PORTFOLIO PERFORMANCE ANALYSIS EXERCISE FOR RISK & ANALYTICS

Yufei Chen

Abstract

This report analyzes credit risk analytics and formulates loss rate expectations for a set of loan products. By analyzing historical loan-level data, I evaluated the risk profiles of three hypothetical loan products, conducted analysis, made forecasts, and provided recommendations on how to utilize this data to enhance risk management.

Overview

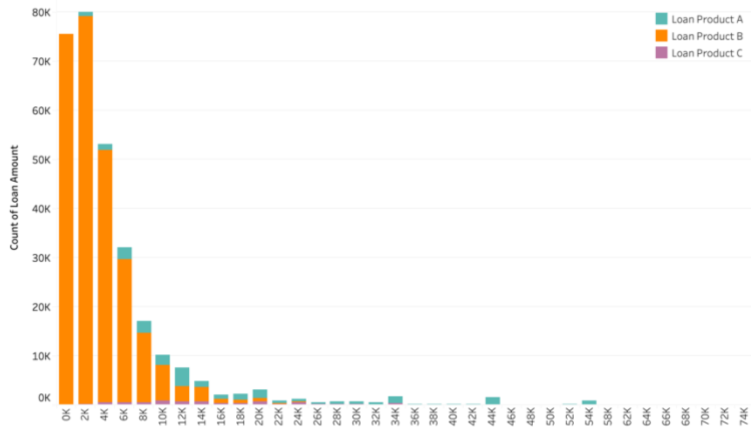
The datasets provided information on three different loan products, loan products A, B, and C, each with unique structures and characteristics. After a thorough inspection of the variables, I was able to identify key insights that can inform risk management decisions.

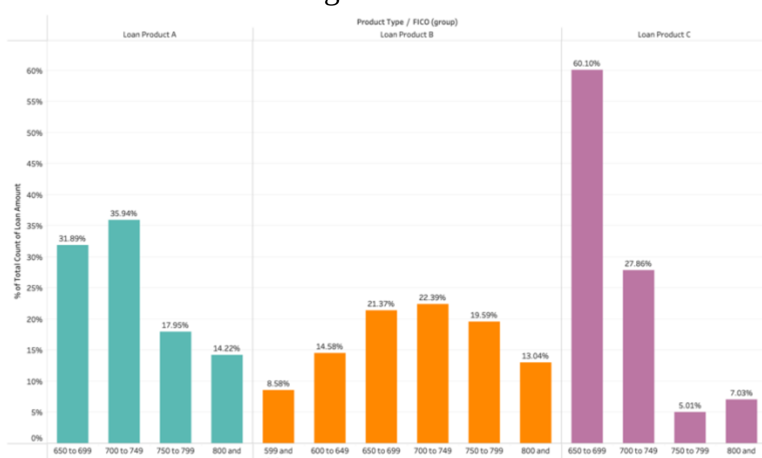
The chart below summarizes the size, composition, and key risk characteristics of the different loan products.

	Product A	Product B	Product C
Portfolio Stratification			
Total Loan Amount	$512,748,514	$1,106,815,600	$111,653,300
Number of Loans	25,645	265,470	7,149
Average Loan Amount	$19,994	$4,169	$15,618
Weighted Average FICO	743	724	707
Average Current Loan Balance	$4,538	$3,498	$7,665
Weighted Average Loan Term (months)	152	33	46
Prepayment Rate	-	24%	51.3%
FICO Distribution			
599 and below	0	8.58%	0
600 to 649	0	14.58%	0
650 to 699	31.89%	21.37%	60.10%
700 to 749	35.94%	22.39%	27.86%
750 to 799	17.95%	19.59%	5.01%
800 and greater	14.22%	13.04%	7.03%

