Data Analysis Task

Basic terminology:

- Fraud - the use of a stolen credit card to purchase online
- A chargeback (CHB) - a charge that is returned to a payment card after a customer successfully disputes an item on their account transactions report

Riskified's business model:

We collect a small fraction of the amount of all approved orders and give a chargeback guarantee for the full order amount if an order we approve comes back as a chargeback.

Riskified's decision-engine:

All orders are given a score by the model ("classification score")

For each merchant, we define a threshold between 0 to 1 for each model.

- When the classification score is above the threshold, the model's action will be "approve" and order will be auto-approved
- When the classification score is below the threshold, the model's action will be "decline" and the order will be auto-declined

The following dataset consists of 40,825 orders, classifications for these orders by a single model, and a few of the data points we collect for each order.

Columns in the dataset:

- order_id: a unique identifier for an order
- order_submitted_at: the date in which the order was submitted to Riskified
- order_status : the actual status of an order (in this dataset we included only approvals and chargebacks)
- price: order total amount
- digital_product: whether the product is digital or tangible
- customer_account_age: the interval between the date the account was created and the order was placed
- order_source: the type of device used to place the order (web/mobile device app/phone/etc.)
- billing_zip: the zip code filled in the billing details
- shipping_name_length: shipping name number of characters

- classification_score: the score given by the model

<u>Task</u>

1. Rely exclusively on model scores, and set a decline threshold that will provide a 90% approval rate
2. Plot the model scores distribution
3. Assuming we aim at a proportion of 50% between the CHB cost and the total revenue (sum amount of CHBs divided by the total revenue), what would have to be the fee?