DataDriven
PORTFOLIO SOLUTIONS

Stage F Report: Evaluation/Deployment

In this final step of our project, we aim to address the business questions outlined in our initial stage and provide comprehensive answers to the inquiries raised by Walter. Our focus extends beyond statistical and scientific evaluation to uncover the business significance of our findings. By delving deeper into the outcomes of our modeling and evaluation stage, we can derive practical applications and suggest tangible ways to apply the results. Through our meticulous investigation, we strive to provide valuable insights that can guide decision-making and bring about positive impacts within the context of our project. This report serves as a culmination of our efforts, bridging the gap between data analysis and actionable recommendations.

Project Development Overview:

Over the past ten weeks, our project has progressed through a number of key steps. Initially, we focused on preparing and cleaning the data, during which we uncovered many potentially informative columns (refer to report B-C for more details). Subsequently, we moved on to building the model, fine-tuning its parameters, and conducting statistical evaluations (see Reports D-E for in-depth insights). Our current model operates in two stages: firstly, it classifies loans above 2% while providing the corresponding probability, and secondly, it utilizes a regression model to predict the numerical yield. We believe that the combination of these two models will offer a comprehensive understanding of the model's ability to improve loan selection by grade and integrate peer lending into GreatYields' portfolio. This report addresses Walter the CIO's questions and evaluates our model from a business perspective. Please note that while we have chosen to present the answers to your questions in a slightly different order, our intention is to provide a comprehensive and coherent understanding of the topic from a business perspective, to highlight the logical sequence of events and offer unambiguous answers that align with the broader context of the subject matter. We believe that this approach will enable a clearer comprehension of the information presented and enhance the overall effectiveness of the report.

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

To begin our analysis, we will outline the approach used to determine the expected realized returns:

$$E(x) = \frac{12}{T} * \frac{1}{f}\left\{\left[\frac{\mathrm{p}}{\mathrm{m}}\left(\frac{1\text{-}(1+i)^{\mathrm{m}}}{1-(1+i)}\right)\right] * (1+i)^{\mathrm{t\text{-}m}} - f\right\}$$

$T\ -\ Fixed\ time\ horizon\ in\ months$

$F\ -\ Funded\ amount$

$P - Payments\ recoveris$

$\mathrm{M} - \mathrm{Actual\ duration\ (issued\ date\ to\ last\ payment)}$

$I\ -\ Interest\ rate\ for\ reinvest\ the\ money$

The yield calculation formula used in this analysis incorporates the time value of money and adjusts the annual return to facilitate comparisons between different loan grades. It considers the initial investment amount, periodic cash flows, and the various loan statuses to provide accurate projections of realized yields. For detailed methodology, refer to Appendix A, and for a visual representation of expected returns, see Appendix B.

The loan grades have different yield distributions, which we analyzed using standard deviation (std) to understand the variation and associated risk. A higher STD indicates a wider range of potential outcomes and higher risk, implying more unpredictable and volatile returns and conversely. See Appendix C for a graph displaying the STD for each grade and analysis. We used box plots to visually compare returns' minimum, maximum, and median values across grades, providing insights into return variation. Additionally, we assessed loan amounts for each grade to determine the total loan scope. To understand the yield distribution for each grade, we conducted statistical tests like the Shapiro-Wilk test to check for normality. By combining these analyses, we gained a comprehensive understanding of how returns are distributed across grades. Please refer to Appendix D for graphical representations and insights derived from these analyses.

In general, we observed that riskier loan grades tend to have lower returns and higher standard deviations. It is worth noting that riskier loans exhibit lower availability in our dataset, both in terms of quantity and loan amount. This may suggest that these loans are less attractive to investors. Additionally, a majority of the loans exhibit a left tail, indicating potential downside risks from a financial standpoint.

From a business perspective, we conducted an analysis of the distribution of loan grades to address a specific problem. Our focus was to understand the percentage of loans within each grade that yielded actual returns above or below 2% (GreatYields average annual return). By examining this distribution (refer to Appendix E for the graph), we observed that for riskier loans (grades E-G), the distribution was nearly evenly split (50-50) between loans with returns above and below 2%. This finding highlights the challenge of predicting outcomes for these high-risk loans with the basic model (grade). In response to this business problem, we have undertaken efforts over the past 12 weeks to develop a predictive model aimed at enhancing our ability to forecast such cases. Detailed information on the progress and findings of these endeavors can be found in reports B-E.
<u>Potential pitfalls to consider include the following:</u>

- Ignoring macroeconomic factors.
- Over-reliance on past performance.
- Making assumptions in yield calculations.
- Limited availability of loans with respect to investment amounts, particularly for grades G and F.

For a more comprehensive and detailed examination of these potential pitfalls, please refer to Appendix F.

<u>**Question 5** - What is the risk level entailed in such investment (as measured by the volatility)? - by grade</u>
While our current dataset is exclusively focused on peer-to-peer lending, this doesn't obstruct us from conducting a valuable analysis of volatility, albeit within certain limitations. Volatility is typically ascertained by scrutinizing the oscillations in returns over time and juxtaposing them against various market benchmarks, such as stock and bond market indices. However, our dataset lacks these benchmarks and historical data, impeding the use of the traditional method for volatility analysis. In the context of peer-to-peer lending, we can utilize the variations in the loans' rate of return as a measure of volatility. Quantification of the dispersion or variation of returns through standard deviation allows us to assess the level of risk involved in an investment because a higher standard deviation indicates greater fluctuations in returns, indicating a higher degree of uncertainty and potential for losses. To enhance the precision of our analysis, we'll compute the weighted standard deviation of these returns for each loan grade. Weighted yields are used to calculate the average yield of a portfolio of loans or investments. In the context of peer-to-peer loans, the

weight represents the proportion of each loan's amount in relation to the total portfolio. Using weighted yields allows for a more accurate representation of the overall performance of the loan portfolio, taking into account the varying sizes of individual loans. It gives more weight to loans with larger amounts and less weight to loans with smaller amounts, aligning with the principle of diversification in investment portfolios. By assigning higher weight to larger portfolios, the impact of each individual investment on the overall portfolio performance is appropriately reflected, ensuring that the portfolio's risk and return characteristics are more accurately represented. This approach helps mitigate the potential negative impact of any single investment and contributes to achieving a more balanced and resilient investment strategy (For detailed methodology and analysis of the volatility graphs, please refer to Appendix G). A higher standard deviation would indicate higher volatility and, thereby, higher risk. However, it's crucial to acknowledge that this methodology provides a somewhat narrowed perspective on volatility, particularly when contrasted with conventional measures of volatility in the broader financial market. Nonetheless, our analysis still provides valuable insights into the risk and volatility inherent in investing in loans on our platform given our dataset.

<u>Assumptions:</u>

1. Weighted standard deviation is a suitable measure of volatility for each loan grade.
2. The holder bought the loan from the issue date until maturity.

