Payment Matching

At Shopee, bank transfer is a payment method in most countries. When a buyer chooses to place an order using bank transfer, he/she is supposed to make the transfer within 2 days after he/she places the order.

After he/she makes the transfer, Shopee will receive a bank statement from the bank and Shopee needs to compare and match the bank statement with the checkout information in order to confirm that this particular order has been paid. This process is called payment matching.

Two criteria need to be met in order to match a bank statement with a checkout:

- Amount match: Statement amount equals checkout amount.
- Name match: Statement description "matches" checkout buyer name (Note: statement description usually contains buyer name)

A proper match occurs when both the amount and the name matches on both bank statement and checkout list.

Task

Based on the two criteria, please match the Bank Statement IDs to the Checkout IDs. Please note that in the dataset provided, there are unique matches of each bank statement and checkout.

Example

Please find below a few examples of matches:

Statement description	Statement amount	Buyer name	Checkout amount	Name match logic
SUK UNAISI ANMIS V* QOXXEQ QEYIXX * HCQOXXEQ QEYIXX - TRANSFER TO A/C	10.40	Suk Unaisi	10.40	Statement description "contains" buyer's name.
BVALY JIEN HAL * QEHHW * CKQEHHW -	32.84	Bvaly Hal	32.84	Statement description does not have to contain every word in buyer's name.
SHOPEE FROE FAI MIA ROBJ	14.52	Froe Mia Robj Fai	14.52	Order of the words in names does not matter.
ROHANA MOMAME* Rohana.com.my	9.00	Rohana Momamed	9.00	The spelling of the words in names are not exactly the same, but very similar.

Dataset Description

bank_statement.csv:

- stmt_id: Bank statement ID
- stmt_amount: Bank statement amount
- description: Statement description received from bank. It is supposed to be matched to checkout buyer name.

checkout.csv:

- ckt_id: Checkout ID
- ckt_amount: Checkout amount
- buyer_name: Buyer name that buyer fills in during checkout. It is supposed to be matched to bank statement description.

Evaluation

Submissions are scored based on Categorisation Accuracy:

Evaluation Description

Your submission will be evaluated based on categorisation correctness. The score is the number of correct results that your submitted file contains.

$$score = \frac{1}{N}\sum_{i=0}^{N}p(x_{i}, y_{i})$$

Where<p>

- N is number of test samples.

- x_{i} is the predicted category for i th test sample.

- y_{i} is the ground truth for i th test sample.

- $p(x_{i}, y_{i})$ is calculated as 1 if $x_{i} = y_{i}$ and 0 otherwise.

Submission Format

Two columns required:

- stmt_id
- ckt_id

Example:

stmt_id	ckt_id
1	955
2	10027

Your submission should have **240,000 rows** (excluding the headers) , each with **2 columns**.