Learning Log: Develop your approach to cleaning data

Overview

By this point, you have started working with real data. And you may have noticed that data is often messy-- you can expect raw, primary data to be imperfect. In this learning log, you will develop an approach to cleaning data by creating a cleaning checklist, considering your preferred methods for data cleaning, and deciding on a data cleaning motto. By the time you complete this entry, you will have a stronger understanding of how to approach the data cleaning process methodically. This will help you save time cleaning data in the future and ensure that your data is clean and usable.

Fill out the Data Cleaning Approach Table

The problem with data cleaning is that it usually requires a lot of time, energy, and attention from a junior data analyst. One of the best ways to lessen the negative impacts of data cleaning is to have a plan of action or a specific approach to cleaning the data.

In order to help you develop your own approach, you'll use the instructions from this learning log to fill out a Data Cleaning Approach Table in your learning log template. ⬀ The table will appear like this in the template:

Data cleaning checklist	Preferred cleaning methods	Data cleaning motto
• • • •	• • • •	

Once you have completed your Data Cleaning Approach Table, you will spend some time reflecting on the data cleaning process and your own approach.

Access your learning log

To use the learning log for this course item, click the link below and select "Use Template."

Link to learning log template: Develop your approach to data cleaning ⬀

OR

If you don't have a Google account, you can download the template directly from the attachment below.

> 📎 **Learning Log Template_ Develop your approach to cleaning data**
> DOCX File

Step 1: Create your checklist

You can start developing your personal approach to cleaning data by creating a standard checklist to use before your data cleaning process. Think of this checklist as your default "what to search for" list.

With a good checklist, you can efficiently and, hopefully, swiftly identify all the problem spots without getting sidetracked. You can also use the checklist to identify the scale and scope of the dataset itself.

Some things you might include in your checklist:

- Size of the data set
- Number of categories or labels
- Missing data
- Unformatted data
- The different data types

You can use your own experiences so far to help you decide what else you want to include in your checklist!

Step 2: List your preferred cleaning methods

After you have compiled your personal checklist, you can create a list of activities you like to perform when cleaning data. This list is a collection of procedures that you will implement when you encounter specific issues present in the data related to your checklist or every time you clean a new dataset.

For example, suppose that you have a dataset with missing data, how would you handle it? Moreover, if the data set is very large, what would you do to check for missing data? Outlining some of your preferred methods for cleaning data can help save you time and energy.

Step 3: Choose a data cleaning motto

Now that you have a personal checklist and your preferred data cleaning methods, you can create a data cleaning motto to help guide and explain your process. The motto is a short one or two sentence summary of your philosophy towards cleaning data. For example, here are a few data cleaning mottos from other data analysts:

1. "Not all data is the same, so don't treat it all the same."
2. "Be prepared for things to not go as planned. Have a backup plan."
3. "Avoid applying complicated solutions to simple problems."
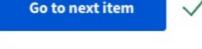
The data you encounter as an analyst won't always conform to your checklist or activities list regardless of how comprehensive they are. Data cleaning can be an involved and complicated process, but surprisingly most data has similar problems. A solid personal motto and explanation can make the more common data cleaning tasks easy to understand and complete.

Reflection

Now that you have completed your Data Cleaning Approach Table, take a moment to reflect on the decisions you made about your data cleaning approach. Write 1-2 sentences (20-40 words) answering each of the following questions:

- What items did you add to your data cleaning checklist? Why did you decide these were important to check for?
- 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?
- How did you decide on your data cleaning motto?

[Go to next item]   ✓ Completed

👍 Like   👎 Dislike   🚩 Report an issue