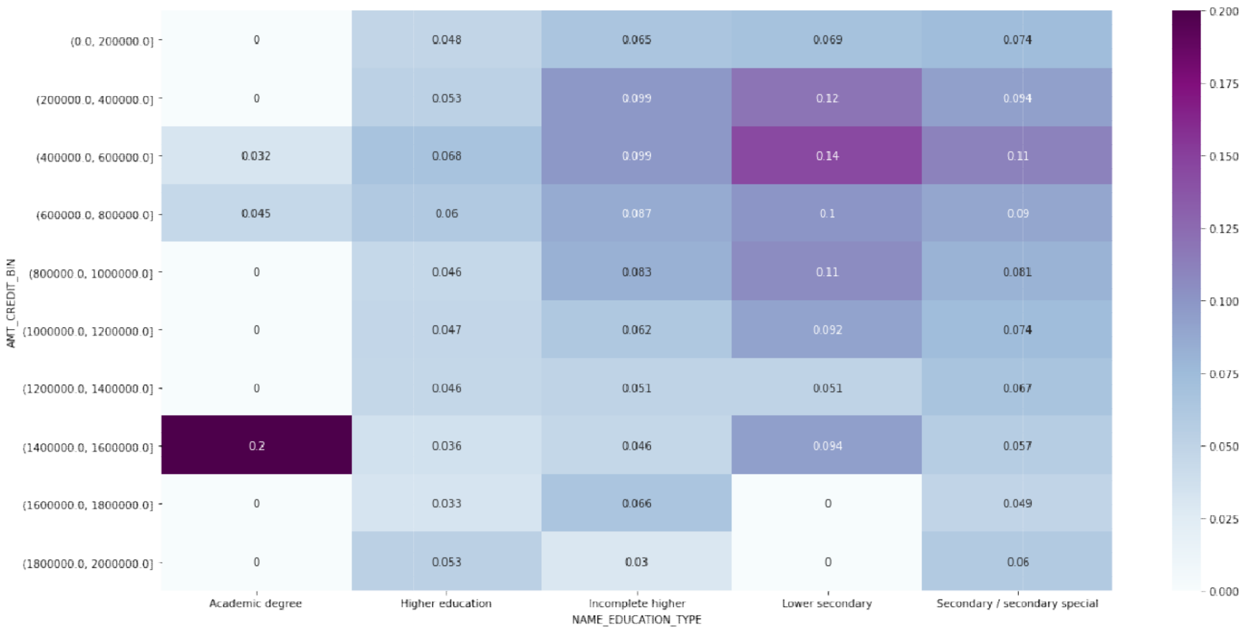
Normalised data from external source are reliable indicators for predicting likelihood of payment. Score above 0.5 indicates more likelihood of payment. Score below 0.4 indicates that applicant might face difficulty in repayment

Region rating and likelyhood of payment



People living in region 1 & 2 are more likely to repay than in region 3

Ananlysis of relationship between Education,ticket size and likelyhood of payment



Earlier we had stated that:

People having a academic degree or a higher education level are more likely to repay

