1. What would you say drives the variance between actual and predicted repayment?

See the python notebook for the results.

As data is not loan level (per consumer), there is limitation on the analysis that can be performed as predicted and actual repayment are loan level(per consumer) data points. What's provided, is the merchant level performance data.

I did some analysis on the data provided.

- *Calculated the total $ over prediction and %.*
- *Separated the merchant level data into over prediction and under prediction. Ran descriptive statistics(mean, median, std dev, quartiles) on the separated data to find any obvious trends or data errors. Didn't find any obvious issues.*
- *As data is merchant level, I cannot comment or see trends for loan level.*

The factors that drive the variance between actual and predicted repayment:

- *The model being used by XYZ is over predicting the repayment. Overprediction is when model overpredicts repayment, but actual repayment is lesser. Thus, losses/bad rate is more than expected.*
- *Change in consumer's credit profile over time like*
    - *Change in FICO*
    - *Change in debt-to-income ratio*
    - *Unexpected events like sudden expenses, job loss or medical emergencies*
- *Government influence like*
    - *Stimulus payments*
    - *Covid relief programmes*
    - *Pause in student loan payment, etc*
- *Credit Score Inflation – Artificial increase in credit score(like FICO). Results in situations where consumer are getting better scores than they should actually have. This is a major reason for models overpredicting repayment. Thus, XYZ ends up with a higher bad rate/losses than expected.*
- *External Economic factors: Shift in macro-economic trends due to global factors, geopolitical concerns, etc.*
- *Shortcomings with the Quantitative model being used to predict repayment rates. Lenders' risk assessment models may not accurately capture all relevant factors, leading to prediction errors.*
- *Incorrect/Misleading sales: Borrowers are sold incorrect credit products. Products or Loans they cannot service, loan terms and financial obligations they don't understand, thus impacting repayment behavior.*

2. What could drive the difference between auth_amount and loan_amount?

auth_amount: total amount of user requests / apply for
loan_amount: total amount of the loan

- *Ability to pay the amount customer apply for.*
- *Credit Score/Profile of the customer. E.g. FICO Score*
- *Debt to income ratio*
- *Employment Status*
- *Total Income*
- *Rent/Mortgage payment*
- *Underwriting criteria of the lender*
- *Risk Appetite Limits/Framework of lender*
- *Interest rate and Fees*
- *Loan Term*

3. Based on the data, in which areas would you increase or decrease volume?

See the python notebook for the results.

- *Based on data, I have calculated Interest income, bad loss and pnl for each merchant. Sorted them by PNL.*
- *Obviously, volume should be increased with the merchants where we see the maximum profit. I have pulled the top 10 profitable merchants.* ***Volume should be increased at those merchants.***
- *For these top 10 merchants, Although the repayment rates for them are not in the high 90% and credit profile of these consumers is near-prime/prime, the risk is adequately compensated by high APR which translates into high interest income.* ***Volume should be increased at those merchants.***
- *One of the reason that can be attributable to high losses, is that certain merchants have 0% APR. I see that's where the main profit bleeding is taking place. Also, another reason is some merchants have very low average APR(14-16%).* ***Volume should definitely be reduced at such merchants.***

4. Which categories generate the most profit for Company XYZ?

See the python notebook for the results.

Categories that generate most profit:

- *Women's Fashion – Women's clothing*
- *Musical Instrument and Supplies*
- *Home Furnishing - Furniture Stores*
- *Shoes Stores*
- *Personal Service – Dentist office*
- *Health and Personal Care store*
- *Jewelry*

5. What information can Company XYZ gather to further evaluate their merchants' profit?

- Company XYZ can gather loan level/consumer level data to accurately evaluate merchant profits.
- Consumer/Loan level data can further give us insight into what is driving the PNL. Is it consumer's credit profile or merchant related attributes
- Merchant level data can be further broken down geographically to assess what regions in the country are driving higher profits
- Any data relating to additional fees and charges other than the interest income will also help with accurate estimation of PNL
- XYZ can also gather information about the exact product that is being purchased on the installment loan. This can further drive insights about demand for that product/merchant and if any new customer acquisition/marketing strategies can be used to further drive profits.

6. What types of analysis, evaluation, or diligence should Company XYZ do?

Company XYZ should look for the following:

- Portfolio trends over time and where do they seem the most opportunity, demand and profit
- Build the portfolio PNL and take business decisions based on the profit numbers
- Fraud analysis to check if no fraudulent transactions are occurring
- Analysis for gaming behavior – As I see a lot of the merchants are offering 0% APR loans, we should be checking if gamers are not taking advantage of the offers. Repeated gamers should be banned from the underwriting process
- Roll rate analysis – To check if bad accounts could have been made current if timely action was done. Bucket 1,2,3 is the most important time to work with the customer and reduce bad losses.
- Reject Inference – For consumers that were rejected from availing credit, Company XYZ can conduct reject inference to see how the consumer performed. Business can gain valuable customers and increase profit if it sees that good consumer are being rejected.
- Repeat purchases – Once the loan level data is available, Company XYZ can check if consumers are coming back on the APP/Website/Merchant and availing the credit from XYZ again. Increase in repeat purchases can be attributes to good customer experience, good UI, ease of checkout, etc.
- Marketing analysis – Loan level and product level data can help marketing teams to identify potential consumer groups for targeting. E.g. targeted advertisement to Amazon/Walmart consumers, targeted advertisement to consumers buying iPhones.