Jan 23rd Summary

The past week I have been working with the fill patent data and the “DI” Infutor data.

I find

* There are 8,481 inventors with last name starting with “DI”
* When matching them with the Infutor data (matching is based on State, and the first two letters of the First, and the full Last name) I get
  + 1,191 Inventors don’t get matched to any names
  + 1,439 Inventors get matched to only one Infutor ID. For these only 50% have the same city in the Infutor and Patent data. They share 90% of clean first names.
  + For those that match, 50% of Inventors have less than 6 matches, and 50% have more than that
* Seems like some inventors are missing from patent data?
* Seems like PID in Infutor may not truly be unique id? Vincent Dirubbio in the same Zip Code and same City of North East MD, but has two different PID (1)
* Seems like many names are missing in Infutor Data. E.g. Albert Dinnerstein in NY gives me matches with an Allan and Alexander, but no Alberts (2)
* These two cases seem to occur pretty frequently