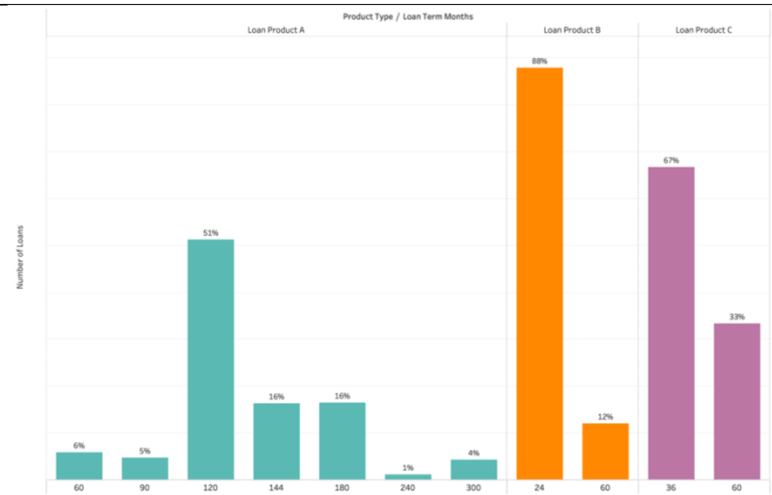
Origination Characteristic

This section focuses on visualizing the origination characteristics of the loan products, which provides a comprehensive overview of the credit mix.

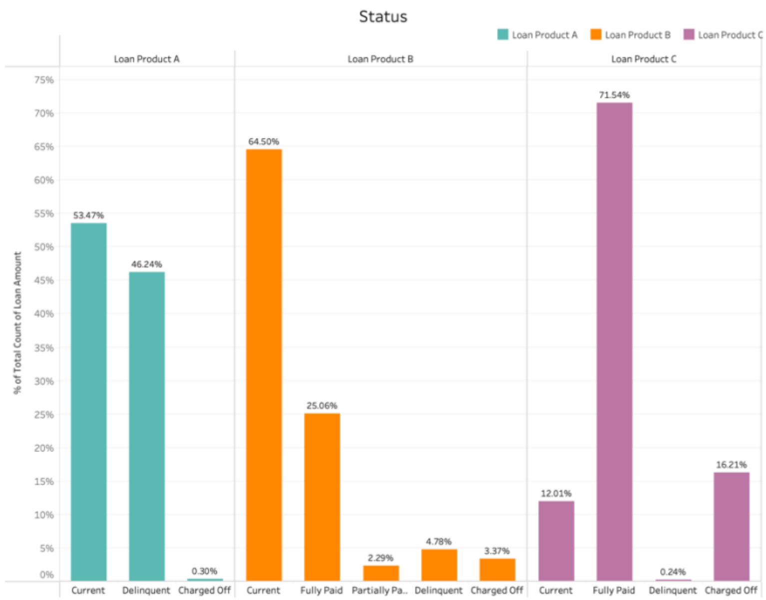
<table>
<tr><td>Size</td><td>Overall, the size of the loans in the pool is skewed to the right, indicating that the majority of loans have a smaller size, with values under $8,000. Additionally, we observed that loan Product B has the highest number of loans within the pool.


</td></tr>
<tr><td>FICO Distribution</td><td>The chart below shows the FICO distribution across the three loan products. Loan Product A displays a relatively even distribution of loan FICO scores which are all above 650. In contrast, Product B exhibits a bell curve FICO distribution ranging from under 599 to above 800, with the highest degree of variation among the three products. Finally, Loan Product C FICO scores are above 650, with a concentration in the 650 to 699 range.


