**Security Manual for Web Based Population Clustering**

**By Nathan Barthen, Connor Barthen, Cole Cassano**

**April 9, 2022**

**Clustering Concerns**:

**Problem**: The polygon size of each school district varies, meaning the cluster

diameter for DBScan cannot be a constant value.

**Solution**: Add a dropdown option on the webpage so the user can specify

the clustering diameter and re-run clustering if they are unhappy with the

results.

**Problem**: School is missing.

**Solution**: Added an ‘Add School’ option so the user can add a school to

the district.

**Problem**: Incorrectly school / incorrect data.

**Solution**: Added an edit option so the user can change the school’s

information before clustering.

**Problem**: School cluster / route buses differently.

**Solution**: Added checkbox beside each school. Only checked schools will

be clustered. Then the user can go back and cluster/route more schools.

**Problem**: DMScan Clustering is inaccurate / might not be working.

**Solution**: Added a rudimentary display to show how the school was

clustered.

**Problem**: User accidentally selected the wrong school to be clustered.

**Solution**: Added a delete feature in the ‘Display all clustering’ page to

delete a clustering.

**What We Could Have Done Better**:

1. The US Census Bureau did not have the countries listed inside of their school’s datasets… This meant that we had to use another site to get the county names (which also told us which school districts were in each county. Then we compared those two datasets when getting the counties / school districts. **As a result**, some of the school districts may not have school data (but data can be added by the user).
2. Currently, when a school gets clustered, it must look at all of the addresses inside of the state (looking only at the county would not work because some schools are in multiple counties). This increased load times for the clustering by about 15-25 seconds. It would be ideal to not look at all of the addresses in Pennsylvania or read the Pennsylvania addresses in parallel.

**References for Project**

**DBScan (Used with Algorithm, Cluster, DataPoint, and DBScan.java):**

 Incorrect example (but useful): <https://www.dataonfocus.com/dbscan-java-code/>

 \* Pseudocode example:

<https://www.researchgate.net/publication/325059373/figure/fig2/AS:624653831790593@1525940487951/Pseudocode-of-the-DBSCAN-algorithm.png>

**Getting School’s Lon/Lat Values**:

How to access information from url - <https://www.javatpoint.com/java-get-data-from-url>

**Reading shapefiles**:

<https://docs.geotools.org/stable/userguide/library/data/shape.html>

-Example on how to read/iterate over a .dbf file (part of a shapefile).

**CSVReader Example (library used for read csv/text files quickly**:

<https://www.journaldev.com/12014/opencsv-csvreader-csvwriter-example>

**Importing / Reading CSV Files (used for CountyDataset.java)**:

<https://www.youtube.com/watch?v=jOKsxtyZrxw&ab_channel=SlashCode>

**HTML/JavaScript**:

<https://www.computerhope.com/issues/ch000317.html>

<https://developer.mozilla.org/en-US/docs/Learn/Forms/Form_validation>

-Used with for edit/add school javascript exception handling

**SMTP Protocol Example**:

<https://www.journaldev.com/2532/javamail-example-send-mail-in-java-smtp>

-Referenced when emailing the user their routing file.