<u>Potential issues:</u>

1. Lack of external benchmarks limits the assessment of relative risk.
2. Data quality and accuracy impact the validity of the analysis.
3. External factors like market conditions are not explicitly considered, forward-looking limitations as historical data may not reflect future volatility.

<u>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).</u>

Based on the analysis presented and the methodology outlined in Appendix H, It can be seen that the graph depicts the weighted annual returns for various investment amounts. It can be inferred demonstrates that GreatYields consistently surpasses the benchmark return of 2% currently achieved by the company. The graph reveals that by utilizing our two-stage model, GreatYields has the potential to achieve an average annual return of 2.5%, considering the maximum investment amount of $3,500,000 (given the data available for analysis). These findings highlight the capability of GreatYields to optimize returns. With these promising results in mind, let us now proceed to compare our model with the existing model based on loan grades.

<u>Potential Pitfalls</u>:

1. The data from 2019 on which the model is based may not reflect current market conditions.
2. Assumptions and model limitations: The analysis relies on specific assumptions and a two-stage model, which may have inherent limitations or not fully capture all relevant factors. Reports B-E detail the potential pitfalls for the models.

<u>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)?</u>

In our previous analysis, we observed that our model consistently constructs an investment portfolio with a return exceeding 2%. Now, we aim to investigate whether our model's loan selection approach surpasses the existing grade-based model. Appendix I presents a detailed comparison between the two models, outlining the respective

methodologies for each analysis. Notably, our model outperforms the grade-based model by yielding the best return for each investment amount. Furthermore, Appendix J elucidates that our model surpasses the standard approach not only in terms of potential return on investment but also by significantly reducing the weighted standard deviation (also as an alternative measure of volatility). This reduction in standard deviation is particularly crucial in the realm of investment strategy, as it provides a more secure foundation for making investment decisions. The methodology behind this analysis is thoroughly detailed in the appendix. To further establish the superiority of our model over the conventional ranking method, we utilize the Sharpe ratio as our benchmark, as described in Appendix J also. The Sharpe ratio is a widely recognized tool in the investment community that assesses the risk-adjusted return of an investment. It recognizes that higher returns over a given period may be indicative of increased volatility and risk, rather than skilled investment management. By employing the Sharpe ratio, we can effectively quantify and compare the risk and reward balance. Notably, our model consistently exhibits a higher Sharpe ratio, indicating a better risk-to-reward balance and superior risk-adjusted performance when compared to Soft Landings scoring model. For detailed calculations of the Sharpe ratio and its methodology, please refer to Appendix K. It is important to note that we used a risk-free rate of 0 in our calculations.
Potential Pitfalls:
1. Sharpe Ratio Limitations: Assumes normal returns, overlooks all risk aspects.
2. Lack of Liquidity Risk Consideration: Ignores risk related to investment liquidity.

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)?
Based on the previous responses, it is evident that the data provided is highly informative. Leveraging this data, we successfully developed a model capable of constructing an investment portfolio that consistently outperforms the grade-based model in terms of returns. Furthermore, the returns generated by our model surpass the current 2% benchmark established by the company.
Potential Pitfalls:
1. **Assumptions and Simplifications:** The model and analysis may rely on certain assumptions and simplifications that might not fully capture the complexities of real-world investment scenarios.
2. **Overfitting:** The model may have been overfit to the specific dataset, which could result in reduced performance when applied to new or unseen data.

In conclusion, our analysis of the data has provided compelling evidence that validates the platform's viability in generating returns higher than 2%. Our meticulously developed model not only surpasses the existing grade-based model in terms of performance but also consistently delivers superior returns. It is important to acknowledge the limitations we have highlighted throughout the report.
We sincerely appreciate the opportunity to collaborate on this project and contribute our expertise to its success. It has been a privilege to work with you, and we hope to have the opportunity to partner again in the future.

Best regards,
The DataDriven Portfolio Solutions Management Team

Appendices:

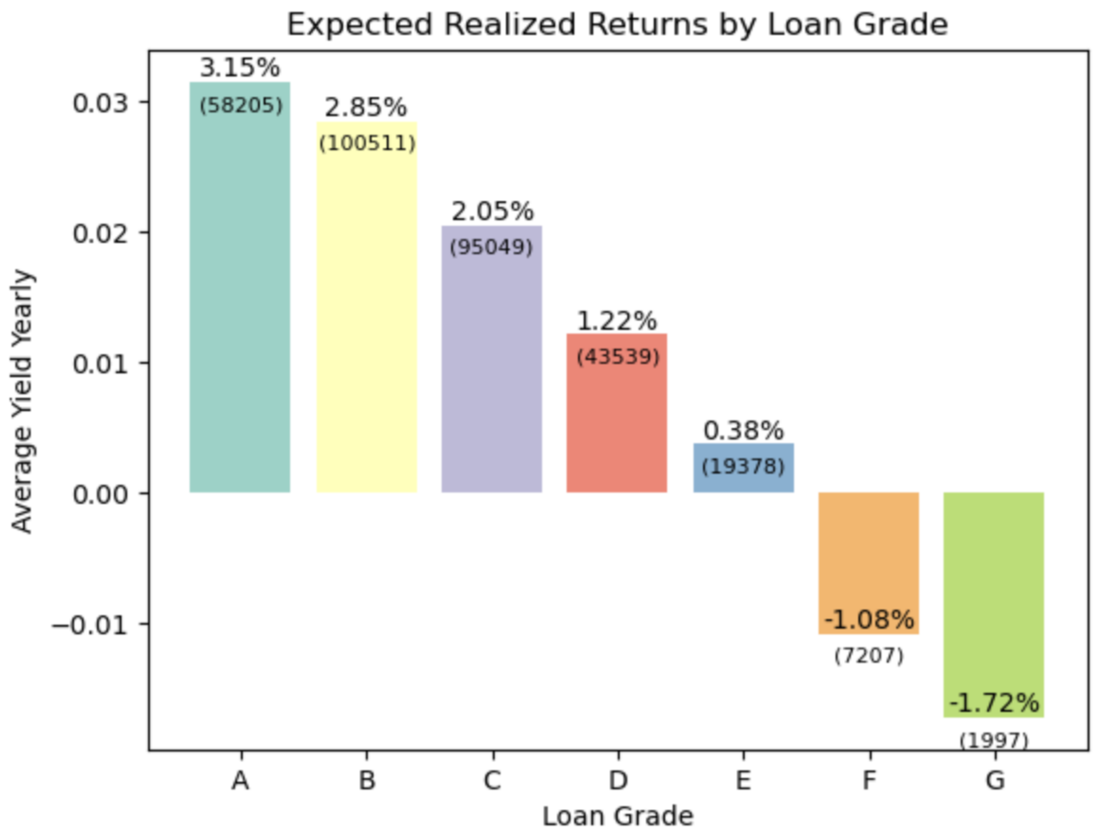
Appendix A- Methodology of the yield calculation formula:

$$E(x) = \frac{12}{T} * \frac{1}{f}\{\left[\frac{\mathrm{p}}{\mathrm{m}}\left(\frac{1\text{-}(1+i)^{\mathrm{m}}}{1-(1+i)}\right)\right] * (1+i)^{\mathrm{t\text{-}m}} - f\}$$