</td></tr>
<tr><td>Term Distribution</td><td>The chart below displays the distribution of loan terms among the three loan products. Loan Product A has a greater variety of loan terms compared to Products B and C, which only offer two loan terms. Furthermore, Products B and C have a higher proportion of loans with shorter loan terms.</td></tr>
</table>

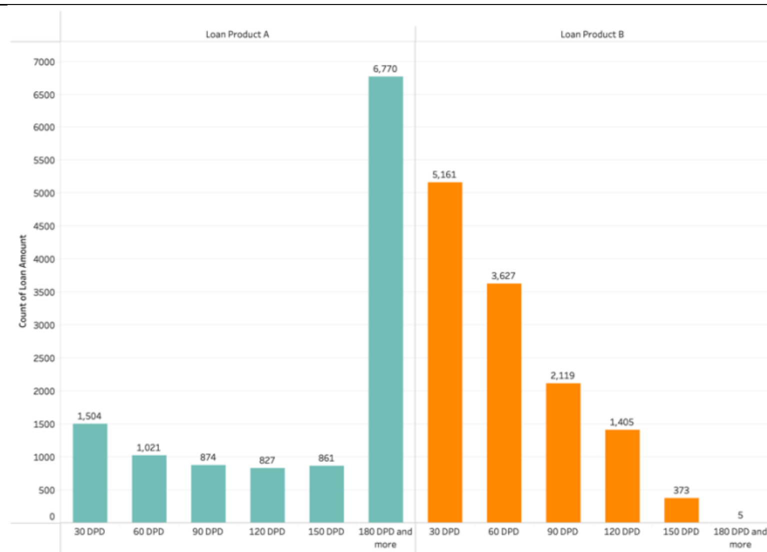
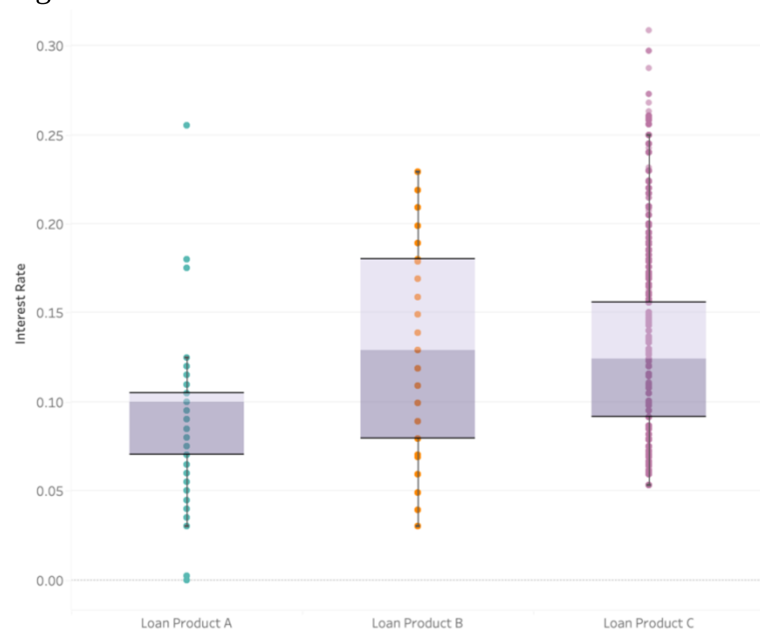
<table>
<tr>
<td></td>
<td>


</td>
</tr>
<tr>
<td>Status Distribution</td>
<td>
The chart below illustrates that most of the loan products are either current or paid in full. However, Loan Product A has the highest delinquency rate among all products, indicating a higher rate of late or missed payments. In contrast, Loan Product C has the highest charged-off rate, suggesting a greater frequency of loans that are unlikely to be repaid.



