**Learning Log: Develop your approach to cleaning data**

**Instructions**You can use this document as a template for the learning log activity: Consider how data analysts approach tasks. Type your answers in this document, and save it on your computer or Google Drive.

We recommend that you save every learning log in one folder and include a date in the file name to help you stay organized. Important information like course number, title, and activity name are already included. After you finish your learning log entry, you can come back and reread your responses later to understand how your opinions on different topics may have changed throughout the courses.

To review detailed instructions on how to complete this activity, please return to Coursera: [Learning Log: Develop your approach to cleaning data](https://www.coursera.org/learn/process-data/supplement/rKHUX/learning-log-develop-your-approach-to-cleaning-data).

| **Date:** <enter date> | **Course/topic:** Course 4: Process Data from Dirty to Clean | | |
| --- | --- | --- | --- |
| **Learning Log:** Develop your approach to cleaning data | | |
| **Think about “clean” and “dirty” data** | Use the instructions from the learning log to complete the Data Cleaning Approach Table below:   | **Data cleaning checklist** | **Preferred cleaning methods** | **Data cleaning motto** | | --- | --- | --- | | * How many rows and columns are there? * What is the schema? * What is the type of each field? * For numerical types, what is the range of this field? * What data is missing? * Are the formats consistent? * Are there any obvious inaccuracies? | * Know the expected ranges, types, and formats. * Use spreadsheet functions to reformat data. * Always make a copy of the raw data for cleaning. Do not modify raw data until the cleaning tasks are done. * Double check that everything has been cleaned correctly. * Write modified data to its own spreadsheet for submission. | * The data always has more tricks, so use your eyes! | | | |
| **Reflection:** | Write 1-2 sentences (20-40 words) answering each of the following questions: | | |
| **Questions and responses:** | * What items did you add to your data cleaning checklist? Why did you decide these were important to check for?   *To see what items I added, look at the list above. I chose to add these items based on both the principles taught in this certificate, as well as experience.*   * How have your own experiences with data cleaning affected your preferred cleaning methods? Can you think of an example where you needed to perform one of these cleaning tasks?   *Yes! My NICU HRV project required all of these tasks. In fact, since that was my first project, I overestimated the amount of cleaning I’d have to do, and that ended up taking a considerable amound of time. There were many mistakes that I didn’t catch until after I had tried to analyze the data, the finding of which resulted in a lot of frustrating backtracking. The situation reminded me of Murphy’s Law, that if it can go wrong, it will. I’ve learned from this that you have to be extremely scrutinizing when you’re cleaning data. The output of every single process you apply to the data has to be double checked.*   * How did you decide on your data cleaning motto?   *I decided my motto based on the frustrating experience described above.* | | |