$T\ -\ Fixed\ time\ horizon\ in\ months$
$F\ -\ Funded\ amount$
$P - Payments\ \mathrm{re}coveris$
$\mathrm{M} - \text{Actual duration (issued date to last payment)}$
$I\ -\ Interest\ rate\ for\ reinvest\ the\ money$

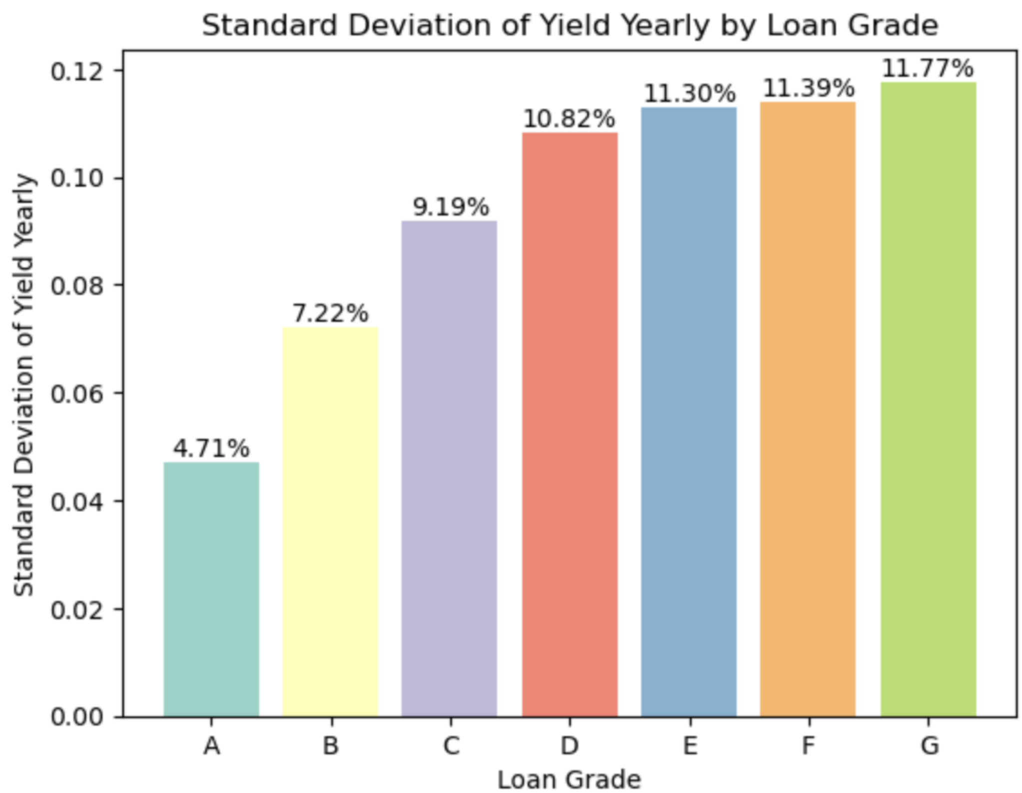
The provided yield calculation formula proves to be an effective and versatile tool for financial analysis, specifically in estimating the anticipated realized yields across different loan grades. Firstly, the formula incorporates the concept of the time value of money, considering the annual return of 2% that GreatYields can generate presently. It assumes that any cash flows received by the investment house are reinvested, acknowledging the higher value of current money compared to future money due to its earning potential. This feature allows for a more comprehensive assessment of the investment's profitability. Additionally, the formula adjusts the annual return to facilitate straightforward comparisons between investments and different loan ratings. This adjustment enables a meaningful assessment of the relative performance of investments in diverse loan grades. Furthermore, the formula takes into account not only the initial investment amount but also periodic cash flows. It provides a comprehensive assessment of the returns, considering the various loan statuses outlined in our initial report (such as prepaid, fully paid, partially paid, and non-payment, and more..). By considering the different loan statuses and the real-time duration of the loan, the formula accommodates the diverse outcomes associated with the loan portfolio. This consideration enhances the accuracy and relevance of the projected realized yields for different loan grades. Ultimately, this versatile formula embodies fundamental financial principles and serves as an invaluable tool for evaluating investment profitability. It significantly contributes to our comprehensive assessment process.

Appendix B- Expected Realized Returns:



As anticipated, there is a downward trend in the Expected Realized Return as the loan grade increases. Notably, for F and G grade loans, the yield is negative, indicating potential losses.

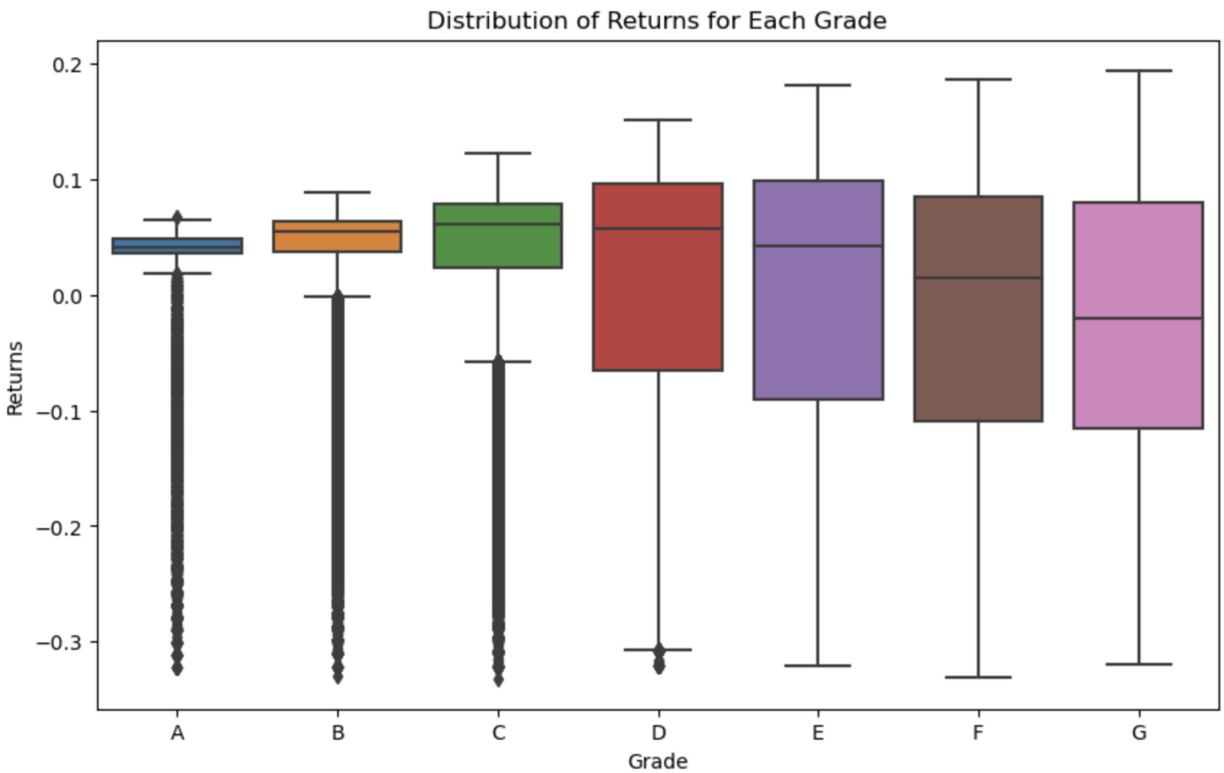
Appendix C - STD for each grade:



As expected, the std increases as the grade is higher, but it can be seen that for the higher grades (e,f,g) the relative standard deviation is close to each other.