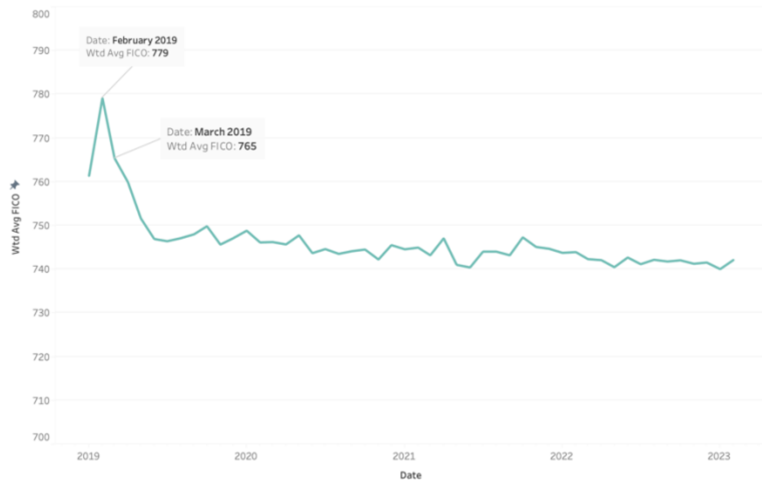
</td>
</tr>
<tr>
<td>Delinquency Distribution</td>
<td>The following chart depicts the distribution of Days Past Due (DPD) among the three loan products. Loan Product A has a higher concentration of loans with higher DPD levels, indicating a greater frequency of missed payments. In contrast, Product B has a higher proportion of loans with lower DPD levels, suggesting a lower incidence of delinquency.[1]</td>
</tr>
</table>

[1] Loan Product C does have some loans with the status of "Delinquent," however, these loans also have a charged-off date, which disqualifies them from being categorized as "Delinquent" in this visualization.

<table>
<tr><td></td><td>
</td></tr>
<tr><td>Interest Rate Distribution</td><td>The chart below displays the interest rate across the three loan products. Product B has the highest median interest rate, while Product A has the lowest median interest rate. Notably, Product A is the only loan product that offers a zero-interest rate option. Furthermore, it's worth noting that Product C has a larger number of outliers compared to the other products, particularly at the higher end, compared to the other products. Given that it also has the highest charged-off rate among the products, there may be a relationship worth exploring further in the upcoming section.

</td></tr>
<tr><td>Credit development</td><td>The following graph presents the weighted average credit score development over time for the three loan products. It reveals that Product B's weighted average credit score has been decreasing over time, whereas Product A and C experienced</td></tr>
</table>

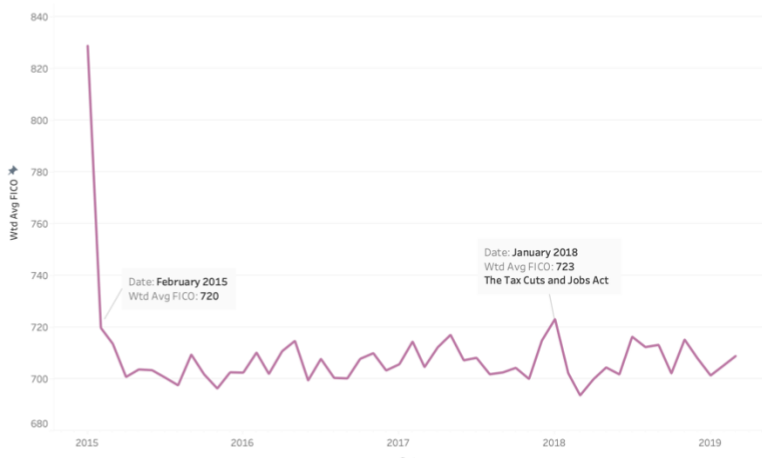
a significant drop in the early period since the origination date and have remained relatively stable over time with slight mean-reversed fluctuations.

