PERCHWELL

Assignment Overview

One of the most important components of Perchwell's business is the marriage of real estate datasets from varying and often unrelated sources. One example is the association of apartment listings for sale with public tax records (i.e. "closings") of sold apartments and homes. Achieving correct association of these record types ensures accurate representation of the real estate market for our users.

Specifications

The attached datasets, `listings.csv` and `closings.csv`, include sample listing and closing data from two sources with different conventions for describing apartment numbers. Using the values of the fields provided, your goal is to write a program to correctly match as many closings to listings as possible. Note that not every record in `listings.csv` has a matching record in `closings.csv` (and vice versa). Additionally, since multiple apartments can be sold at once, one listing may correspond to multiple closings and one closing may correspond to multiple listings.

The following must be true of all associations made between records in `listings.csv` and `closings.csv`:

- Values of the `list_date` field for a listing must precede the value of the `sale_date` field of an associated closing.
- `apartment_number` values (however they may be expressed) describe the same apartment. See `apartment_number_examples.csv` for possible variations on apartment numbers. Feel free to account for as many of the examples as you see fit along with cases that are not included.
- The difference between the value of the `list_price` field and the `close_price` field should be no greater than 40% of the value of the `list_price` field.
- The `building_id` of the listing and the closing must be the same.

Your program should be able to generate a file with the same columns as `matched_listings.csv` containing all matched and unmatched records. Each row should correspond to either an unmatched listing, an unmatched closing, or the combination of a matched listing and closing.

You may implement your solution in whatever programming language you're most comfortable working. And while you're certainly free to use third-party dependencies in the process of completing the assignment, we're more interested in reading and understanding your own approach.

Completion

The code for your completed assignment should be sent to recruiting@perchwell.com as a compressed attachment. It should include:

1. The output CSV file `matched_listings.csv`.
2. Executable code along with instructions for running your program and any other commentary you feel is necessary. We should be able to generate the same `matched_listings.csv` file as the one you provided.

Evaluative Criteria

We're looking for self-documenting, idiomatic code that demonstrates an ability to understand and manipulate datasets. While complete matching of listings and closings is neither expected nor possible, results are important: a strong solution will achieve a high match rate while meeting the stated criteria.

Contact Information
Ellie Nave
Human Resources
ellie@perchwell.com
(914) 282-0141