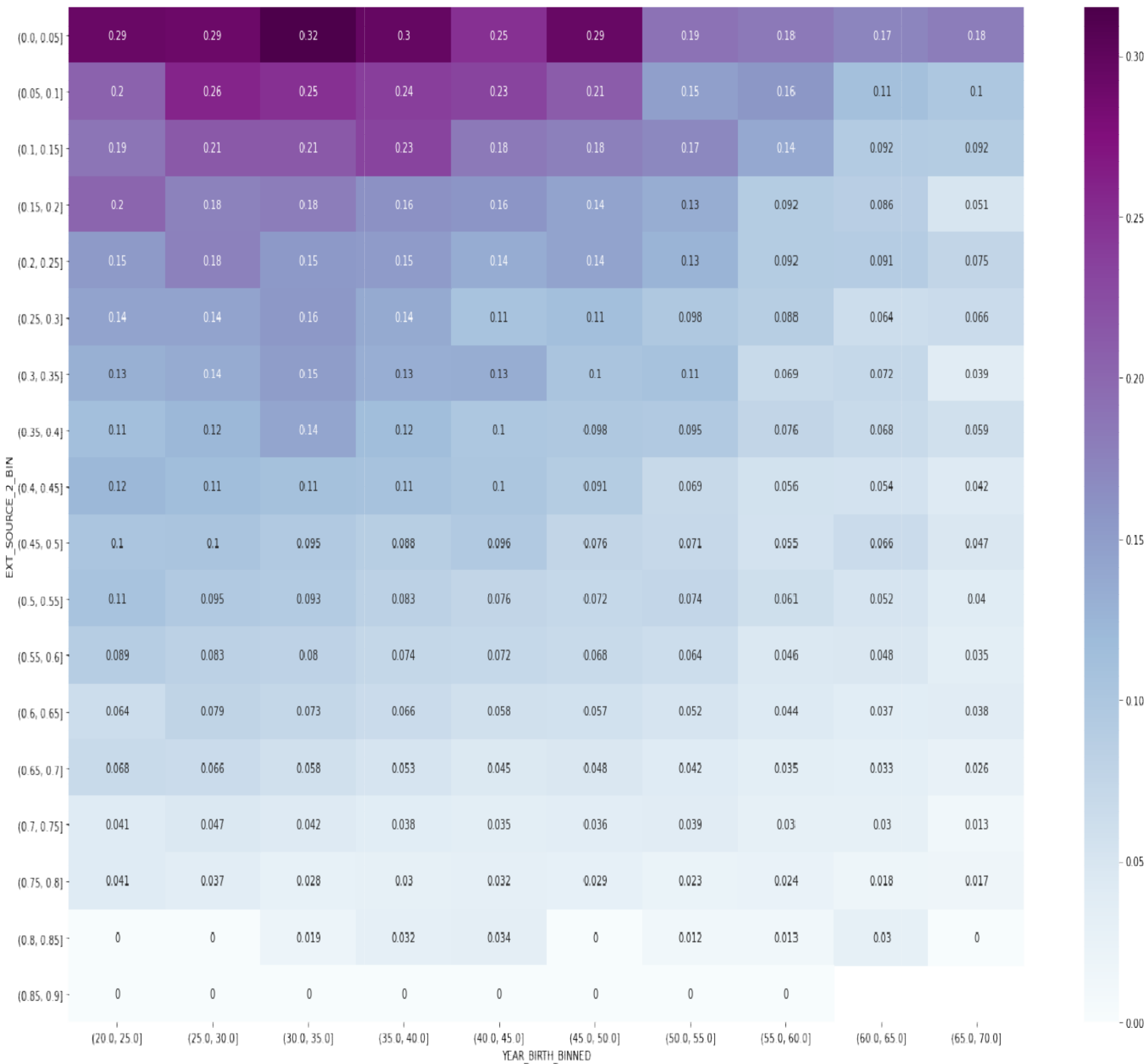
People having incomplete higher or lesser education are less likely to repay

