QR Homework

You can find several csv's uploaded at this link:

https://gist.github.com/alekdot/e0e653bc5b1a0dd2ea0599eefae384b3

Data Dictionary:

- category / Subcategory: Merchant's industry
- name: name of the merchant
- merchant_id: Unique identifier for the merchant
- actual_repayment_pct: actual percentage of loan volume that is repaid
- predicted_repayment_pct: predicted percentage of loan volume that is repaid
- num_trxn: number of loans per merchant
- auth_amt: total amount of user requests / apply for
- loan_amt: total amount of the loan
- fico: score that measures a user's risk, higher score means less risk (range from 300-850)
- term: Duration of the loan
- apr: annual percentage rate (interest rate that charged to the user)

Feel free to use any method of analysis you choose: eg. Python, SQL, R, SAS, Excel and present it in slides, paper, etc.

Premise: You are consulting for Company XYZ which has provided you with a performance summary with some of their loan performance summarized by merchants.

Please answer the following questions:

1. What would you say drives the variance between actual and predicted repayment?
2. What could drive the difference between auth_amount and loan_amount?
3. Based on the data, in which areas would you increase or decrease volume?
4. Which categories generate the most profit for Company XYZ?
5. What information can Company XYZ gather to further evaluate their merchants' profit?
6. What types of analysis, evaluation, or diligence should Company XYZ do?