Product A



Product B

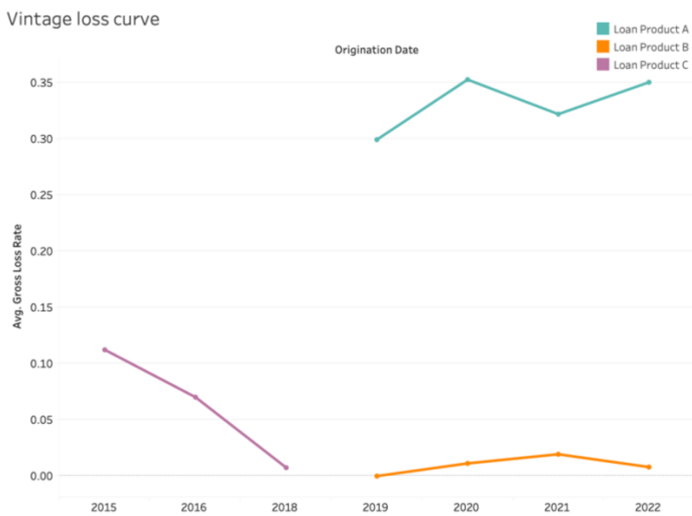
Product C



	Product A and Product C began with a high credit scores soon after the loans originated. However, as more loans were issued, the average credit score decreased and has since remained relatively stable with minor fluctuations. On the other hand, the credit score of Product B has been decreasing since July 2021. It's important to note that credit scoring models are continuously evolving to improve their ability to predict credit risk. In 2021-2022, credit scoring models may incorporate additional data sources, such as rent and utility payments, to provide a more comprehensive picture of a borrower's creditworthiness, which might cause the score to decrease over time.

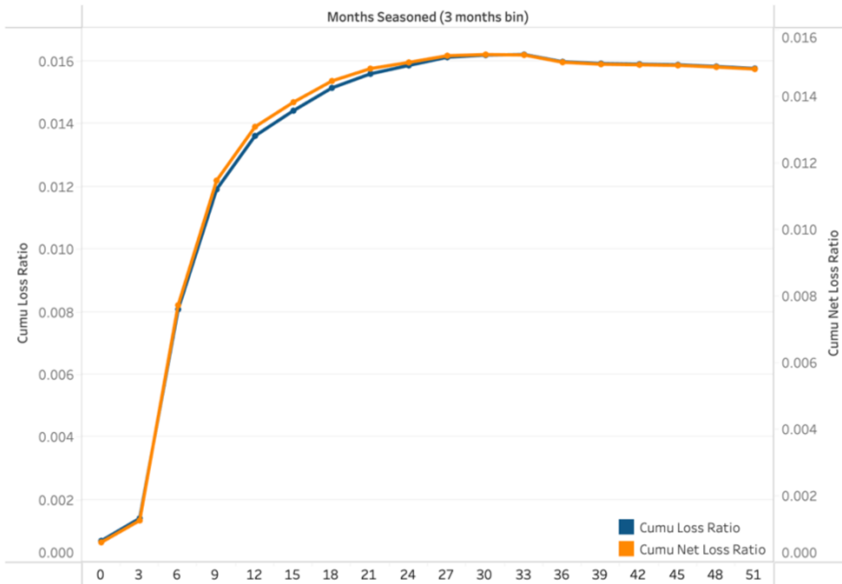
Portfolio Risk Analysis

To begin with, I computed the charge-off rate against the Origination Date and generated a Vintage loss curve to visualize the loan products' historical loss performance. The graph below shows that the overall loss rate for both Product A and B has increased, while Product C has demonstrated a continuous decrease in its loss ratio.

