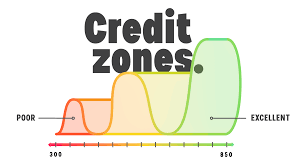
**Credit One**



**Credit Analysis Report Out**

**December 2020**

**Agenda**

1. **Problem Statement**
2. **Project Objective**
3. **Methodology Design**
4. **Key Findings**
5. **Improving Default Scores**
6. **Project Recommendations**

**1. Problem Statement**

An increase in customer default rates is bad for Credit One since its business is approving customers for loans in the first place. This is likely to result in the loss of Credit One's business customers.

**2. Project Objective**

Develop, design, and implement a solution that accurately predicts and classifies if an individual will default or not on a given loan

**3. Methodology Design**

To achieve our objective of developing the right model, the project team used various methods to uncover more information. There were in particular three techniques that were very important to design the model and provide valuable insight to the solution. These were:

**Data Visualization:** specifically insightful the analysis on what factors impacted the default status. It helped built a customer profile to understand who is more likely to default.

**Correlation Matrix:** very important as it provided an initial understanding of what variables were important or not for the model, and helped trimmed the dataset to improve its performance.

**Recursive Feature Elimination:** this technique provided an additional insight into which variables were statistically better and therefore improved the model accuracy.

**4. Key Findings**

**Variable Correlations**

Based on the analysis of the data it was observed that there is no strong correlation between the dependent (Default) and independent variables.  This means that there is no strong relationship, and that in fact variables are hardly related to it.  Nevertheless, collinearity was observed between independent variables.  This is important to identify and take out as it could cause problems when fitting the model and interpreting the results.

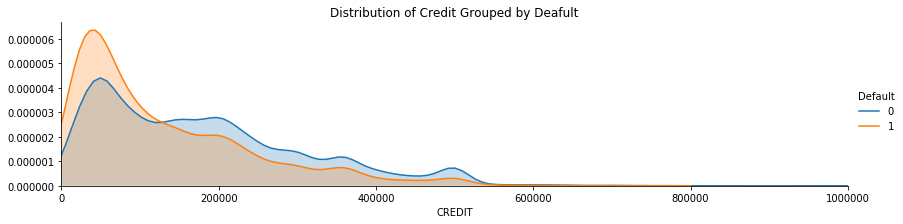
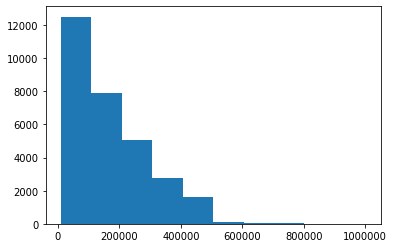
**Statistically Important Variables**

All variables are not equal, and the analysis demonstrated that the most important variables to develop a predictive model were all those related to customer characteristics and historic trend.

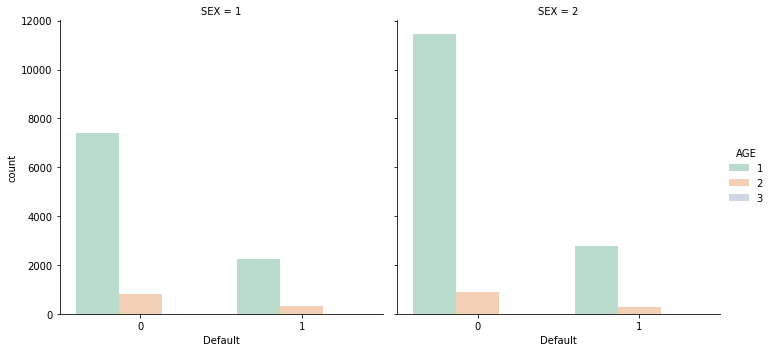
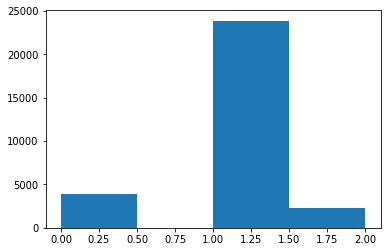
|  |  |  |  |
| --- | --- | --- | --- |
| **Model Key Variables** | | | |
| Age | Sex | Education | Marriage |
| Credit $ Amount | Historic trend (May, Jul, Aug, Sep) | | |

**Credit Amounts and Age**

From a credit perspective, it was noted that the majority of the loans given are under $200K.  Following a similar trend, people default more on smaller loans (under $100K) rather than in the larger ones.



