

## Project: Capstone Project 1: Data Wrangling

 6 - 10 Hours

Now that you have a basic ideas of the various data wrangling steps and techniques available, let's apply it to your capstone project.

The first step in completing your capstone project is to collect data. Depending on your dataset, you may apply some of the data wrangling techniques that you learned in this unit. Some of you may be using standard datasets and sources, such as Kaggle or Yelp, where minimal or no data wrangling is required. Students often find that this part of the project takes a lot longer than they estimated, which is completely normal. The more work you put in, the more you'll learn. Data wrangling is an important tool in a data scientist's toolbox!

### Steps:

1. Create a Google Doc (1-2 pages) describing the data wrangling steps you took to clean the dataset. Include answers to these questions in your submission:
  1. What kind of cleaning steps did you perform?
  2. How did you deal with missing values, if any?
  3. Were there outliers, and how did you handle them?
2. Submit a link to the document.

3. Discuss it with your mentor at the next call.
4. Revise and resubmit if needed.
5. Convert the final document to a .pdf and add it to your GitHub repository for this project. This document will eventually become part of your milestone report.

Your project will be evaluated using this [rubric](#).

You've started building your data science portfolio! Keep it neat and well-organized to make it easier for you later on in the course and when you interview for jobs. At this point, your portfolio should include two mini-projects, a folder for a capstone project with a project proposal, and the data wrangling report. Additionally, your dataset should be ready to start exploratory data analysis.