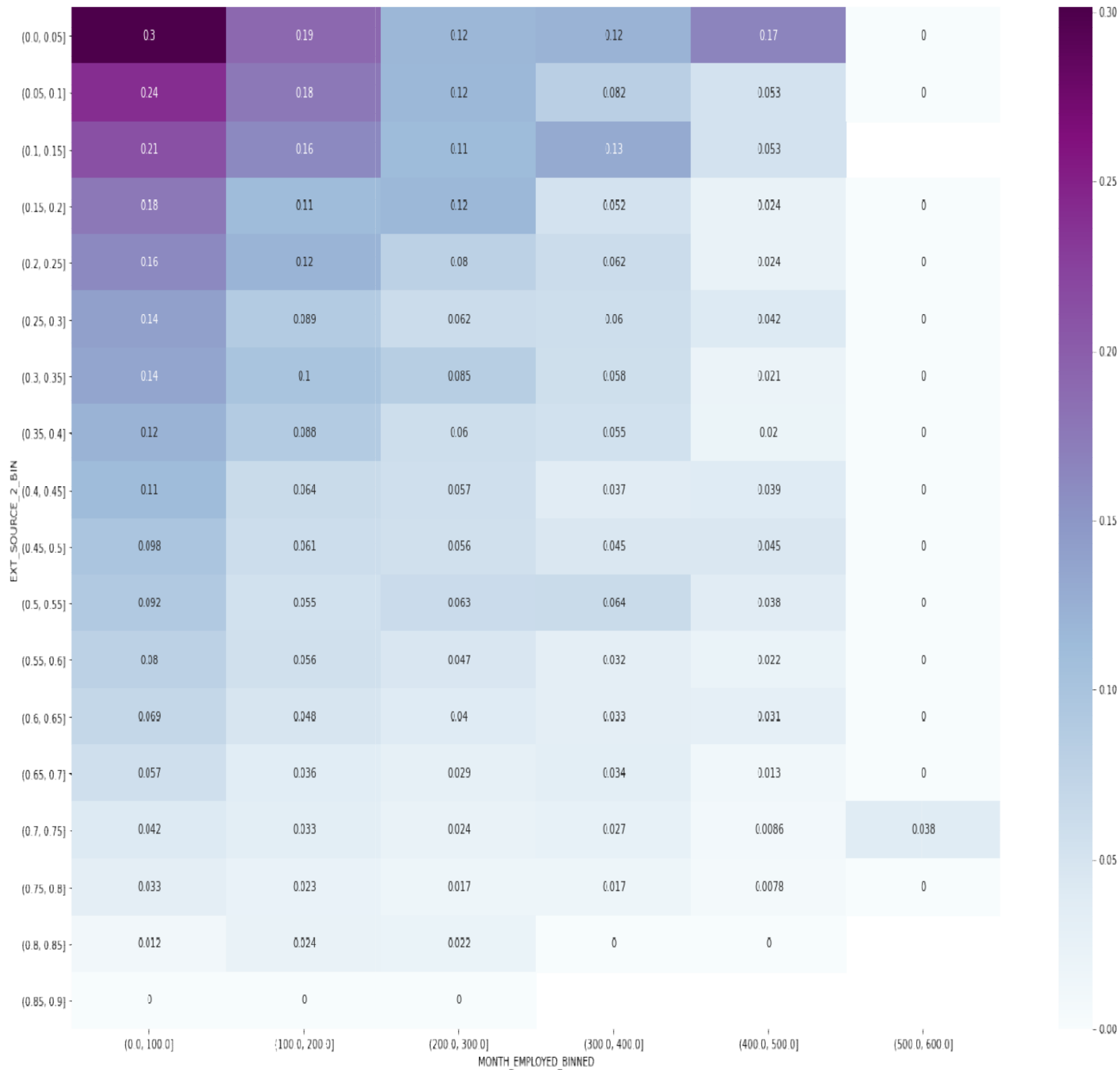
The applicants taking a credit of upto 200000 even when they are less educated are more likely to pay

The checks and balances seem to be in place when it comes to loans of ticket size higher than 1000000.

