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

There are factors that could impact the variance in predicted vs actual repayment. Such factors include:

i) Model overfitting and underfitting

Underfitting and overfitting are common causes of model variance. High variance is mainly due to the algorithm being more sensitive to the training dataset, whereas for low variance, the model is less sensitive to the specific dataset used during training. If bias-variance trade-off is not carefully implemented, this could lead to a less reliable model.

ii) Model deterioration/drift

Due to constant changes in economic conditions, such as inflationary or recessionary pressures, consumer behavior could change to adapt to such changes in the economy. If the models are not tested and updated regularly, this could lead to significant differences between predicted and actual output.

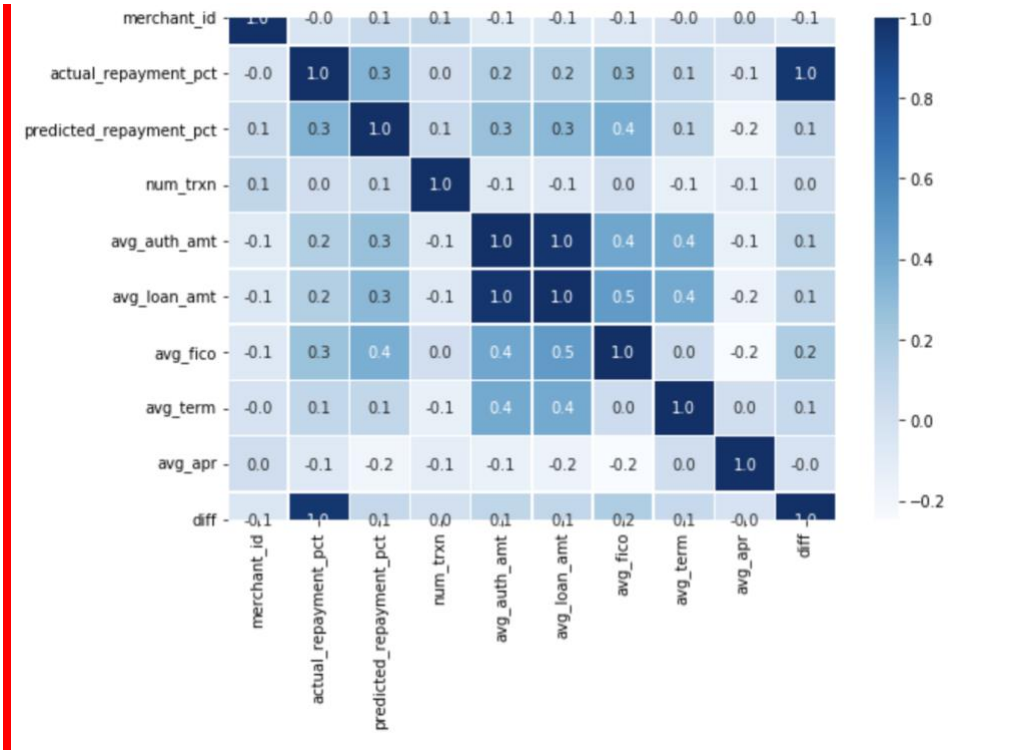
iii) Data limitations

As with any other machine learning algorithms, credit risk models utilize data in training and validation. Factors such as lack of data, poor quality data or biased datasets could negatively impact the model's accuracy, leading to huge differences in observed vs predicted output.

This dataset reveals an average repayment difference of 2% (absolute value), which could be attributed to any of the above reasons. It is worth mentioning that the difference in expected vs actual repayment could also be a normal, as it's always expected that the model would not perfectly mimic the real-world conditions.

2) **What could drive the difference between auth_amount and loan_amount?**

Based on classical consumer behavior, it is expected that one would ask for a loan amount that is higher than they need. Moreover, higher risk customers are more likely to apply for higher loan amounts. This is clearly demonstrated in the dataset, as there is a slightly stronger correlation between the merchant's fico score and the amount approved for vs the amount they applied for (0.5 vs 0.4).



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

Looking at the distribution of volume by category, it's apparent that the top three categories ('Other', 'Jewelleries' and 'Home Furnishings') make up about 80% of the overall volume. When broken down by subcategory, their top four subcategories make up about 60% of the volume, meaning that about 50% of the overall demand comes from the only four subcategories. This could potentially be a great opportunity for the business to invest in and focus more in these areas.

Likewise, the volume distribution by category shows the lowest two categories (Men's Fashion and Beauty) each have less than 1% volume. A further look into their subcategories reveals there are only two subcategories, each with a 50% distribution. Since there are only two subcategories, the business could reduce any sales and marketing efforts in these two categories. Alternatively, it may be worth research and analysis to learn more about why these subcategories have low volume.

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

Categories with the higher volume are likely to be more profitable than those with less volume. However, the data shows that a combination of higher volume and longer loan term would result in the highest profits. The results indicate that of the top three afore-motioned categories, merchants who average term was greater than the mean (9.24) have a 3% higher repayment rate than those whose term was less than 9.24.

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

- Merchant tenure: Merchants with a longer tenure, combined with good payment history, could be targeted with specific offers such as lower interest rates and higher loan amounts as they are less risky and thus more revenue likely to be earned from them. Similarly, low tenure merchants could pose a slightly higher risk, hence the company could be more restrictive with the loan offers.
- Average Rate of Return (ARR): Increasing ARR indicates growing business, hence an opportunity to target such merchants with more loan offers and special customer service treatment.
- Merchant's costs: Understanding the merchants' costs would help to better evaluate their profitability as well as the level of risk exposure. For instance, higher variable costs could eat into the revenues earned, and over long periods of time, this could lead to slow growth and inability to repay loans.

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

- Diligence in modelling (how well you predict) and customer evaluation. This involves constantly testing and updating the models to make sure their output reflects the current market and consumer behaviors.
- Experimentation (such as A/B testing) for some of the variables, to try and optimize revenue and profits. The company could also launch test and learn (pilot) programs that run for a relatively short period as a way to test the market response to new products and services.