

Submission 1:

5377 rows are missing a value in the state column. By looking at the state facets there is a value for (blank) at the bottom of the state list which lists the number of rows that have state blank.

Submission 2:

4362 having missing ZIP codes.

Submission 3:

345175 valid, 34961 invalid, and 4362 blank

Submission 4:

With a radius of 3.0, some of the values in the clusters are not related. Alaska and California clusters could be merged, but the clusters for Tajikistan and Indonesia would not be merged, because they include different countries.

Submission 5:

With a block size of 2 these two clusters would be worth merging:

```
Alaska(791 rows), alaska(4 rows), Alska(1 rows) => Alaska
```

```
California(84 rows), Caliofrnia(1 rows), Calfiornia(1 rows),  
Cailfornia(1 rows) => California
```

Submission 6:

The strings are a lot longer and there are thousands of them. Calculating the distances for each of them will take a very long time. The application could check how many items you want to cluster and warn you that you have too many or it could provide another distance measurement that's not as resource intensive.

Submission 7:

		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	2	3	4
8	R	7	6	5	5	4	3	2	1	2	3
9	E	8	7	6	6	5	4	3	2	1	2
10	L	9	8	7	7	6	5	4	3	2	1
11	L	10	9	8	8	7	6	5	4	3	2