DataDriven
PORTFOLIO SOLUTIONS

Appendix D- Another in-depth analysis of the distribution of grades:



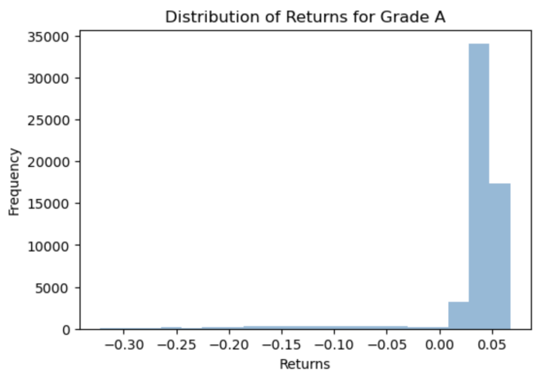
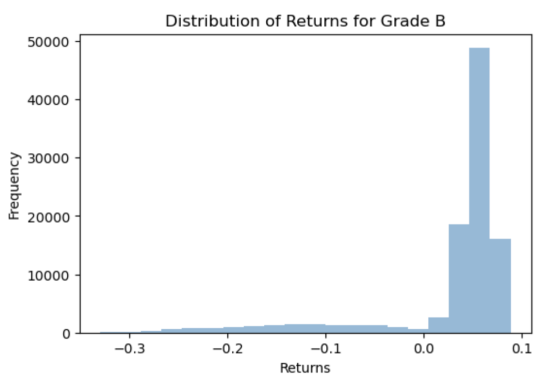
Examining the quartiles, we can see that the 25th and 75th percentiles reflect the range within which the majority of the data falls. Better-grade (A-C) loans tend to have narrower interquartile ranges, suggesting a more concentrated distribution of returns. In contrast, higher risk-grade loans (D-G) exhibit wider interquartile ranges, indicating greater dispersion of returns. Better-grade loans may provide a more predictable and potentially safer investment, with generally positive expected returns. However, higher risk-grade loans offer the potential for higher returns but with increased risk and a wider range of possible outcomes. You should carefully consider their risk tolerance and investment objectives when allocating funds across different loan grades.

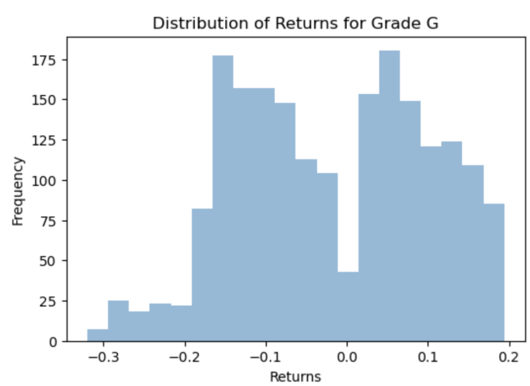
Grade	Expected Realized Returns	SD of Yield	Number of Loans	Total Loan Amount
A	3.15%	4.71%	58,205	$827,136,325
B	2.85%	7.22%	100,511	$1,297,360,550
C	2.05%	9.19%	95,049	$1,340,526,675
D	1.22%	10.82%	43,539	$674,768,850
E	0.38%	11.30%	19,378	$338,367,400
F	-1.08%	11.39%	7,207	$136,699,850
G	-1.72%	11.77%	1,997	$38,856,500

When analyzing loan grades from A to G, a clear trend of decreasing loan volume and total loan amounts becomes evident. Higher-grade loans, such as grades A, B, and C, exhibit a larger number of loans and significant loan amounts. This suggests that loans with better grades are more popular among lenders, indicating substantial loan activity. Conversely, as we move towards lower grades (grades D to G), there is a noticeable decline in both the number of loans and the total loan amounts. For example, Grade G, the lowest grade in the dataset, has a significantly smaller loan volume, with only 1,997 loans and a total loan amount of $38,856,500. The diminishing loan volumes and total loan amounts for subprime loans imply that these riskier loans are less favored by borrowers, likely due to higher levels of risk and uncertainty. Consequently, lenders and investors may approach subprime loans with caution, considering the reduced demand and increased exposure.

Overall, the analysis highlights a consistent pattern of decreasing loan volume and total loan amounts as we move from higher-grade loans to lower-grade loans, indicating shifting borrower preferences and lender risk assessments.

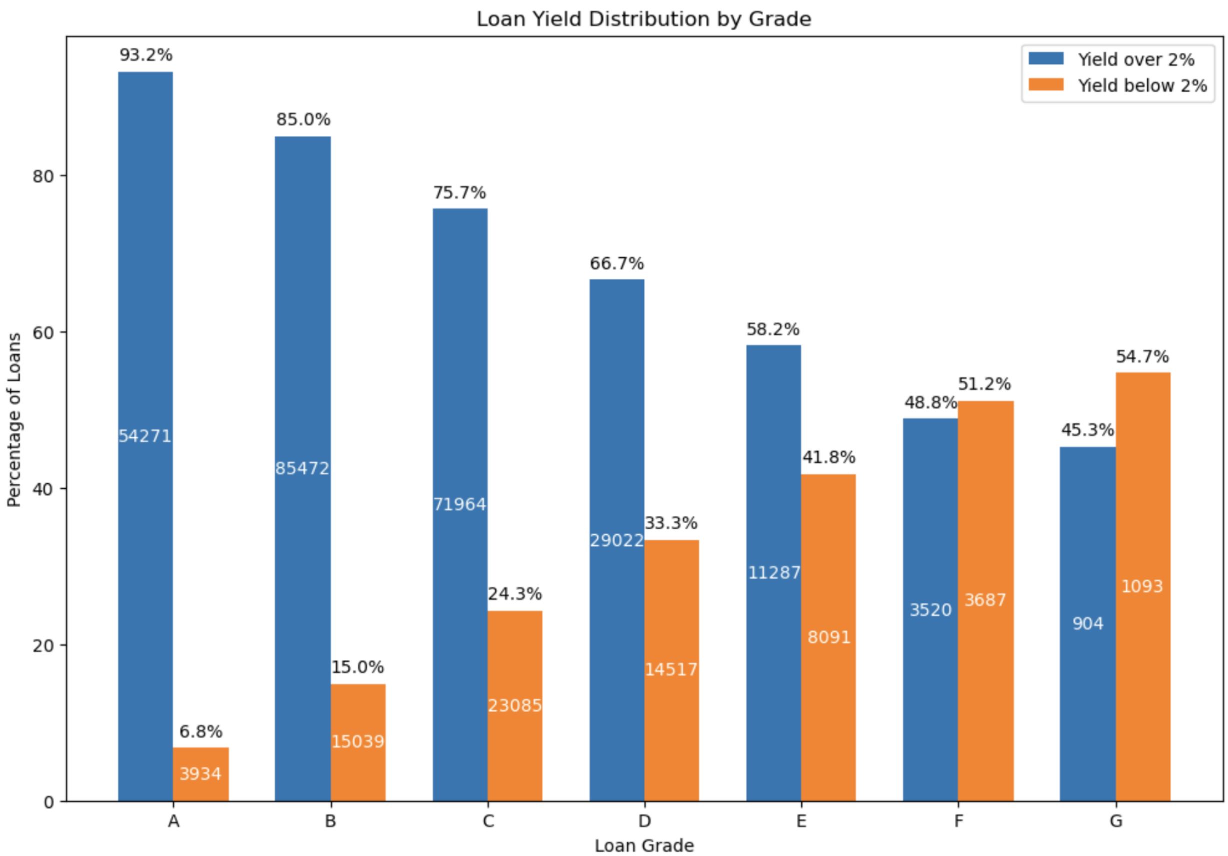
Individual analysis according to the Shapiro Wilk test:

