

Challenge: To perform a ‘fuzzy’ lookup against a lookup table to extract additional information

Situation: In this example the source record could contain one of the values for Entity_Name (see table below) and the requirement is match this against the Entity_Name column in the target table to obtain the Entity_Type.

SOURCE	TARGET		Desired result (value assigned to source record)
Entity_Name	Entity_Name	Entity_Type	Entity_Type
L D'OTTAVIO NURSERIES LLC	L D'OTTAVIO NURSERIES, L.L.C.	B	B
LAWRENCE D'OTTAVIO JR	D'OTTAVIO LAWRENCE JR	P	P
2 FAIRFIELD CRESCENT L L C	2 FAIRFIELD CRESCENT, L.L.C.	B	B
SOMERSET HILLS Y M C A	SOMERSET HILLS Y.M.C.A.	B	B
PETER D THOMPSON	THOMPSON PETER D.	P	P
PETER D THOMPSON	THOMPSON PETER D.	P	P
WILLIAM H PETTIT JR	PETTIT WILLIAM H. JR	P	P
JEFFREY B REICHLE	REICHLE JEFFREY B.	P	P

B: Business

P: Person

The Entity_Name column could have multi token strings with individual tokens transposed. For example, ‘PETER D THOMPSON’ may exist as ‘THOMPSON PETER D.’ in the lookup table. The requirement is to be able to ‘fuzzy’ match against the lookup table and establish that these records satisfy the ‘lookup’ condition. Based on this the Entity_Type column is to be assigned to the record from the source.

Solution: Since the match has to account for multi-token strings with possible transposition of tokens it is recommended to use the Identity Match Option (IMO). If the customer is not licensed for IMO it is possible to use Bigram Matching too. The key here is to get match outputs in ‘Matched Pairs’ instead of ‘Clusters’. The screenshot below shows where option can be selected.

Properties

General

Ports

Match Type

Strategies

Match Output

Advanced

Data Viewer

Tags

Progress

Match output type:

Match Output Type	
Clusters	
☒ Matched Pairs	
Best Match	

Once the matched pairs are obtained it is necessary to weed out to the poor matches by retaining the matches with the highest scores.