Merchant Risk Score Monitoring Tool

Peter Kaiser

CEU (Central European University) Business Analytics MSc

2022 May

Capstone Project Summary

Introduction

I made a risk score monitoring tool for my client for my capstone project. The client is a financial services provider, who provides Point of Sale terminals to merchants. They wanted to calculate their operational risk. For them, the operational risk meant how much money they would lose if a specific merchant went out of the business or terminated its contract. This of course can be meaningful for multiple time intervals: daily, weekly, monthly, quarterly, half-yearly and yearly. They wanted the risk for each of their merchants and the risk of the portfolio, which would be the sum of the individual merchants. The report was to be in power BI, where the riskiest merchants and riskiest merchant categories were shown. By the end of the project, I delivered a report, which had all the usable data they provided in a master table. It also included some charts and tables which I thought were important. There could be improvements as there were data issues and the timeframe or the project was short. But they were happy with the product as it has shown them what they wanted to see, and something which is even more important: the shortcomings of their data.

Calculation

They wanted the risk to be calculated as the product of three factors: the sum of every transaction (transaction amount) which went through every terminal of the merchant, a risk category assigned by the company, and a credit score from an outside source. The transaction amount took the absolute value of the transactions, so the ones where the merchant paid for example refund, would be still added up. The risk category assigned was a factor variable which could take three values: low, medium, high. I mapped numerical values to these categories. These were in the small percent ranges as these values correspond to the amount the company gets after each transaction. The last factor, the credit score was sourced from an outside source. It was binned into 4 bins. This is the variable that should tell us how likely the company will go out of business and based on that it will get a value between 0.5 and 2. It is more likely that we lose merchants whose credit score is higher, than those for which the

credit score is lower. We can see that the factor which will impact the risk most is the transaction amount. This is because the other two factors are bounded, while the transaction amount is not. This comes from the fact what they wanted the risk be: amount of money lost if the merchant goes out of business, which corresponds to the merchants total transaction amount.

Outcomes

There are four levels of hierarchy in the merchant ecosystem: terminal level, merchant level, merchant category subcategory, and merchant category. At the start of my work there was a file which had the terminal level and merchant category level transaction amount. This is an Excel Cube file, it has an underlying data and we can have different pivot tables on it. The obvious problem is that there is no merchant level data. It can be solved with a map that has each merchant with their terminals. The other files were: a file with the required calculations, and a file with the merchant category subcategories and their risk categories. I made a first prototype on the merchant category level data. I calculated a merchant category level average risk category from the merchant category subcategory level risk values. I then joined the transactional data, in Excel with LOOKUP functions to the risk categories. Four weeks later I got a merchant to terminal level map, and a merchant onboarding data set. There were three problems with these. The terminal level transaction data had 10% of the total transactional amount not attributed to terminals, the merchant to terminal map was missing 10% of the merchants and the onboarding data which was supposed to have the merchant risk category and the credit score had 2% completion rate. Later I learned that this is a still ongoing process and getting the credit scores from the outside source is time-consuming and costly. This meant for the final report, I could not use the credit scores, so we just removed that factor from the calculation. Also, I kept using the average risk values for merchant categories. I read in all the data with power BI, joined the tables and made tables and graphs for two power BI sheets.

Improvements

There are three categories of improvement which can be made: data improvements, automatization, and report improvements. The first and most obvious one is improving the data. There are two ways to improve the Cube transactional data set. One would be to add merchant level transactional data, the other would be to map all the transactions to a terminal. The first would be obviously better, as then they would not need the merchant terminal map.

The second improvement for data would be to improve the merchant terminal map to include all merchants. The third would be to fill the on-boarding data so it has risk categories and credit scores for all merchants. The second improvement category would be to automate the pipeline. As for now power BI scans a file path to read in the data. It reads in the new data, but for example the Cube data has format changes when there is a new month. That would be solvable if they would use the long data format. The last improvement I suggest is using more data in the tool. Maybe the age of the merchant or their location. They could even create more data, for example ratios of refunds and payments or how a merchant compares to the average risk in the merchant category.

Benefits

For the company, the benefits were that they had a report which shows them the risky merchants. This is a first kind of a report for them. This report was done for a country level data, and they are in multiple countries, they could use this for more countries or for all. The second key factor they mentioned is that they now know how fragmented their data is, and that there is a lot of missing data. As we could see 10% of the transactions were not mapped to a terminal, the merchant terminal map was missing 10% of the data, and the onboarding data set was missing 98% of the data. This report was important for them, but they did not have any people with the right skillset to make this for them.

What I learned

The skills I developed were mostly from using power BI and I also refreshed my Excel skills. I communicated with my project manager and learned how a mid-sized company operates as well. I already knew Tableau but learning power BI fast was still a challenge. There were a couple of interesting functions and information I learned about the software. For instance, power BI scans a file path, unlike Tableau, where you can save the data with the report. I learned about this when I finished the project, I deleted everything from my notebook, their report went blank, and I had to rewrite the read in the new file paths. The next important lesson I learned was that if one read a text Excel or read a numerical cell as numerical then converted it to text, for the same input the results would not be the same. Meaning the table keys need to have the same format in the Excel file. The non-technical aspect was interesting too, my contact was kind and helpful, but sometimes our requests were still slowly processed. I think this is because sometimes a mid-sized company just does not have the needed resources to do it all.