Grade	Histogram	Shapiro- Wilk Statistic	Conclusions
A		0.393	The returns for Grade A display substantial departures from a typical distribution. The data exhibits significant non-standard characteristics.
B		0.584	The returns for Grade B exhibit significant deviations from a symmetric distribution. The data displays substantial non-symmetrical characteristics.
C		0.732	Suggests that the returns for Grade C do not follow a symmetric distribution. There is evidence of departure from a symmetric distribution in the data.

D		0.841	Indicates that the returns for Grade D deviate from a typical distribution. The data does not adhere to the assumptions of a standard distribution.
E		0.915	Suggests that the returns for Grade E do not conform to a standard distribution. The data displays noticeable deviations from the assumptions of a typical distribution.
F		0.955	Indicates that the returns for Grade F do not adhere to a standard distribution. The data shows pronounced deviations from the assumptions of a typical distribution
G		0.965	The returns for Grade G exhibit substantial deviations from a standard distribution. The data displays significant non-standard characteristics

Over all, The data provides an interesting comparison of the distribution of expected realized returns across different loan grades. When examining the yield_2 values, we observe variations in both the mean and standard deviation among the grades. Higher-grade loans, such as grades A and B, tend to exhibit higher mean returns compared to lower-grade loans. This suggests that investors may expect relatively better performance and lower risk for these loan grades.

Appendix E- Loan Distribution by Grade with Actual Returns Above/Below 2%:



The graph presents the percentage of loans for each grade categorized by whether their yield is above or below 2%. The white portion of the bars represents the count of these loans.

The analysis reveals a clear trend in loan performance based on their rating tiers. Loans with the highest ratings (A-C) predominantly yield returns above 2%. On the other hand, for loans with lower ratings (E-G), the distribution appears almost random, with an equal likelihood of yielding above or below 2%. It's worth noting that these lower-rated loans also receive higher starting interest rates from the platform.

This information provides cautious optimism that our predictive model may have the potential to improve the selection of these lower-rated loans, resulting in an increase in the average yield achievable from these portfolios compared to a random loan selection within the same rating tier.

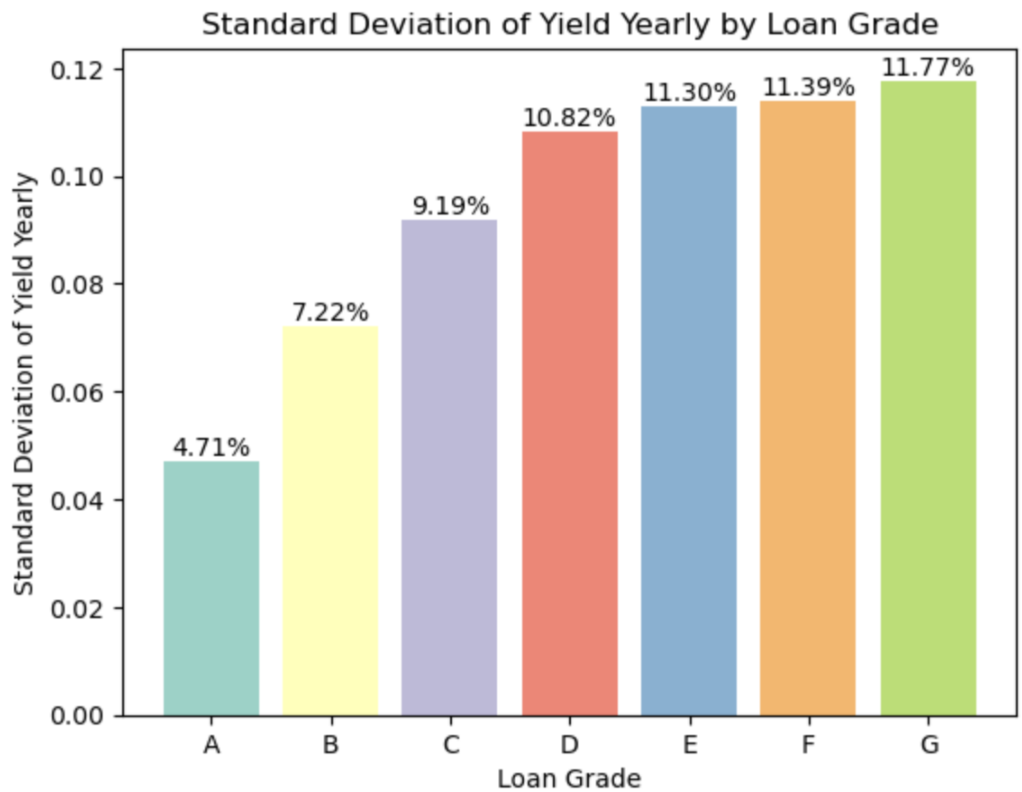
Appendix F - Potential Pitfalls Q1 :

Potential Pitfalls:

