**Background**

The Department of Game, Fish and Parks in South Dakota have a ton of data to run through that has errors and duplicates. Instead of running through all this data manually, I’d like to be able to automate it as much as possible. Most if not all the data is written into csv files, but these files often have different fields. I’m wondering if I can make something that gets rid of duplicates by iterating it through a list based on a field in some user interface. I don’t know of anything like this is ArcPro, running through databases, though I know we do it in code with arcpy. I’m not sure what the viability is for the interface to be with ArcMap or ArcPro, but I don’t really know any others.

This is essentially a giant for loop through the rows of the csv file, looking in the specified field and then deleting the row if it is equal to one of the previous rows. Which I know is doable in a script, but I want to try and make available in an interface with the option to run more than one file through with the same parameters.

**Data and parameters**

Users would need a csv file and the field name they want to remove duplicates from.

**Outputs**

An edited csv file with the rows deleted that were duplicates

**Limitations**

1. This tool would only be used to remove duplicates, not other errors in data, such as spatial reference, integer vs. float vs. string, random points outside of the norm without explanation.
2. User would have to select the field name for each file and the value type of that field name, making it still a tedious endeavor.
3. Users would have to make a copy of the original csv file so that they can go back if a mistake is made.

**Solutions**

1. Could integrate some sort of batch option for csv files that have to same field names, but I don’t know if they would also have to be in the same location.
2. Depending on the interface could have the user specify if the field contains numeric values or text, then use that information in the code to turn whatever is in that row into a string, integer, or float.
3. Could possibly have it take all the rows that aren’t duplicates and use them to create another csv file so that the original stays the original. This is a lot more work than deleting rows is though.

**Problems/Roadblocks**

1. I don’t know of an interface that I could use. Would the csv files have to first be turned into shapefiles for ArcPro/ArcMap to use them? I don’t think so, but the csv files might need to have something done to them for cursors and readers to be used outside of a script.
2. Do the field names need to be in the same location on the row for multiple files to be processed at once in a batch?
   1. With cursors I don’t think they do; the field names just need to be specified.