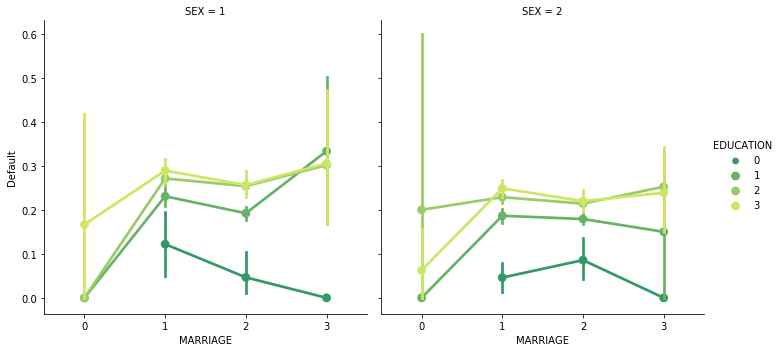
In case of the age variable, most of the loans are given to customers that range from 25 to 50 years old (Bin #1).  It is not a surprise that most of the people that defaulted are under this age range, regardless of sex or other variable.



**Age:** 1. 0 to 25, 2. 25 to 50, 3. 50 to 75, 4. 75+

**Customer Profile - Default**

Based on the previous analysis, default loans are centered on lower $ amounts and from customers between 25 to 50 years old.  By analyzing the other profile variables, we are able to provide an estimation of what type of customers are more likely to default their loans.



**Default:** 0. No default, 1. Yes Default **/ Sex:** *1. Male, 2. Female* **/ Marriage:** *0. Other, 1. Married, 2. Single, 3: Divorce* **/**

**Education:** *0. Other, 1. Graduate, 2. University, 3. High School*

In the graph the higher the “Default” criteria, the more likely scenario that a person will not pay their loan

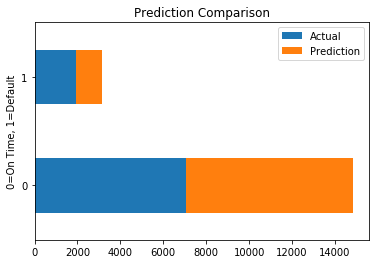
The analysis above shows us that divorced males with “Graduate”, “University”, and “High School” education are the most likely to default their credit.  Important to notice too that single males with the same education are not far behind.

On the other hand, divorced females with “University” and “High School” education are more likely to default, following a similar trend to their male counterparts in the next categories.

**5. Improving Default Scores**

The team designed a solution to improve Credit One default scores. A model was created to properly classify potential candidates as loan subjects or not. It provides a proactive analysis before a credit is given.

The model uses the variables outline above, and with an accuracy of around 80% it can detect potential problematic customers or not. Below you can see the test results.



**80% Accuracy**

The tool is intended not to be used in a silo, it should be complimented with the use of the customer profiling explained above and the expert knowledge of the company.

**6. Project Recommendations**

**Customer Profiling**

As we analyze the data, it becomes apparent that being able to profile customers or potential clients becomes an important tool or exercise for the business.  It provides valuable insight to support credit decisions in terms of the amount and type of credit, but also if a person is subject to a loan or not.  This will improve the default indicators the company is facing.

**More requirements for smaller credits**

The data is showing us that most of the defaulted credits are under $100K.  Additionally to the profile exercise, the company could add additional requirements to people requesting this type of loans, like warranties or guarantors.  With this action the company could cover itself, and limit the potential loss.  It could also recover all or part of the defaulted loan.

**Model Implementation**

As a continuation of the customer profiling, the next step would be to run the model to quickly predict what is the likelihood of the person to default or not. This would give Credit One an additional tool to assess and make a decision. Tool has an estimated accuracy of around 80%, so additional recommendations as the ones exposed here, plus the company´s expertise would be key to drive the most appropriate decision.