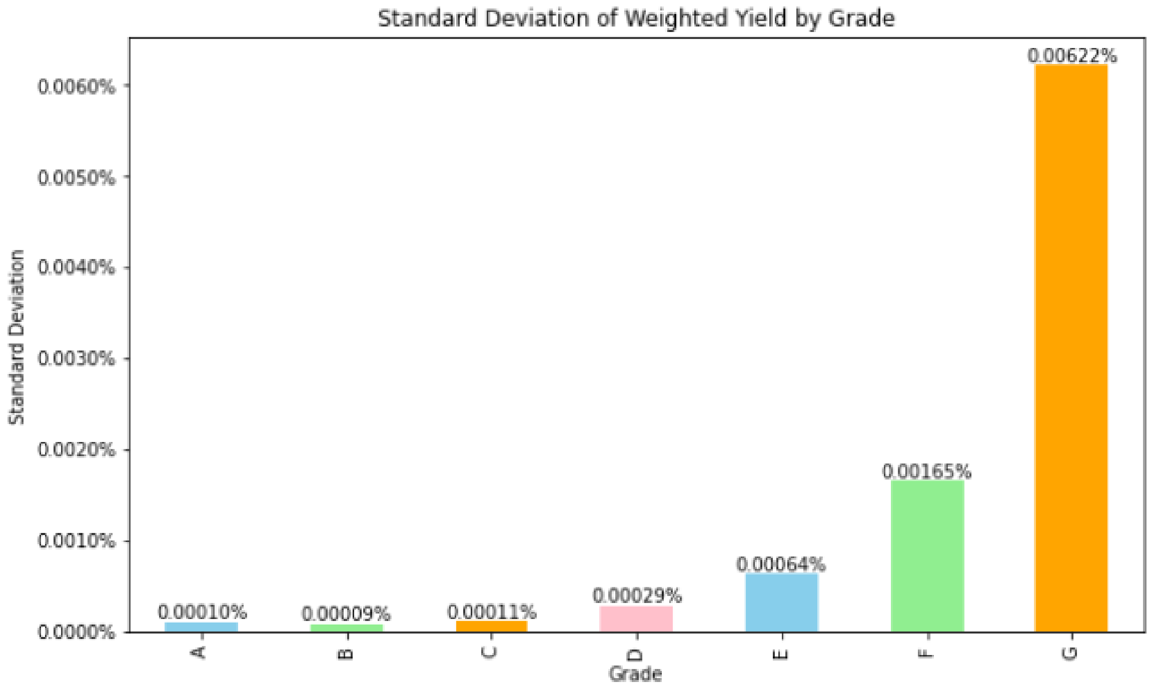
- **Assumptions in Yield Calculation:** The yield calculation formula used to estimate realized yields incorporates certain assumptions, such as reinvestment of cash flows and the 2% annual return benchmark. However, it is important to acknowledge that these assumptions may not perfectly align with real-world scenarios or investor preferences.
- **Over-Reliance on Past Performance:** It is crucial to remember that past performance may not reliably predict future results, especially considering the dynamic nature of economic conditions and borrower behavior. Relying solely on historical data may lead to inaccurate estimations of future loan performance and returns.
- **Macroeconomic Factors and External Factors:** While the analysis primarily focuses on loan grades and their associated returns, it is essential to recognize that external factors beyond the loan grades, such as macroeconomic trends, regulatory changes, or unforeseen events, can significantly influence loan performance and returns. Neglecting to account for these factors may result in incomplete or inaccurate estimations. It is important to consider and analyze broader economic conditions, including macroeconomic trends and indicators, when evaluating loan performance and estimating returns.
- **Limited Range of Riskier Loan Grades:** The analysis reveals a limited range of loans for riskier grades, especially grades G and F, in terms of the investment amount. This limitation could stem from a lower supply of such loans in the market, investor aversion to taking higher-risk loans, or incomplete lifespan of those loans. It is important to consider how this limitation affects the ability to predict yield reliably for riskier grades.
- **Risk-Return Tradeoff:** Higher-grade loans typically offer lower returns but come with less risk. It is essential to exercise caution when chasing high returns without considering the increased risk associated with lower-grade loans.
- **Market Timing and Liquidity Risks:** The analysis assumes a static investment horizon and does not account for potential market timing or liquidity risks. Fluctuations in interest rates, market conditions, or investor sentiment can impact the actual returns realized over time.
- **Risk and Return Relationship:** The analysis suggests that riskier loan grades tend to have lower returns and higher standard deviations. However, it is important to recognize that the risk and return relationship can vary over time and be influenced by market conditions, economic factors, and changes in borrower behavior. Failing to account for such dynamics could lead to inaccurate predictions and expectations of future returns.
- **Data Quality and Reliability:** The accuracy and reliability of the data used in the analysis are crucial. Ensuring that the data is clean, consistent, and free from errors or biases is important. Inaccurate or incomplete data can significantly impact the analysis and undermine the validity of the conclusions drawn.
- **Sample Bias:** The dataset used for analysis may not be fully representative of the entire population of loans or may have inherent biases. Disproportionate inclusion of certain loan grades or specific periods of time in the dataset can lead to estimations that do not accurately reflect the broader loan market.

● **Generalizability:** The findings regarding the distribution of loan grades and their returns are specific to the dataset and timeframe analyzed. It is important to assess the generalizability of the results to broader contexts or different time periods. Changes in lending practices, economic conditions, or regulatory environments could affect the performance of different loan grades and their returns.
● **Interpretation of Risk Measures:** The use of standard deviation as a measure of risk assumes a normal distribution of returns. However, in practice, returns may not follow a perfect normal distribution. As we did it is important to employ other statistical measures or techniques to capture the true risk profile of different loan grades accurately and avoid misinformed decisions.

<u>Appendix G- Methodology and Analysis of the Volatility Graphs:</u>



DataDriven
PORTFOLIO SOLUTIONS



	Grade	Standard Deviation_weighted	Standard Deviation
0	A	0.0001%	4.7114%
1	B	0.0001%	7.2184%
2	C	0.0001%	9.1868%
3	D	0.0003%	10.8245%
4	E	0.0006%	11.3032%
5	F	0.0017%	11.3900%
6	G	0.0062%	11.7708%