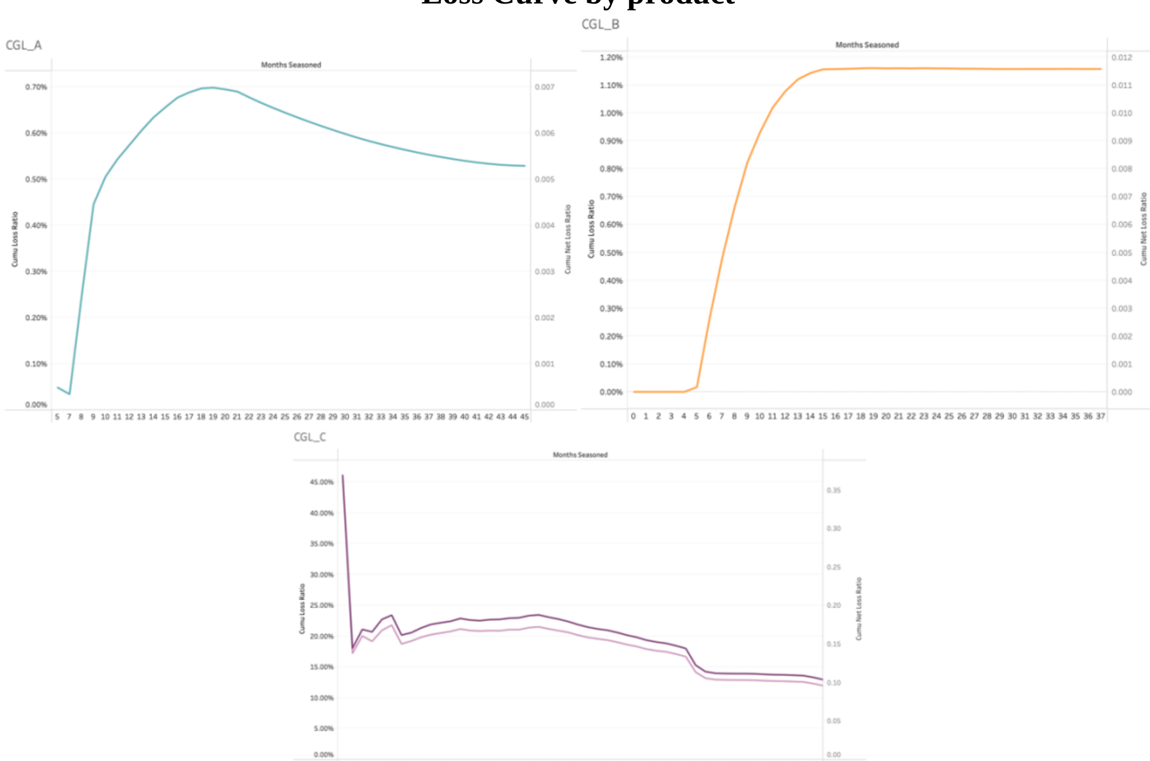


Next, I created a plot of the cumulative charge-off rate against Months Seasoned, which is a newly created variable representing the number of months that have passed since the loan was originated. This was done to gain insight into the possibility of default after loan issuance and to observe how the loss rate changes over time. The resulting curve indicates that the likelihood and magnitude of recovery after a write-off are low, as the cumulative gross loss and cumulative net loss are relatively close. It is also important to note that there is no recovery data available for Products A and B, and so for the purposes of this analysis, we assume that there is no recovery for these two products. The chart below shows that default rates drastically increased three months after loan origination until month 21, at which point they started stabilizing and eventually reaching a plateau. This indicates that loans are at the highest risk of default or non-repayment 3 months after origination and that effective risk management strategies should be put in place to minimize losses during this period.

