Credit One

Memo

|  |  |
| --- | --- |
| To: | Guido Rossum |
| From: | Yiching Huang |
| Date: | 2018/05/25 |
| Re: | Lessons Learned Report |

As you requested, this memo presents the result of my lessons learned from projects based on Python. I will detail the key takeaways that I learned when operating Python and potential business value from this analysis. I will separate the following discussion into different sections based on learning projects.

**Exploratory Data Analysis (EDA)**

## EDA is an important data analysis step before we dig into doing any model training, testing and prediction. EDA can help us to familiarize the data set, observe the data quality, and find out what is the important variables in the data. It is always the priority to understand our data before moving to the next steps. The main idea of EDA is to use different visualization ways to help us understand the data from a variety of aspects efficiently. We used the EDA method to explore the default of credit card clients and have provided a summary below.

* We first explored the characteristics of our clients and focused on the relationship between education, sex and default. We counted the data and figured that female with university degree has the most defaults in our data. However, it’s more precise if we look at the default ratio instead of the count number. We noticed male with university degree (26%) and male with high school degree (27%) has the highest default rate based on the data.
* Second, we bin age into decade and create a new variable (age\_by\_decade) to help us explore the data efficiently. We noticed most of our clients are in age 20-40, and if we compare age and education with default number, we can see clients in age 20-40 who has university degree has the most default number. As we mentioned before that might be because most of our clients are from that range, so we also need to see the default ratio.
* Third, we also explored the distribution of limit balance, bill amount 1, and pay amount 1. We noticed most of the clients’ limit balance is less than 200,000, and bill amount and pay amount is less than 50,000. From the three distribution charts, we can see the default and no default curve has almost the same shape.
* Next, we move to analyzing factors that affect the default. We tried different variables and noticed the client that has other education degree in our data set has lower chance to default. The result is consistant when we group by age and sex. Thus, it is important for us to find out what kind of education degree is for this observation, is it doctor degree or others? We also look at the history of past payment and figured when clients have records of no consumption, paid in full and the use of revolving credit, the chance to default is low. However, as client past records turn to payment delay 1 and 2 months, we can see the chance to default increase greatly.
* Last, we see the relationship between variables in the linear plot. We noticed as the payment record increases, the chance to default increases, too, which makes sense to us since if clients have records for delayed payment, it means they are failing to meet their obligation in time. Also, we look at the relationship between default and pay amount 1 grouped by sex. Both sexes showed, as pay amount gets larger, the less chance the client will default.

**Recommendation**

## EDA is a useful tool before we start analyzing data and helps us to find interesting facts that help us for solving the business problem. As we mentioned before, it is necessary that we do EDA first to understand our data and become familiar with it. After all, if we don’t understand the data, it is hard for us to find the right answer to solve the business problem. This task helps us to review clients’ data for us to understand which affect the default in order to help us gain back our customers and enhance the revenue. We also can use EDA in a different task, too. We also can provide the service to our partners to help them decide which is their next valuable customer that they should acquire. We can help them look for customers that use credit cards frequently and always pay in time. For this kind of task, we might not have their payment record since they are totally new customers to us. However, we can collect other information such as yearly income, number in househols, region where clients live, or credit score. With this information and the EDA process, we can find interesting facts for business insight.