

## Exercise 05 - Due Friday, November 18, at 12pm

*Objectives:* Gain experience wrangling data using an automated tool, Trifacta. Identify transformations that will make a source dataset easier to work with, and implement them. Evaluate the tool in context.

*Deadline:* Friday, November 18, 12pm. Submit a zip file containing your writeup, information about your source dataset, your wrangle script, and the resulting transformed dataset to Blackboard. You are not required to post this to GitHub, although you may, after the deadline.

### Part 1 - Select a dataset

Choose a dataset from either [us-city.census.okfn.org](http://us-city.census.okfn.org) or [data.gov](http://data.gov). Download the data as either a CSV or Excel file.

### Part 2 - Wrangling with Trifacta

Install Trifacta Wrangler on your own machine. Note that while it asks you to establish an account, the product is free of charge.

Add your dataset to Trifacta and transform two or more columns to make them more useful for analysis.

### Part 3 - Export results

Use the “Generate Results” function to export your data. Be sure to “Profile Results”. Review the results. Also, be sure to save your transform script.

### Part 4 - Writeup

Write a 1-2 page summary of your process. It should include the following:

- A high-level description of your dataset
- An explanation of your transformations, in English
- Your wrangle script (exported from Trifacta)
- An evaluation of your results. Did the transformations work? Did the full results show you something you missed when writing the script?

**Submit a zip file** containing your writeup, including all the components specified in Part 4, to Blackboard by the deadline.