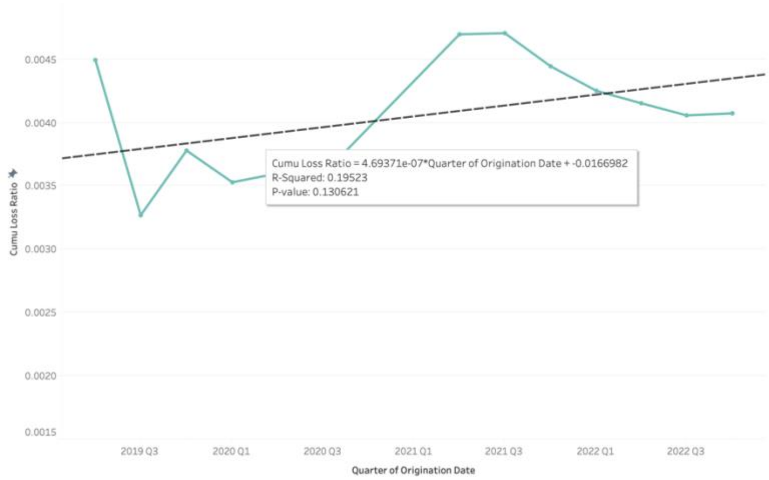
Total Products Mix

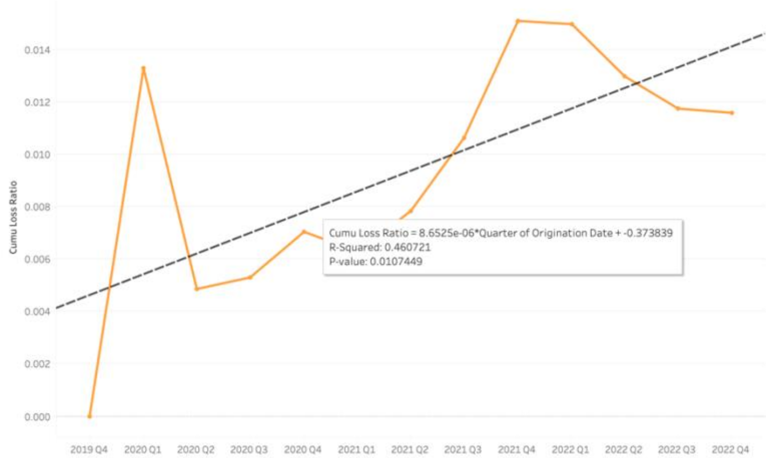


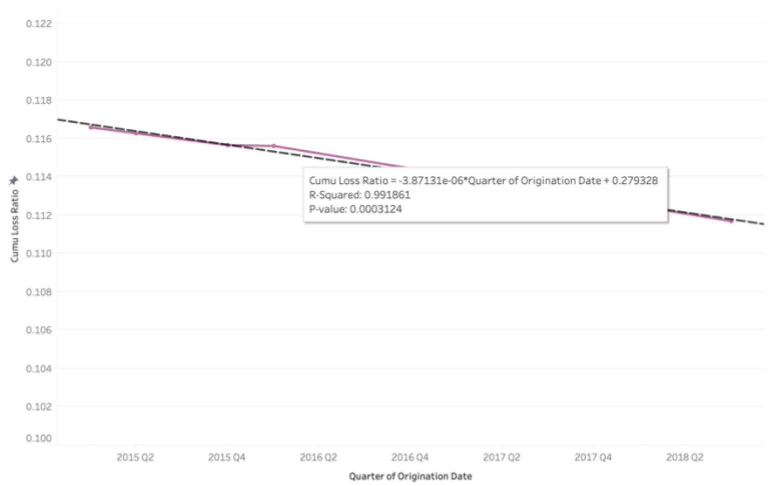
Loss Curve by product



When examining the cumulative loss ratio development over time for each product, we can observe a clear trend in each product. So, by employing linear regression estimation, we can forecast the estimated cumulative loss ratio for each product in the future.