But in the range of 200000 to 1000000 there is high probablity of default in peole with low education

Ananlysis of relationship between Age ,External Source Data ,Employment Duration size and likelyhood of payment



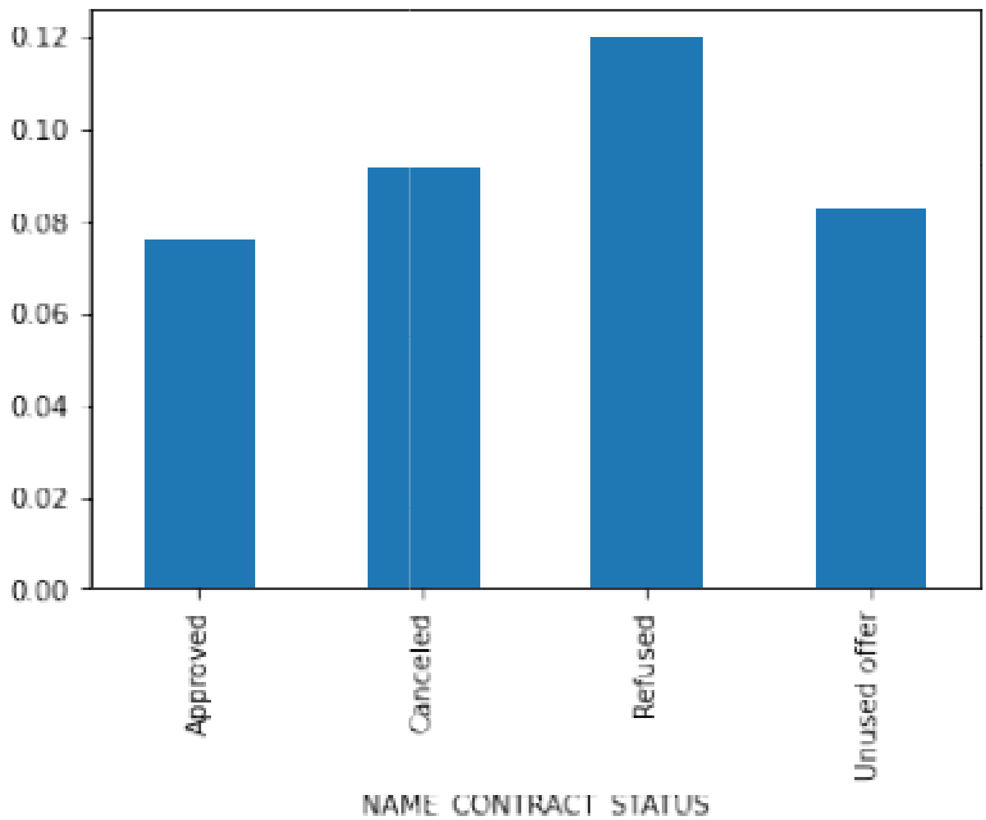


We had suggested that younger people were more likely to default on credit than older people. Looking for the heat maps we can say that for applicants under the age of 40 having score of 0.6 or greater are more likely to pay.

The heat map is dark in the upper triangular portion indicating a very high default rate in the group. Map can be used for reference on case to case basis