Methodology of the analysis-

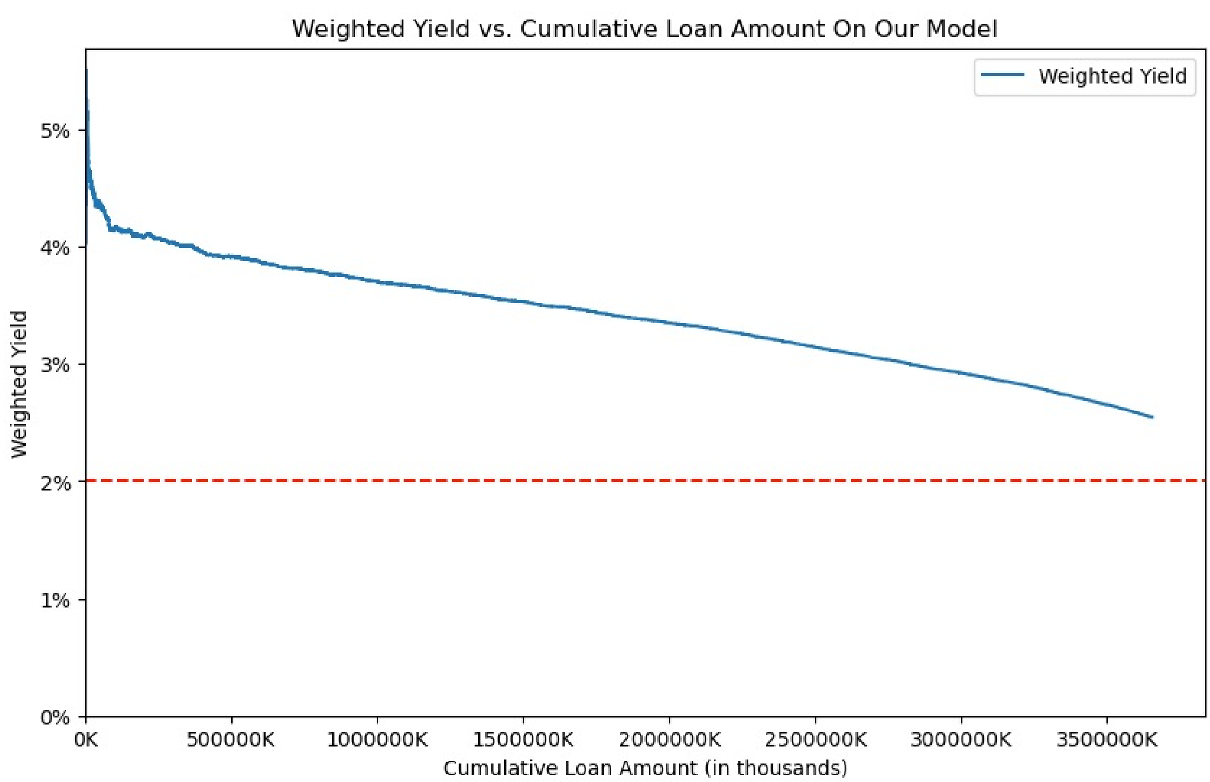
The comparison between weighted and unweighted STD provides valuable insights into the impact of weighting on the variability of yields across different loan grades. When analyzing the standard deviation values, we observe a mixed effect of weighting on yield variability.

In some cases, the application of weights to the yield calculations results in a lower standard deviation for the weighted yield compared to the unweighted yield. This indicates that applying weights helps to reduce overall variability in yields within certain loan grades (in relation to the size of the investment). The lower standard deviation of the weighted yield suggests a more consistent and stable performance within those grades, with the weighted yield values less spread out from the mean. This reduction in variability indicates a more reliable measure of performance as the weights account for the influence of individual loan sizes or other relevant factors, providing a more accurate representation of the average yield within each grade.

However, in other cases, we may observe a different trend where the standard deviation of the weighted yield is higher than that of the unweighted yield. This suggests that applying weights introduces additional variability to the yield calculations within certain grades. The higher standard deviation of the weighted yield indicates a wider range of performance outcomes for loans within those grades, as the weights emphasize the influence of certain loan characteristics or factors. This increased variability in the weighted yield values highlights the importance of considering the impact of weights on risk and performance assessment within those specific grades.

Overall, the comparison between weighted and unweighted yields underscores the importance of carefully considering the effects of weighting on yield variability. While weighting can help reduce variability and provide a more accurate measure of average performance within certain loan grades, it can also introduce additional variability in other cases. Understanding the nuances of these effects enables more informed decision-making and risk assessment, allowing for a comprehensive analysis of yield performance across different loan grades in the lending or investment context.

Appendix H - Average returns with our platform:

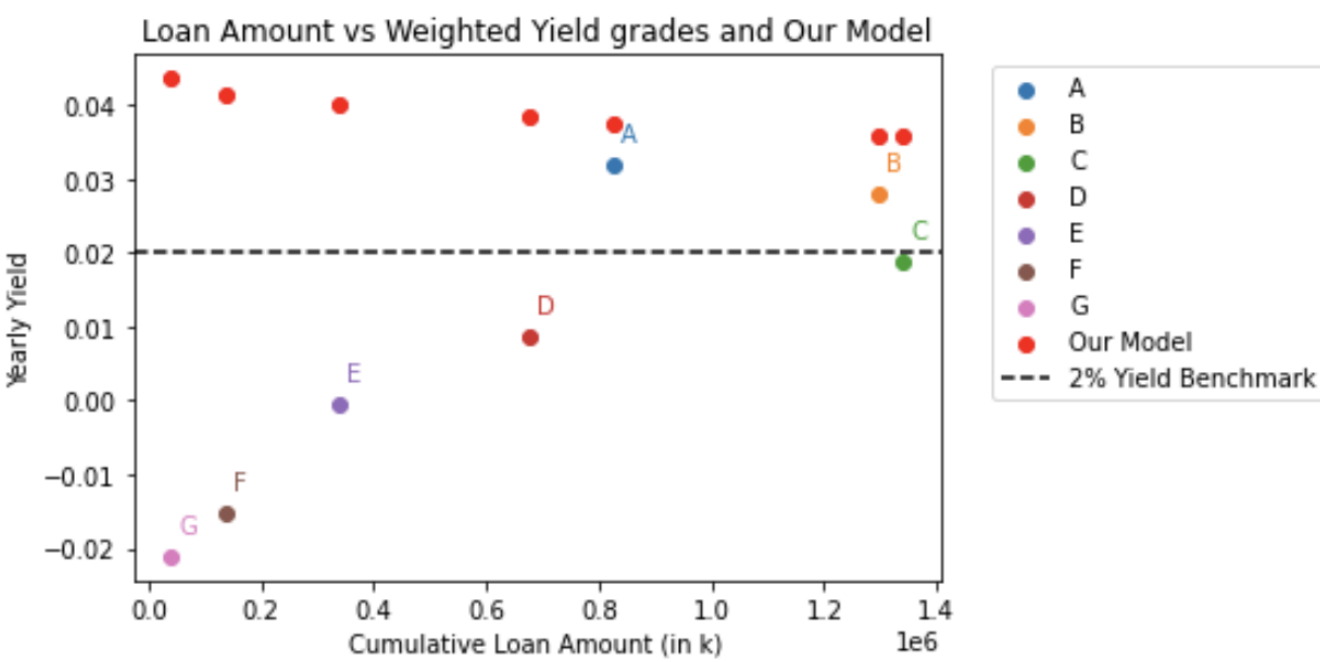


Methodology of the analysis-

To ensure a more accurate representation of the returns, we opted to calculate a weighted return instead of a simple average return in our analysis methodology. The reason for this choice is that an average return fails to consider the varying sizes of investments. Treating a small investment with a higher return the same as a larger investment with a moderate return would overlook the impact of investment size on overall returns. To address this, we employed a weighted approach by dividing the loan amount by the total amount of loans and multiplying this ratio by the previously calculated return (as described in Question 1).

