

Task 1 --- Framing:

Understanding client, available data and exploratory data analysis.

By Team 12

On the basis of our understanding to the data of Lending club, what we need to complete is educate our client on the insights we draw from the data. This report will be divided into three parts.

- ◆ Research and summarize what are the client's needs.
- ◆ Conclude keys from the graphs which Lending club has generated.
- ◆ Explore the data and comment on the data quality and features.

Part 1: Research and summarize what are the client's needs.

Our client is an economic consulting firm which interested in understanding how customers of lending club can provide leading indicator on the economy and lending. In order to understand our client, we make an explanation about what is main economic consulting. Economic consulting is a type of paralegal work, it offers expert to testify in litigation matters related to economics and finance. In practical terms, it means economic consulting firms produce research, analysis, and testimony to be used in court cases.

Take SMITH Economics Group as an example to analyze the client's needs. SMITH Group deliver top-notch economic and financial analysis, testimony, problem solving and creativity. Now, it wants to obtain leading indicator on the economy and lending, and it chooses to complete this case study according to the dataset of customers of Lending club. Therefore, our goal is to find the most stable customer's attributes at the same time they can get more profits as possible. At last, we feedback our model about predict rates and teach client how to use the model.

Part 2: Conclude keys from the graphs which Lending club has generated.

At first, we summarize the three charts about Lending club's overall situation.

- The overall scale of the platform's main point is the total number of loans per year. According to TOTAL LOAN INSURANCE given by the platform, we can know that the total loan issuance is gradually increasing every year, and the rate of increase is accelerating. It represents that the company still has a lot of space to improvement and the market is still not saturated.
- According to the lender's main purpose, we can know 63%, more than half of them are for refinancing existing loans or pay off their credit cards.
- Depending on the total amount of loans in different states, we can briefly judge that the degree of development is probably related to the economic level of each state. But if we want to analyze it further, we need to calculate the parameters such as the population density of each state.

Second, there are two factors influence the Adjusted Net Annualized Return, the average age of portfolio and diversification of investment.

- According to the Average Age of Portfolio, we can know that the higher average age of the Notes in the portfolio, the more stable the annual interest rate of the investment. Basically, after nine months, you can get a return approximately 5 percent. On the one hand, we can simply summarize that investors have determined their investment methods based on previous investment experience, and in order to maintain stability, their interest rates have also stabilized.
- In addition to the investment time, another aspect we need to pay attention to is the diversification of investment. Diversified accounts have generally experienced less volatility and more solid returns than investors with more concentrated holdings. So, for risk-averse investors, it is a good idea to choose to invest their own funds separately.

Third, there are four charts to describe different grade's development and composition.

- The average annual interest rate for different grades has always had a clear dividing line. Now the interest rate between different grades difference 4%. B is higher 4% than A and C is higher than 4% than B.
- The composition of different levels also has significant changes. There are more and more people choosing ABC grades with high stability. High-risk levels are gradually abandoned by the public.
- According to the NET ANNUALIZED RETURN BY VINTAGE, returns are gradually decreasing every year.

Part 3: Explore the data and comment on the data quality and features.

All the data we have including 890 thousand records formed by 75 attributes. The data dictionary provides ample explanation for the meanings of each attribute. We need to understand the data dictionary at first, and then manual analysis the dataset.

At first, in order to complete data cleaning, we should know that real world data tend to be incomplete, noisy and inconsistent. Data cleaning routines attempt to fill in missing value, smooth out noise while identifying outliers, and correct inconsistent in the data. There are 78 variables in the data dictionary, and with 4 meaningless which means there are no corresponding attributes in the table. And then we can find there are 20 variables are mainly empty so we can delete it first. There are 55 remaining attributes to analysis.

At second, we classify dataset's variables according to the actual meaning of the attributes and delete meaningless attributes. We divided the remained 21 attributes into three classes.

- The data related to the member's basic information and inherent capital.
- The data related to the grade and interest rate.
- The data related to the existing lending record's state.