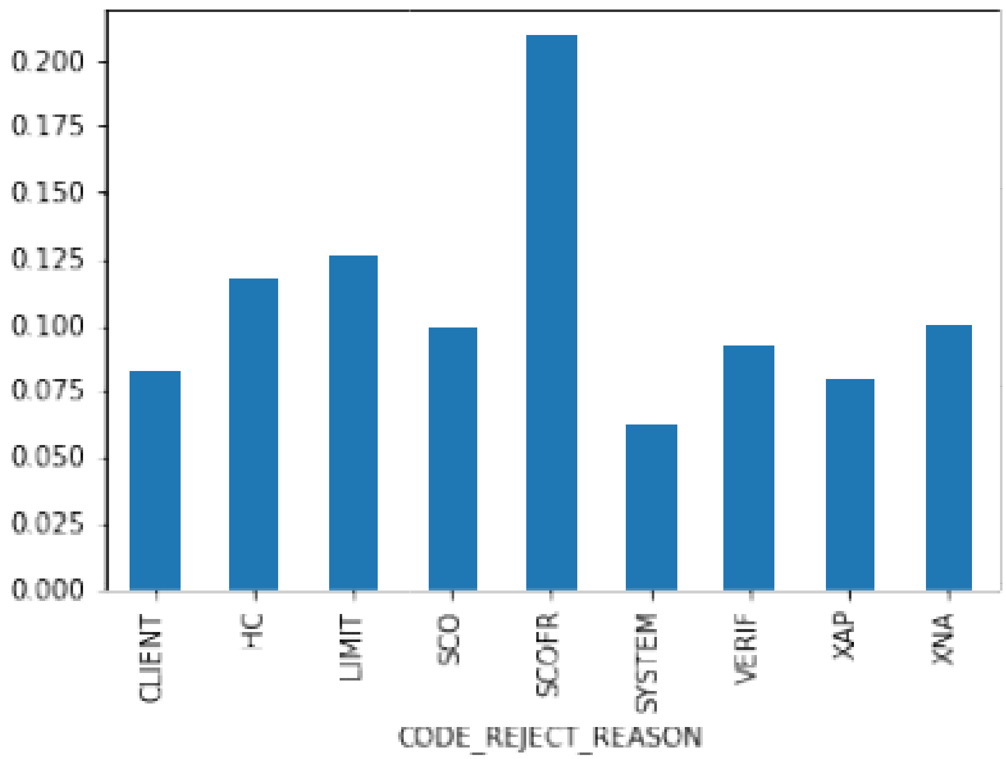
We have the historical application data of applicants. We will use this data for analysis

Analysis of relationship between Contract Status and likelyhood of payment



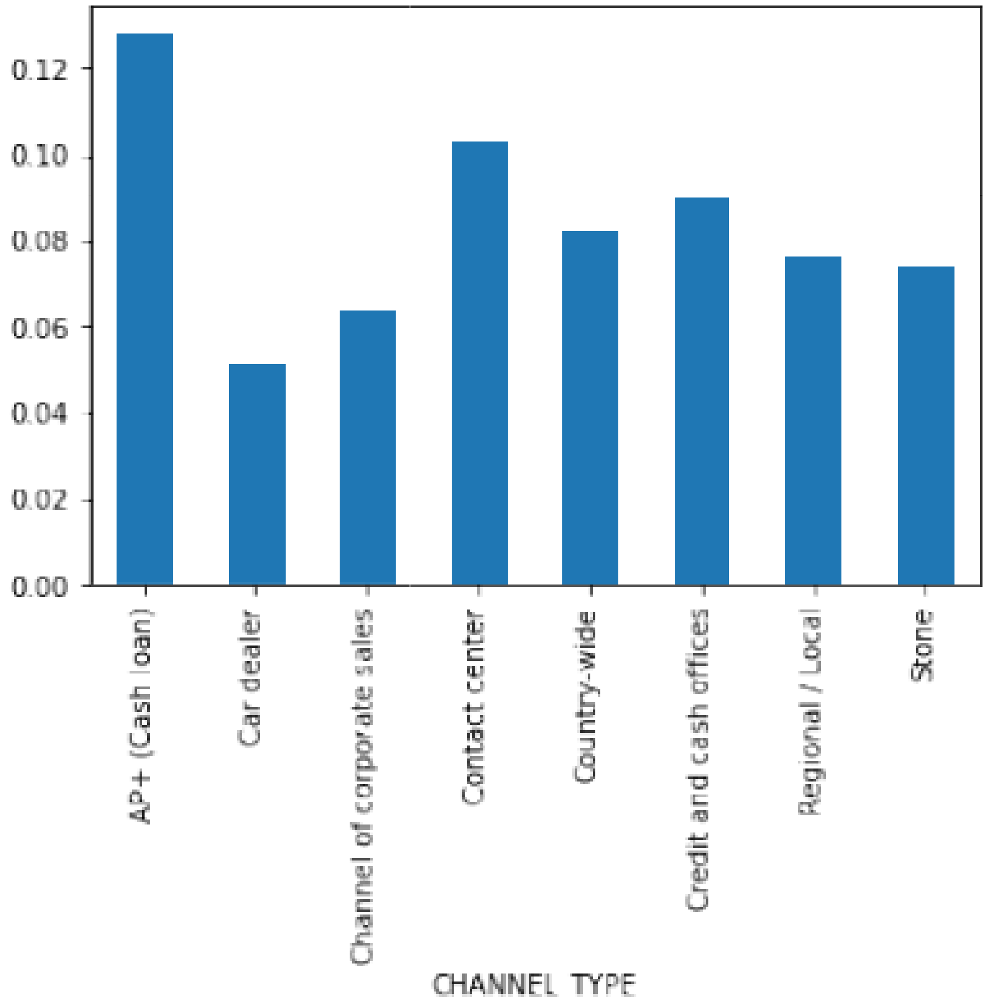
Applicants who were refused a loan earlier were were more likely to default than those who were offered

