Data Science – Problem 1

Customer Matching

# Problem Statement:

Customer has two different datasets one with the customer level data and the other based on the aggregated transaction level data. The issue here is they have not been able to directly map both the datasets since customer data has the customer ID but misses the customer Name and the transactions data has the customer Name but misses the customer ID.

In order for the Customer to identify revenue across their customer base, could you match data between two datasets.

# Requirement:

Can you provide a solution to perform the data triangulation between the two datasets? (The datasets have been attached)

We have the following data in both the datasets. But the format of the data might be different.

* Geography
  + 4 fields: ANZ, Asia, Europe, NA
* Time period
  + 4 field: last 12 months, last 24 months, last 36 months, lifetime
  + There’s not date provided, but the last recorded transaction looks to be 2/9/2021, so I assume the reference periods are anchored around that date
* Revenue components
  + 3 fields: transaction count, volume (turnover), revenue
* Start date, end date (the header label might be different)

**The data points for each customer that you have to compare are**

* Start date
* End date
* Geography
* L12M total transaction count, principal, revenue (Use L12M spot transaction + L12M forward transaction)
* L24M total transaction count, principal, revenue
* L36M total transaction count, principal, revenue

Your solution should make sure data anomalies are handled. You can use Fuzzy matching or any other approach.

Also, please provide the percentage match between the two datasets.