**COURSE 4 MODULE 4**

When you clean data, you make changes to the original dataset. It’s important to verify the changes you make are accurate and to let your teammates know about the changes. In this part of the course, you’ll learn to verify that data is clean and report your data cleaning results. With verified clean data, you’re ready to begin analyzing!

### **Learning Objectives**

* Describe the process involved in verifying the results of cleaning data
* Describe what is involved in manually cleaning data
* Discuss the elements and importance of data-cleaning reports
* Describe the benefits of documenting data cleaning process

**MANUALLY CLEANING DATA**

[**VERIFY AND REPORT RESULTS**](https://www.coursera.org/learn/process-data/lecture/eKsSc/verify-and-report-results)

You've been learning a lot about the importance of clean data and explored some tools and strategies to help you throughout the cleaning process.

In these videos, we'll be covering the next step in the process: **verifying and reporting on the integrity of your clean data**.

**Verification** is a process to confirm that a data cleaning effort was well-executed and the resulting data is accurate and reliable.

It involves :

1. Rechecking your clean dataset, doing some manual clean ups if needed
2. Taking a moment to sit back and really think about the original purpose of the project.

That way, you can be confident that the data you collected is credible and appropriate for your purposes.

**Making sure your data is properly verified is so important because it allows you to double-check that the work you did to clean up your data was thorough and accurate.**

For example, you might have referenced an incorrect cell phone number or accidentally keyed in a typo. Verification lets you catch mistakes before you begin analysis. Without it, any insights you gain from analysis can't be trusted for decision-making. You might even risk misrepresenting populations or damaging the outcome of a product that you're actually trying to improve.

I remember working on a project where I thought the data I had was sparkling clean because I'd use all the right tools and processes, but when I went through the steps to verify the data's integrity, I discovered a semicolon that I had forgotten to remove. Sounds like a really tiny error, I know, but if I hadn't caught the semicolon during verification and removed it, it would have led to some big changes in my results. That, of course, could have led to different business decisions.

There's an example of **why verification is so crucial**. But that's not all.

The other big part of the **verification process** is reporting on your efforts. Open communication is a lifeline for any data analytics project. Reports are a super effective way to show your team that you're being 100 percent transparent about your data cleaning.

**Reporting is also a great opportunity to show stakeholders that you're accountable, build trust with your team, and make sure you're all on the same page of important project details**.

Coming up, you'll learn different strategies for reporting, like creating data-cleaning reports, documenting your cleaning process, and using something called the **changelog**.

A **changelog** is a file containing a chronologically ordered list of modifications made to a project.

It's usually organized by version and includes the date followed by a list of added, improved, and removed features.

**Changelogs are very useful for keeping track of how a dataset evolved over the course of a project.** They're also another great way to communicate and report on data to others. Along the way, you'll also see some examples of how verification and reporting can help you avoid repeating mistakes and save you and your team time.

Let's go!

[**CONFIRM DATA-CLEANING MEETS BUSINESS EXPECTATIONS**](https://www.coursera.org/learn/process-data/lecture/wmjwi/confirm-data-cleaning-meets-business-expectations)

re

[**STEP-BY-STEP: VERIFICATION OF DATA CLEANING**](https://www.coursera.org/learn/process-data/supplement/Bc9rE/step-by-step-verification-of-data-cleaning)

re

[**VERIFICATION OF DATA CLEANING**](https://www.coursera.org/learn/process-data/lecture/Hx69i/verification-of-data-cleaning)

re

[**DATA-CLEANING VERIFICATION CHECKLIST**](https://www.coursera.org/learn/process-data/supplement/c9vis/data-cleaning-verification-checklist)

re

[**TEST YOUR KNOWLEDGE ON MANUAL DATA CLEANING**](https://www.coursera.org/learn/process-data/quiz/5Ntpp/test-your-knowledge-on-manual-data-cleaning)

re

**DOCUMENT THE CLEANING PROCESS**

[**CAPTURE CLEANING CHANGES**](https://www.coursera.org/learn/process-data/lecture/jjOrO/capture-cleaning-changes)

re

[**EMBRACE CHANGELOGS**](https://www.coursera.org/learn/process-data/supplement/FvuSF/embrace-changelogs)

re

[**SELF-REFLECTION: CREATING A CHANGELOG**](https://www.coursera.org/learn/process-data/quiz/ejgYq/self-reflection-creating-a-changelog)

re

[**WHY DOCUMENTATION IS IMPORTANT**](https://www.coursera.org/learn/process-data/lecture/TvyoD/why-documentation-is-important)

re

[**FEEDBACK AND CLEANING**](https://www.coursera.org/learn/process-data/lecture/k3c6N/feedback-and-cleaning)

re

[**ADVANCED FUNCTIONS FOR SPEEDY DATA CLEANING**](https://www.coursera.org/learn/process-data/supplement/PLnRS/advanced-functions-for-speedy-data-cleaning)

re

[**TEST YOUR KNOWLEDGE ON DOCUMENTING THE CLEANING PROCESS**](https://www.coursera.org/learn/process-data/quiz/1t2Tf/test-your-knowledge-on-documenting-the-cleaning-process)

re

**MODULE 4 CHALLENGE**

[**GLOSSARY TERMS FROM MODULE 4**](https://www.coursera.org/learn/process-data/supplement/JIVvr/glossary-terms-from-module-4)

**[MODULE 4 CHALLENGE](https://www.coursera.org/learn/process-data/exam/7OicY/module-4-challenge)**

**MODULE 4 CHALLENGE**

**TERMS AND DEFINITIONS FOR COURSE 4, MODULE 3**