Analysis of relationship between Rejection Reason and likelihood of payment



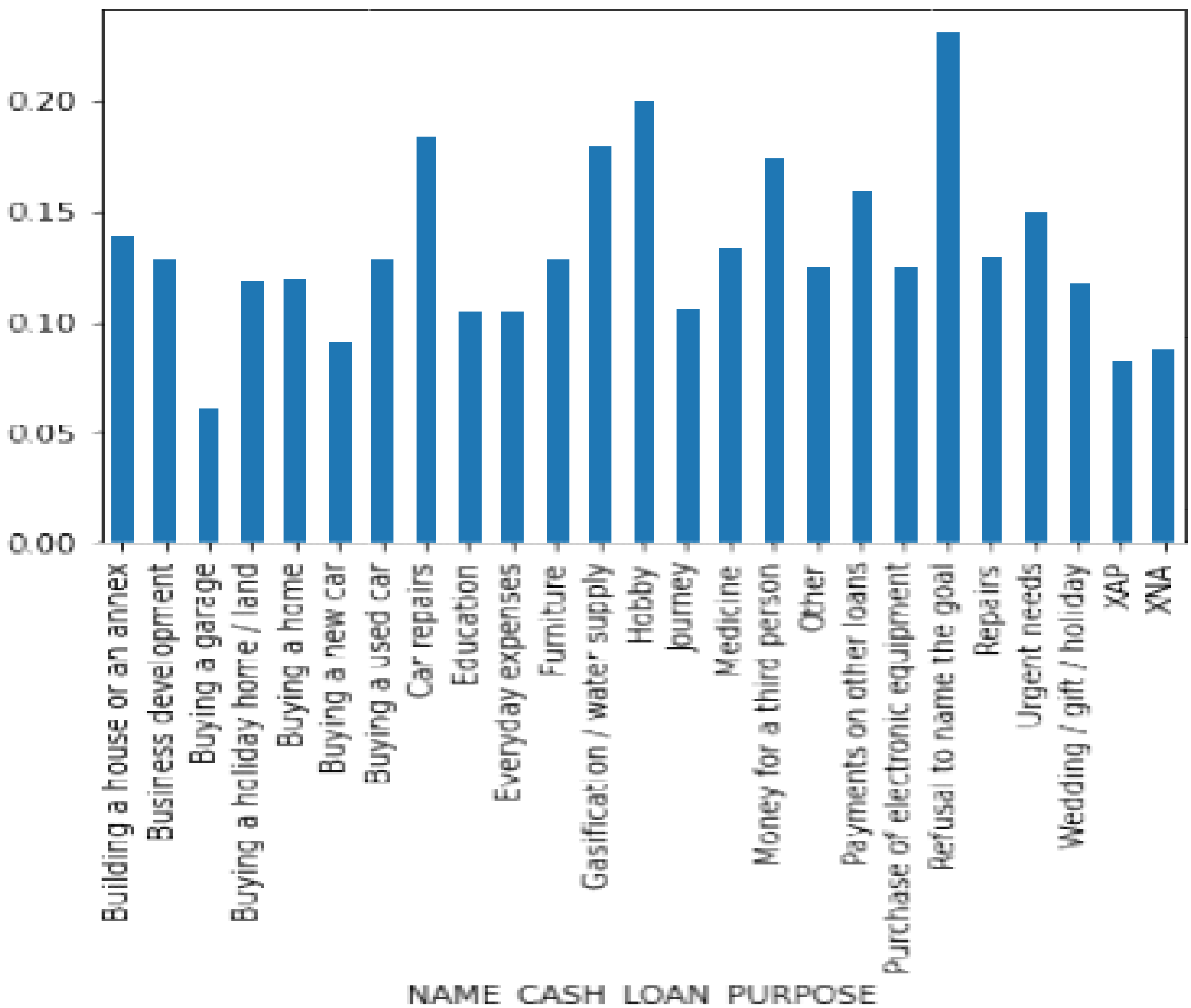
Among applicants whose previous loan applications were rejected, those rejected on SCOFR, HC OR LIMIT grounds were more likely to default.

