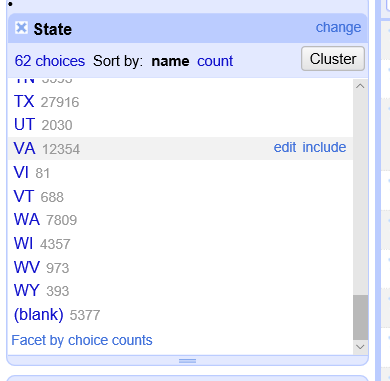
Lab 10: OpenRefine

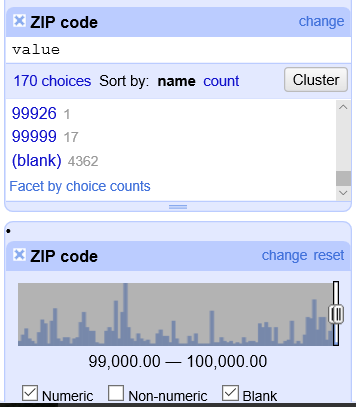
Submission 1:

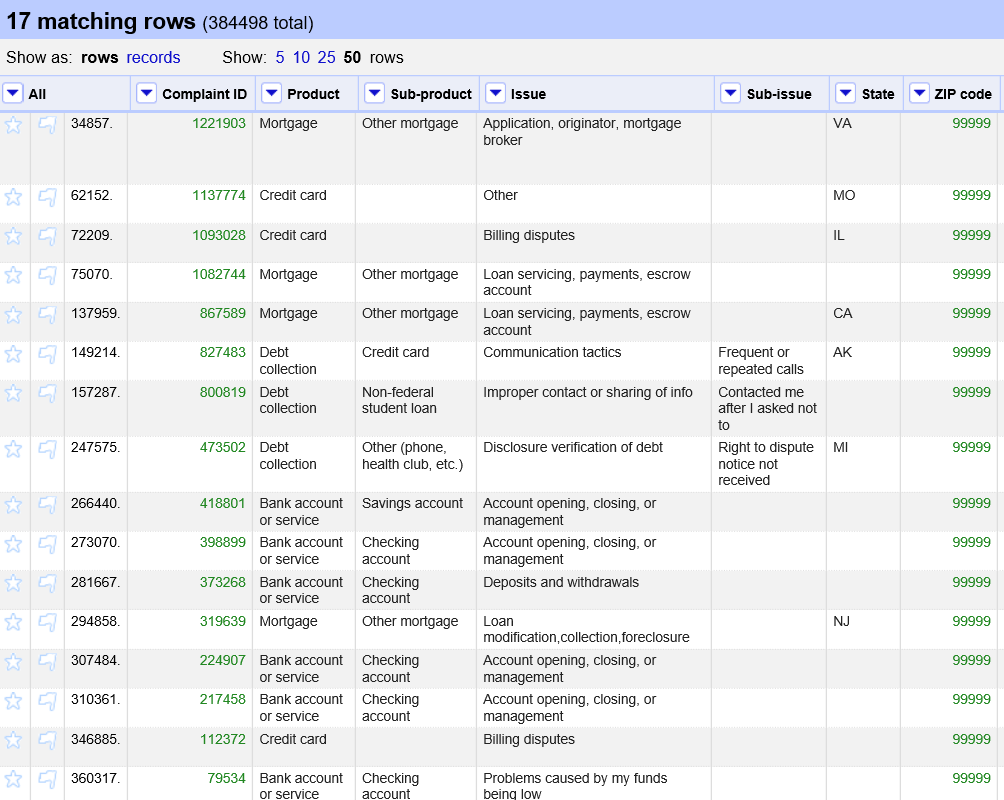
There are 5377 rows without a value for State



Submission 2:

There are 4362 blank zip code entries with 0 non-numeric and some are invalid zip codes (e.g. 99999)