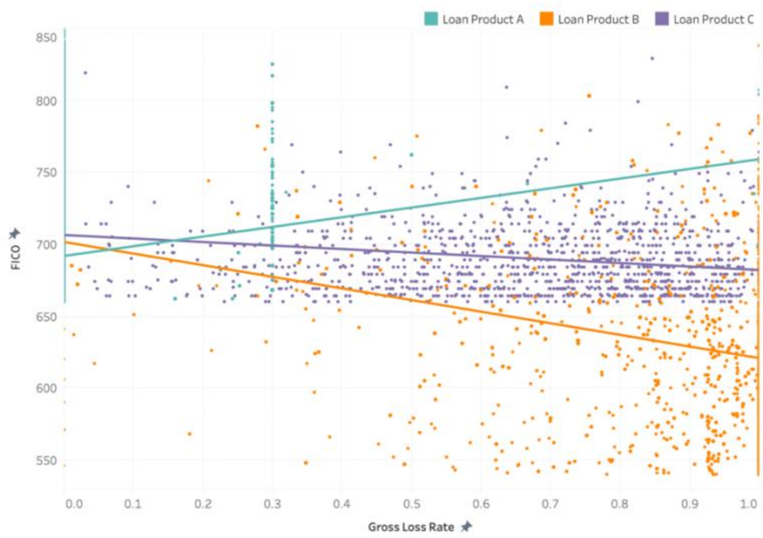


Variables Breakdown

After calculating the correlations between all variables, we found that Interest Rate, Loan Term Months, FICO, and Months Seasoned have the highest correlation with Loss Rate for each of the three products under investigation.

Correlation	Loss Rate		
	Product A	Product B	Product C
Interest Rate	0.05	0.10	0.28
Loan Term Months	-0.03	-0.03	0.18
FICO	0.17	-0.14	-0.16
Months Seasoned	-0.12	0.08	-0.30

For Product A, Loan Term Months and Months Seasoned were found to have a negative correlation with Loss Rate, while Interest Rate and FICO were positively correlated with Loss Rate. However, it's important to note that Product A had a more concentrated distribution of FICO scores, which may impact the accuracy of the correlation with FICO (as showed below).



For Product B, Loan Term Months and FICO had a negative correlation with Loss Rate, while Interest Rate and Months Seasoned had a positive correlation with Loss Rate.

For Product C, Loan Term Months and Interest Rate had a negative correlation with Loss Rate, while FICO and Months Seasoned had a positive correlation with Loss Rate.

Based on the given dataset, I conducted a preliminary calculation to estimate the rate of return:

	Net Loss	Interest Income[2]	Rate of Return
Product A	2,088,384	368,021,034	99.43%
Product B	12,761,324	731,383,029	98.26%
Product C	10,799,755	339,386,642	96.82%

[2] Interest Income is calculated by Interest Rate * Months Seasoned * Loan Amount, under the condition that the loan is current or fully paid.

Conclusion

Firstly, from the above plot of cumulative charge-off rate against Months Seasoned, we know that the risk of default rises after 3 months and keeps rising until month 21. Therefore, I recommend implementing strict risk monitoring during this window.

Secondly, Product A is the highest performing product among the three loan products, with the highest rate of return, highest weighted average FICO score, and slower cumulative loss rate increase. Therefore, the company should allocate more resources to Product A to further improve its performance.

Thirdly, Product B has a rapidly increasing cumulative loss rate, even though its weighted average FICO score is not the lowest. However, it contain the lowest FICO scores in the pool, and the weighted average FICO score has decreased over the years. To address the issue, the company should conduct a more detailed analysis of the underlying factors contributing to the high loss rate within Product B. Furthermore, the company should implement more stringent underwriting standards for new loans within Product B and consider scaling back its size while reallocating resources to Product A.

Lastly, Product C has the lowest rate of return and the highest charged-off rate among the three products despite showing a decreasing trend. Additionally, Product C has fewer accounts with a FICO score over 700, although its median interest rate is comparable to Product B. Since all Product C loans were issued before October 2018 and are unlikely to issue new loans, the company should focus on managing the risks associated with outstanding loans. One potential strategy could be offering incentives to borrowers within Product C who maintain a strong payment history and FICO score. Moreover, the company should conduct a more detailed analysis of the borrower profile for Product C, given the highly fluctuating FICO score, even though FICO score distribution is more concentrated and there is less number of loans in the first place.

Limitation

One major limitation of this analysis is the presence of missing data and insufficient variables in the existing datasets. Advanced tools such as machine learning would require additional factors, including information on the loanee (such as the purpose of the loan and household income), economic indicators, and market data, to generate more accurate analyses and forecasts. Also, the regression model used for the analysis indicates that Product A has an R-square below 20% and Product C has an R-square over 99%, both of which are not the perfect result. Given more data and more time, the company could consider incorporating more detailed analysis with machine learning tools and adding more relevant factors to improve the accuracy of the results.