Analysis of relationship between Channel Type and likelihood of payment



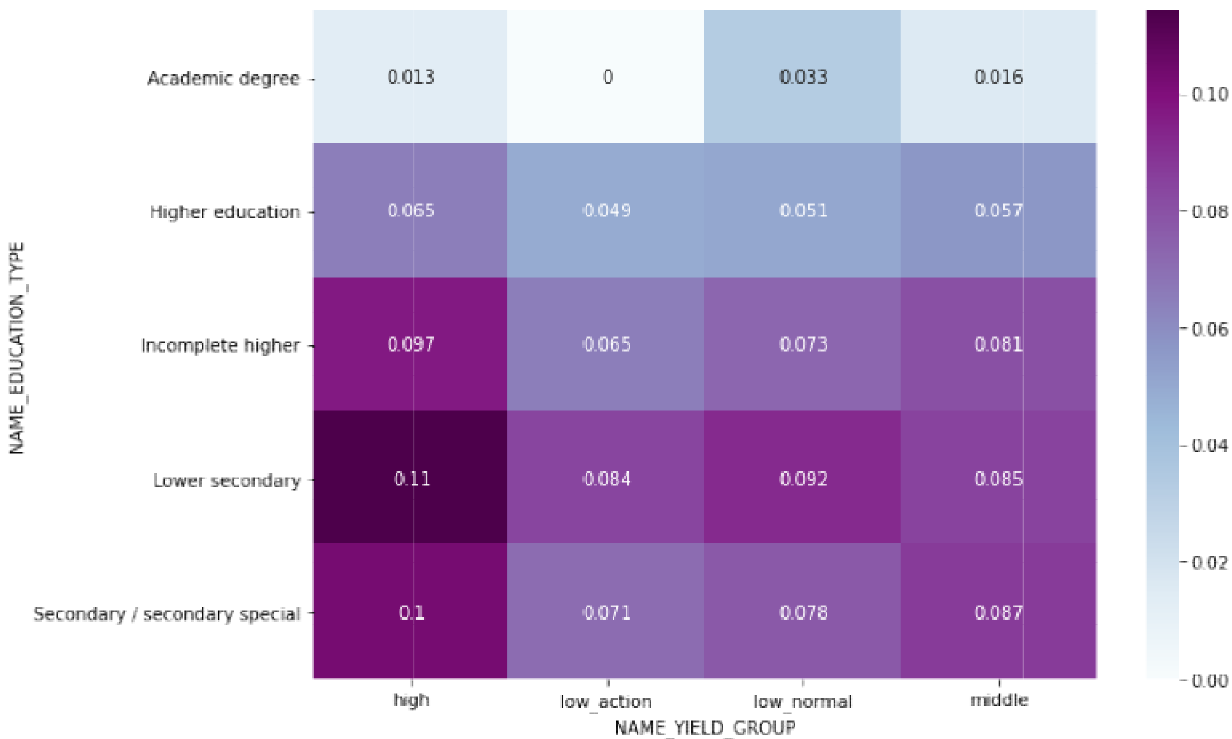
Applicants applying from channel partners for loan dispensing such as car dealers, stone are less likely to default

