Background: One of the harder problems we work on is joining records from data sources that do not have a join key. We refer to this process as conflation.

Instructions:

Write a program that takes the two .csv files of Points of Interest (POI) data as inputs and outputs a third .csv of matches in the form of **id1, id2** where id1 is the id from Provider 1 and id2 is the id of Provider 2. A match is when a POI from Provider 1 and a POI from Provider 2 are the same POI.

While you are expected to complete the making of a match list, it is not expected and is likely impossible to get a 100% match rate. This is ok. The purpose of this exercise is not for you to match all of the records perfectly, but for you to attempt the problem so that we can see how you approach it. We've provided a hint by providing the mapbox_id for each of the POIs so you can see which POIs should match each other (the datasets can be joined on mapbox_id). **You should not use this field in your program** and it is only included for your evaluation purposes on how well your program does in creating matches and possibly as labeled data.

Once you've got a working program that you are happy with please answer the following questions:

1. What are the metrics for your program (accuracy, etc)? How does your FP rate relate to trying to get more matches? What should you optimize for and why?
2. What was your general approach to this problem? How would you scale your approach to millions of POIs?
3. What were the biggest challenges you faced in trying to get a high match rate?
4. If you could change something or get more information to try and get a higher match rate, what would it be?