Submission #3

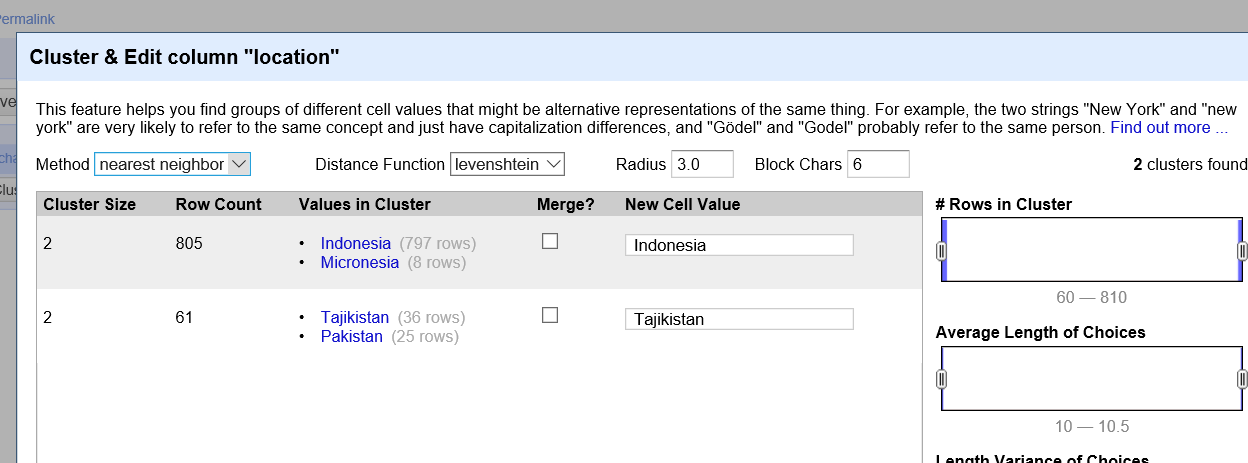
There are 35418 invalid zip codes

Submission #4???

Didn’t exist, I assume it is regarding what to do with nst. I would fill with 1’s because the sample indicates at least one station recording the info (otherwise, how would we have it?)

Submission #5

Don’t want to merge these results

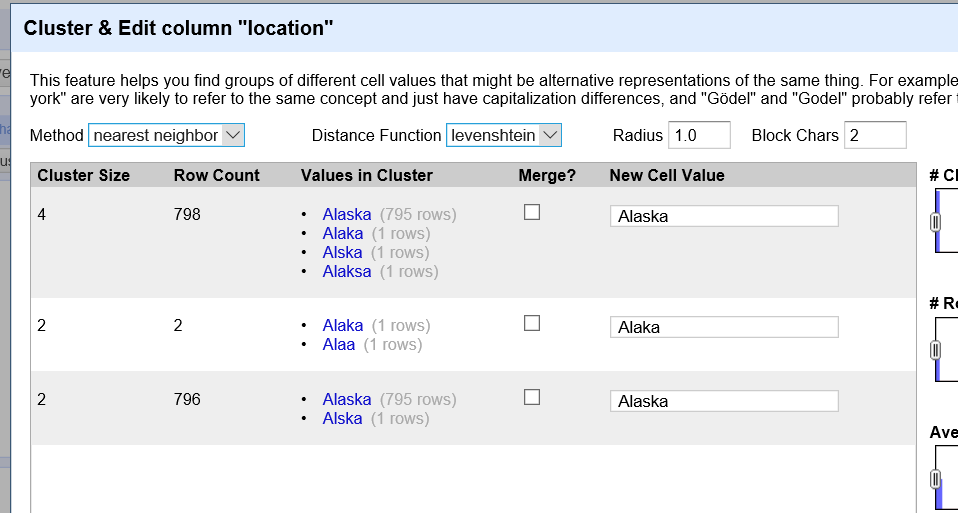




Submission #6

This tool is lovely. Thanks for showing it to us!

I would merge all of these.



Submission #7

This is struggling because of the length of the strings and the block window. Setting it to a very large search window helps the search end quickly. OpenRefine could handle this situation better by auto-suggesting a reasonable block char size based upon compute time (e.g. best search that finishes in a few seconds)

Submission #8

Distance of 3

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|  |  |  | G | U | M | B | A | R | R | E | L |
| 1 |  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 2 | G | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | U | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4 | N | 3 | 2 | 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5 | B | 4 | 3 | 2 | 2 | 1 | 2 | 3 | 4 | 5 | 6 |
| 6 | A | 5 | 4 | 3 | 3 | 2 | 1 | 2 | 3 | 4 | 5 |
| 7 | R | 6 | 5 | 4 | 4 | 3 | 2 | 1 | 1 | 2 | 3 |
| 8 | E | 7 | 6 | 5 | 5 | 4 | 3 | 2 | 2 | 2 | 3 |
| 9 | L | 8 | 7 | 6 | 6 | 5 | 4 | 3 | 3 | 3 | 3 |
| 10 | L | 9 | 8 | 7 | 7 | 6 | 5 | 4 | 4 | 4 | **3** |

