



Risk and Analytics Team – Dataset Exercise for Candidates

Introduction

Thank you for taking the time to interview with us so far — we're excited about your interest in a role on our team and would like to invite you back for another discussion. In advance of our next meeting, we'd like to ask you to look at a raw dataset and perform a small analytical project so that we can see how you work. We'll ask you to make a small, informal presentation of your work at our next meeting.

Before diving into the project, please go ahead and schedule your next interview with our recruiter. We prefer that you set a time approximately one week after receiving these instructions, but the recruiter will work with you one-on-one to arrange a time that is convenient for everyone

Scenario

Spendy is an innovative new payment service that has just launched. It has partnered with top online shops throughout Japan to provide pay later credit to the shops customers

During the checkout stage, all customers now have the option to use Spendy as a payment method and is only required to fill out a phone number and email address to use the service.

When the customer uses Spendy, Spendy sends 97.5% of the purchase amount to the shop and keeps 2.5% as a payment facilitation fee.

Spendy has 2 ways for the customer to settle their debt the next month:

- At the convenience store in cash, which attracts a fee of 200 yen collected by Spendy
- Through bank transfer, which attracts no fees

In addition to the phone number and email, the shop also sends Spendy some extra information detailed below:

Field	Description
amount	Amount of the purchase
browser_type	Type of web browser the customer is using
has_points_card	Do they have a points card for the shop?
high_risk_item	Is the item considered high risk?
merchant_customer_age	Number of months customer has had an account with the shop (7 = greater than 6 months)
item_category	What type of goods is the customer buying
purchase_time	What time of day did the customer purchase the goods
age_range	Customers age range

For the first 20,000 customers, spendy had an approval rate of 100% in order to collect data as much repayment data as possible.

Of the customers that repaid the debt, 80% did so by convenience store.

For customers that missed the payment due date, Spendy decided to call and remind them of their obligation to pay. On average this call cost Spendy 300 yen per customer called.

A flag to indicate default has been added to the data for these 20,000 customers.

Field	Description
flag_default	1 if the customer missed the due date by more than 90 days else 0

Your task

Using the first 20,000 customers (Historical Data sheet), build a limit policy that can be applied in the future to set maximum purchase amounts for new customers

Your target should be to maximise the profitability for Spendy by accepting or rejecting purchases based on the information provided.

Apply this policy to the next 10,000 new customers (New Data sheet)

- How does your model perform?
- How many of the 10,000 customers should we approve?
- How many of these customers do you expect would default on their repayment?
- What would be the expected profit/loss for spendy from these purchases?
- How would you approach this exercise if the business wanted to optimise growth?

Notes

- While candidates are expected to train a machine learning model, there is no need to spend too much time building something perfectly hypertuned and production-worthy – but do please be prepared to discuss any potential shortcomings and what additional steps you would take if this was for a real business application
- Building good models is only one part of managing risk – understanding how to apply the results of the model to various business scenarios in order to optimize for economics other factors is also key.
 - What key information would need to be known before devising a strategy on top of this model? What business strategies would the candidate think are advisable given certain key assumptions?
- It is important that you provide explanations of your work (i.e. why you choose to do things with the data set that you do). You are welcome to present verbally during the interview time, but your output should be a complete representation of your analysis
- It would be great if you could store any code you produce in a place that's easy to review (Github or BitBucket are preferred, but sending files directly is OK too)
- You can choose how to present your results. Please let us know what you'd need to present your findings (laptop with Powerpoint, R, Python notebook, etc)
- If you have any questions whatsoever, please feel free to reach out to us directly or via the recruiter!