This calculation method allows us to appropriately account for the size of investments, providing a more comprehensive and reliable assessment of the weighted returns across different investment amounts.
After establishing the weighted return calculation, we proceeded with the construction of the graph. Firstly, we focused on selecting loans for each investment amount, considering the yield predicted by our model (restricted to loans classified above 2% in the classification model). To ensure optimal investment choices, we sorted the loans in descending order based on their predicted annual yield. This prioritization allows us to identify the loans with the highest expected returns - as we would have chosen given the model alone.
Next, we calculated the cumulative investment amount by progressively adding the loan amounts according to the sorted order. This cumulative approach reflects the incremental growth of the investment portfolio and enables us to observe the relationship between investment amount and expected returns.
Based on these calculations, we created the graph to visually depict the impact of investment amount on expected returns. This graphical representation facilitates a clear understanding of how different investment amounts correlate with the anticipated performance of the investment portfolio.
By following this methodology, we gain valuable insights into the optimal allocation of investments based on predicted yields, enabling us to make informed decisions and maximize returns on our platform.

Appendix I - Average returns with our platform vs the platform models:



Methodology of the analysis-
The graph represents each grade on the x-axis, based on the maximum investment amount (loan amount), while the weighted return is depicted on the y-axis, with each grade represented by a different color. The red dots indicate the weighted return predicted by our model for the corresponding investment amount, Methodology for selecting the loans is the same as described in the previous appendix. Notably, our model consistently outperforms in terms of returns across all grades. However, this leads us to consider the performance of different investment amounts that are not the maximum.

Amount Invested	Our Model	A-Yearly yield	B-Yearly yield	C-Yearly yield	D-Yearly yield	E-Yearly yield	F-Yearly yield	G-Yearly yield
100000	0.062208	0.037645	0.049161	-0.003705	0.060498	-0.018977	-0.063556	-0.014030
10000000	0.044780	0.026158	0.019343	0.004912	0.019022	-0.010838	-0.020639	-0.023323
250000000	0.040588	0.031610	0.024501	0.013314	0.004744	-0.002013	-0.015169	-0.021102
500000000	0.039332	0.032130	0.025045	0.014434	0.007466	-0.000513	-0.015169	-0.021102
800000000	0.037642	0.032001	0.026914	0.015746	0.008635	-0.000513	-0.015169	-0.021102

Methodology of the analysis-

In our analysis, we expanded our investigation to include five distinct investment amounts, analyzing the weighted annual returns associated with each.

Firstly, we sorted the loans (per each grade) in descending order based on the known interest rate (int_rate) at the time of loan origination, which serves as an initial estimate of the expected return corresponding to the associated risk level (each grade).

Next, we calculated the cumulative investment amount by progressively adding the loan amounts according to the sorted order. This cumulative approach reflects the incremental growth of the investment portfolio and enables us to observe the relationship between investment amount and expected returns.

These findings are summarized in the provided table, demonstrating that our model consistently surpasses the grade-based model across all investment amounts.

In the following graph we will present the yield difference:

Diff Our Model vs A	Diff Our Model vs B	Diff Our Model vs C	Diff Our Model vs D	Diff Our Model vs E	Diff Our Model vs F	Diff Our Model vs G
0.024563	0.013047	0.065913	0.001710	0.081185	0.125764	0.076238
0.018623	0.025438	0.039868	0.025758	0.055618	0.065419	0.068103
0.008979	0.016088	0.027274	0.035844	0.042602	0.055758	0.061690
0.007202	0.014287	0.024898	0.031866	0.039845	0.054501	0.060434
0.005641	0.010728	0.021896	0.029007	0.038155	0.052811	0.058744

Appendix J - Over performance STD and yield:

loan_amount	Weighted Yearly Yield	STD Grade	grade	our_model_Yearly_Yield	our_model_STD	Increased Performance std	Increased Performance Yield	Sharpe Ratio model	Sharpe Ratio grades
8.271363e+08	0.031980	0.000032	A	0.037478	1.166709e-06	0.000031	0.005498	32122.628183	992.529528
1.297361e+09	0.028002	0.000032	B	0.035900	7.536317e-07	0.000031	0.007898	47636.281222	869.092835
1.340527e+09	0.018662	0.000032	C	0.035793	7.310777e-07	0.000031	0.017132	48959.461293	579.185399
6.747688e+08	0.008635	0.000032	D	0.038261	1.428923e-06	0.000031	0.029626	26776.153912	268.006598
3.383674e+08	-0.000513	0.000032	E	0.040078	2.984886e-06	0.000029	0.040591	13427.027892	-15.906757
1.366998e+08	-0.015169	0.000032	F	0.041284	8.368438e-06	0.000024	0.056453	4933.239262	-470.797078
3.885650e+07	-0.021102	0.000032	G	0.043556	3.222031e-05	0.000000	0.064658	1351.818608	-654.927630

The provided dataset elucidates information about investments made in varying grades of loans, namely A through G. It details the yearly yield, representing returns for both the SoftLending's grade model (considered as the baseline) and our newly proposed model. Furthermore, the standard deviation (STD), which serves as a proxy for risk associated with each type of loan, is also given for both models. For each loan grade, we calculated the performance increase of our model over the baseline model in terms of both yearly yield and standard deviation. Lastly, Sharpe ratios were computed, a widely recognized measure of risk-adjusted return, for each model.

Let's delve into the data:

Grade A loans : The total amount of investment directed towards these loans is nearly 827 million. The grade model manifests a weighted yearly yield of approximately 3.2%, while the associated standard deviation stands at 0.000032. However, our model demonstrates a superior performance with a yearly yield of around 3.75% and a considerably reduced standard deviation of 1.17e-06. Moreover, the Sharpe ratio, reflecting the risk-adjusted performance of our model, is significantly higher (32122.63) compared to the grade model's ratio (992.53)

Grade B loans : The invested amount in this category is approximately 1.29 billion. Despite the larger investment, the grade model's yearly yield descends to 2.8%. On the contrary, our model enhances this yield to nearly 3.59%, all the while with a minimized standard deviation (7.53e-07). Similar to Grade A loans, the Sharpe ratio for our model (47636.28) surpasses the grade model's (869.09).

Grade C loans : An investment of roughly 1.34 billion is seen in these loans, with the grade model yield further dropping to 1.87%. Despite the decline, our model succeeds in pushing the yield up to 3.58%. The Sharpe ratio continues to reveal our model's superior performance.

Grade D, E, F, G loans : As we proceed towards the lower grades, both the invested amount and the grade model's yield decrease, the latter even becoming negative for grades E, F, and G. Interestingly, our model not only maintains a positive yield but also augments it, despite the grade model's declining performance. Coupled with a reduction in standard deviation, our model's Sharpe ratio significantly outweighs the grade model's, even when the latter becomes negative due to unfavorable yields.

In summary, the data demonstrates that across all loan grades, our model consistently outshines the grade model, both in terms of yield and risk (standard deviation). As the grade model's yield decreases (even turning negative), our model successfully maintains and even enhances the yield while also reducing the associated risk. The consistently higher Sharpe ratios of our model across all grades signify better risk-adjusted returns, suggesting that our model offers more stable performance. This makes our model a .more effective and reliable tool for investment decision-making

Appendix K - Calculations of sharpe ratio:

Assumption: Risk free rate = 0.

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

where:

R_p = return of portfolio

R_f = risk-free rate

σ_p = standard deviation of the portfolio's excess return