Exercise

To link or not to link?

Similar to joins, record linkage is the act of linking data from different sources regarding the same entity. But unlike joins, record linkage does not require exact matches between different pairs of data, and instead can find close matches using string similarity. This is why record linkage is effective when there are no common unique keys between the data sources you can rely upon when linking data sources such as a unique identifier.

In this exercise, you will classify each card whether it is a traditional join problem, or a record linkage one.

Instructions 100XP

- Classify each card into a problem that requires record linkage or regular joins.

Take Hint (-30 XP)

Incorrect

Hmm are you sure about that? One of the other columns does not contain a unique identifier per column, and since names and addresses are strings you may want to opt for

Drag the items into the correct bucket

Drop items here

Record linkage	Regular joins
Using an `address` column to join two DataFrames, with the address in each DataFrame being formatted slightly differently.	Consolidating two DataFrames containing details on DataCamp courses, with each DataCamp course having its own unique identifier.
Two customer DataFrames containing names and address, one with a unique identifier per customer, one without.	Two basketball DataFrames with a common unique identifier per game.
Merging two basketball DataFrames, with columns `team_A`, `team_B`, and `time` and differently formatted team names between each DataFrame.	

Submit Answer