Analysis of relationship between Loan Purpose and likelihood of payment



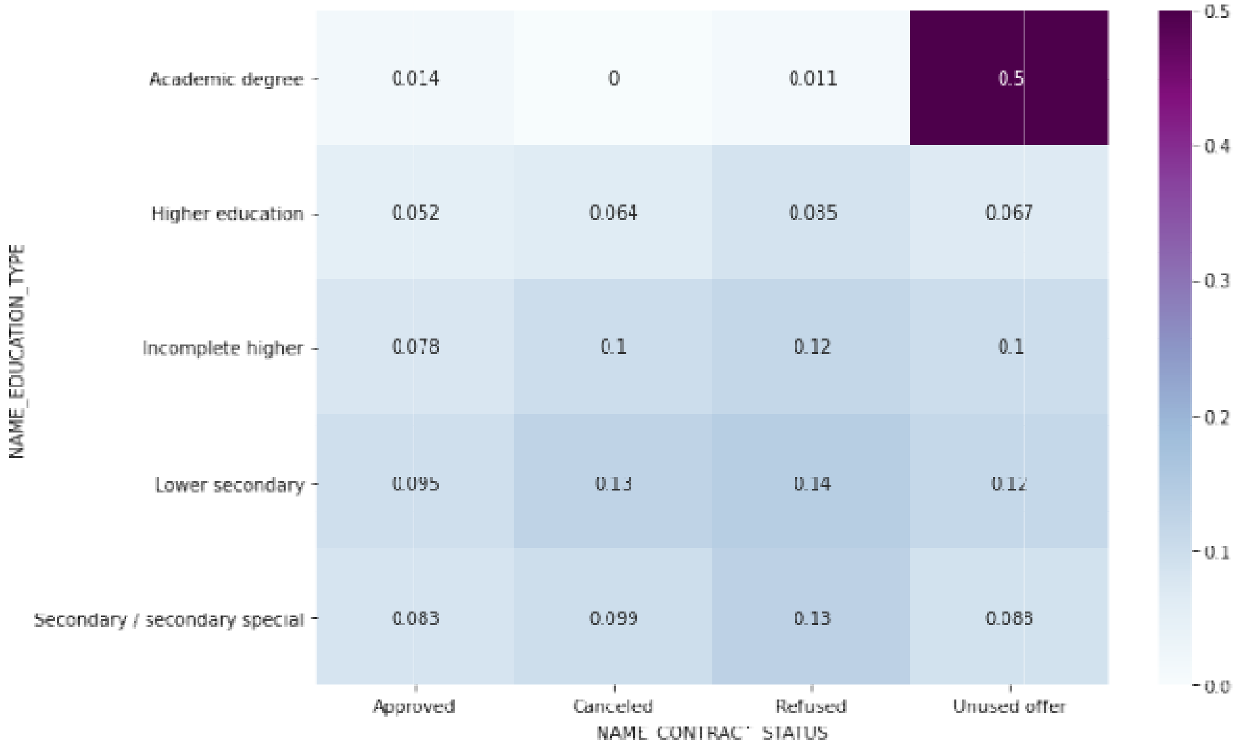
Applicants who had previously applied for purposes such as Car repairs Hobbies ; paying other loans or who had refused to state the purpose of loans were likely to default

Analysis of relationship between Education; Yield Group and likelihood of payment



Earlier we had stated that people with incomplete higher education were likely to default; But those who got a loan with lower yields were less likely to default. The yields can be adjusted appropriately based on above analysis for each group

Analysis of relationship between Education; Application Status and likelihood of payment



Applicants with previous approved loans were less likely to default irrespective of education levels