DataDriven
PORTFOLIO SOLUTIONS

Stage A Report: Business Understanding

Our team is analyzing the potential of integrating peer-lending into investment portfolios. We aiming to answer key questions posed by GreatYields's CIO. While progress has been made, deeper investigation is necessary to fully understand some complex issues. We'll provide time estimates for each question that requires further investigation to ensure transparency. We remain committed to delivering valuable insights and look forward to presenting our findings soon.

Question 1 - What are the expected realized returns for the different loan grades? How are the returns distributed for each grade?

We aim to determine the expected returns for different loan grades by analyzing the "Charged off" and "Fully paid" returns. This approach provides us with a comprehensive understanding of the loan life cycle, allowing us to calculate the expected realized returns more accurately.

To calculate these returns, we use the Rlong formula, which considers the total amount paid and the amount committed to the loan.

Yield for a single loan -

$R_{long} = (total\ paid/loan\ amount) - 1$

*Loan amount-The total amount of the funded loan

*Total paid-the total amount committed

R_{long} refers to the return for the entire loan period, so we need to convert it to an annual return, this will allow us to compare the returns on different loans with varying terms and better understand the overall performance of each loan grade.

$R_{\text{year yield}} = \left(R_{long} - 1\right)^{\frac{1}{actual\ loan's\ life}} - 1$

*yield for one year

Calculation for expected return for each grade:

$E[R]g = \sum_{i=1}^{n} R_i * \frac{1}{N_i}$

*R_i = return in scenario i

*N_i = number of scenarios

To visualize the distribution for each loan grade, we will need to calculate the standard deviation. This can be achieved by performing the following calculation:

$\sigma = \sqrt{\sum_{i=1}^{n} \frac{1}{N} (R_i - E(R)g)^2}$

*N = size of population

*R_i = return in scenario i

*$E[R]g$ = Expected return by grade

Based on the findings presented in Appendix A, we have identified the relevant columns and confirmed that the data is suitable for our intended analysis.
The team requires 1-2 weeks to complete this stage and provide the results.

DataDriven
PORTFOLIO SOLUTIONS

Assumptions & potential issues:

1. Some loans were not paid in full, and some were paid ahead of time, which means the realized return may not necessarily be the stated interest rate.
2. After we find the expected return for each grade, we can convert it to a real rate using the Fisher equation, using the inflation rate provided to us by the US Federal Reserve.
3. The holder bought the loan from the issue date until maturity.

Question 2 - Are the available loan data informative, thus can help selecting loans to invest in (i.e. can the data help choose loans better than random selection, or selection by simple criteria, e.g. loan grade)?

At this time, we cannot definitively say whether the available loan data can be used to select loans to invest in better than random selection or by using simple criteria such as loan grade. However, by modeling the data, we may find patterns for better decisions. We'll define a target variable (e.g., the expected returns of loans or whether a loan is expected to give returns over 2%), preprocess the data, assess its variability, and build a predictive model to evaluate its performance. We'll need 2-3 weeks to complete this and gain a full understanding of the loan data's informativeness.

Question 3 - If the data are indeed informative, what increased performance can be expected, compared to a baseline of simply selecting loans based on their ratings (grades)?

It is impossible to determine the potential impact of the loan data on our investment strategy without constructing and evaluating a model (requires 2-3 weeks). However, using advanced techniques to analyze loan data can yield valuable insights into underlying risk and return factors, potentially revealing patterns and correlations beyond what loan ratings alone can provide. By developing a predictive model that incorporates multiple features beyond loan ratings, we may be able to identify loans that are more likely to meet our target return threshold of over 2%. This approach would enable us to better manage risk and optimize our investment strategy, leading to a stronger and more profitable investment portfolio.

Question 4 - What "average" returns can GreatYields expect from investing in peer lending loans(on our platform)? Keep in mind that ultimately, the goal is to maximize returns (i.e. make as much money as possible).

Currently, we are unable to provide a precise estimate of the average returns for GreatYields, as we still need to fully process and analyze the data. We expect to need 2-3 weeks to calculate the realized returns for various loan grades, as requested in the first question, which will establish a starting point for expected returns. While P2P lending typically yields an average return of 1% to 7.5% according to general information available online, this may not accurately reflect our specific loans and platform. We may be able to improve loan selection and increase yield by utilizing a machine learning model. However, we will need to build and test the model to gain more specific insights into the average returns for GreatYields. It's important to consider the risk of loan defaults and set a budget accordingly.

Potential issues:

1. Data from 2019 may not reflect current market conditions.
2. Possible insufficient investment opportunities (over 2% return) relative to available funds.

Question 5 - What is the risk level entailed in such investment (as measured by the volatility)? (how to calculate) - by grade

Since we don't have access to the series of payments for each loan, we cannot calculate volatility in the traditional way by taking the standard deviation of returns and multiplying it by the square root of the number of periods in the time horizon. Instead, we can use the SD for each loan grade (calculated in Question 1) as our volatility and then we can compare them to similar investments like Softlending, bonds, and stocks to determine the relative risk of investing in peer lending loans on our platform. Another option would be to estimate the probability of loans becoming charged off by using ML method and use this as a measure of risk.

Assumptions & potential issues:

1. We have data on other investment tools.
2. It's important to note that while higher returns may indicate higher risk, this is not always the case, and a thorough analysis is needed to fully understand the risk involved in any investment.

Our team will spend the next two weeks analyzing the data using computer programs, which involves collecting and cleaning the data, and addressing potential issues. Afterward, we will provide a detailed report and presentation of our findings, along with our recommendations for next steps.

Best regards,
The DataDriven Portfolio Solutions Management Team

DataDriven
PORTFOLIO SOLUTIONS

Appendix A:

Value	**Relevant Column Names**
Loan amount	total_pymnt
Total paid	funded_amnt
Num of periods (years term)	last_pymnt_d-issue_date