

Deduplication

Cervantes, Andrew

Names are the same without

‘-‘ or a space

NO

Address or Phone number the same

Any same names when first/last name are switched

Date of birth Is the same

Same first name (married)

Date of birth is the same

Name has “Jr” or “Sr” or a common nickname (ex. Jimmy -> James)

YES

Do Not merge People

More Research Required

Merge People

NO

NO

NO

NO

NO

YES

YES

YES

YES

YES

YES

NO

Is the date of birth Jan 1

NO

YES

Date of birth Similar (Same year, 1 digit off)

YES

NO

Person Merge

Person Merge(Simplified)

Is the Date of Birth the same?

NO

Are the names the same?

YES

Do Not merge People

More Research Required

Merge People

NO

NO

Is Date of birth no more than two digits different and no more than 30 years apart?

NO

YES

YES

YES

Are the Address or Phone Number the same?

Are the names similar (Nicknames, Typos, “Jr”, etc.)?

NO

YES

NO

YES

Are the Address or Phone Number the same?

Algorithm Output

Algorithm is designed to give a percentage that two given people should be merged. From testing, some common percent values **usually** correlate to certain merge situations, such as:

* 1: Means same name and same DOB and same address or phone number
* 0.97: Means same DOB, a nickname/full name combination and same address or phone number
  + Ex. “Rob” and “Robert”
* 0.96: Means same DOB and a slight difference in name or an extra middle initial and same address or phone number
  + Ex. “Jerry R Jones” and “Jerry Jones”
* 0.95: Means DOB is the same and one of the names are one letter off while the other name is exact and same address/phone number
  + Ex. “Bruce Springsteen” and “Bruce Springstean”
* 0.94: Means that the DOB and address or phone number are the same, but the first and last names are flipped for one person (wildcard search)
  + Ex. “Jon Smith” and “Smith Jon”
* 0.93: Means same name and address or phone number, while the DOBs are one number off
  + Ex. “01/01/1967” and “01/01/1968”
* 0.92: Same name and DOB, and one of the addresses is a jail/correctional facility
  + Ex. One home address and one address is “396 Straight Street, Paterson”
* 0.91: Means DOB is off by more than one digit but name and address or phone number match
  + Ex. “01/01/1968” and “01/01/1972” or “01/01/1955” and “01/01/1985”
* 0.9: Means DOB and name are the same, but not address or phone number

**Excel Files to Update**

Files that contain working databases used for the code. Update each of them to your benefit as follows:

* Jail\_addresses.csv: This files contains the street adrress of jails and correctional facilities. If your disease sees a particular address be quite common, add it to this list
* Nicknames\_list.csv: This file contains a list of nicknames and their corresponding full names. If you think of a nickname/fullname combo, add it to this list so the program can catch it for you
* Exclusions.csv: For any IDs that come up that you do not want to match, add one of the IDs to this file so that the same twpo cases will not show up again the next time you run the program

Example Output:

People to Merge

Percent: 0.96

Case IDs: 890355,1596228

Percent: 0.93

Case IDs: 1596701,623780

Percent: 0.93

Case IDs: 1595242,993061

Percent: 0.95

Case IDs: 563459,1175974

Percent: 0.9

Case IDs: 727274,1596177