

File organization guidelines

Every data analyst's goal is to conduct efficient data analysis. One way to increase the efficiency of your analyses is to streamline processes that help save time and energy in the long run. Meaningful, logical, and consistent file names help data analysts organize their data and automate their analysis process. When you use consistent guidelines to describe the content, date, or version of a file and its name, you're using file naming conventions.

In this reading, you'll learn more about best practices for file naming conventions and file organization.



Best practices for naming files

File-naming conventions help you organize, access, process, and analyze data because they act as quick reference points to identify what's in a file. One important practice is to decide on file naming conventions—as a team or company—early in a project. This will prevent you from spending time updating file names later, which can be a time-consuming process. In addition, you should align your project's file names with your team's or company's existing file-naming conventions. You don't want to spend time learning a new file-naming convention each time you look up a file in a new project! It's also critical to ensure that file names are meaningful, consistent, and easy-to-read. File names should include:

- The project's name
- The file creation date
- Revision version
- Consistent style and order

Further, file-naming conventions should act as quick reference points to identify what is in the file. Because of this, they should be short and to the point.

In the following sections, you'll explore each part of a sales report file name that follows an established naming convention, **SalesReport_20231125_v02**. This example will help you understand the key parts of a strong file name and why they're important.

Name

Giving a file a meaningful name to describe its contents makes searching for it straightforward. It also makes it easy to understand the type of data the file contains.

In the example, the file name includes the text **SalesReport**, a succinct description of what the file contains: a sales report.

Creation date

Knowing when a file was created can help you understand if it is relevant to your current analysis. For example, you might want to analyze only data from 2023.

In the example, the year is described as **20231125**. This reads as the sales report from November 25, 2023 following the year, month, and day (YYYYMMDD) format of the international date standard. Keep in mind that different countries follow different date conventions, so make sure you know the date standard your company follows.

Revision version

Including a revision version helps ensure you're working with the correct file. You wouldn't want to make edits to an old version of a file without realizing it! When you include revision numbers in a file name, lead with a zero. This way, if your team reaches more than nine rounds of revisions, double digits are already built into your convention.

In the example, the version is described as **v02**. The v is short for the version of the file, and the number following the v indicates which round of revisions the file is currently in.

Consistent order and style

Make sure the information you include in a file name follows a consistent order. For example, you wouldn't want version three of the sales report in the example to be titled

20231125_v03_SalesReport. It would be difficult to find and compare multiple documents.

When you use spaces and special characters in a file name, software may not be able to recognize them, which causes problems and errors in some applications. An alternative is to use hyphens, underscores, and capital letters. The example includes underscores between each piece of information, but your team could choose to use hyphens between year, month, and date, too:

SalesReport_2023_11_25_v02.

Ensure team consistency

To ensure all team members use the agreed-upon file naming conventions, create a text file as a sample that includes all of the naming conventions on a project. This can benefit new team members to help them quickly get up to speed or a current team member who just needs a refresher on the file naming conventions.

File organization

To keep your files organized, create folders and subfolders—in a logical hierarchy—to ensure related files are stored together and can be found easily later. A hierarchy is a way of organizing files and folders. Broader-topic folders are located at the top of the hierarchy, and more specific subfolders and files are contained within those folders. Each folder can contain other folders and files. This allows you to group related files together and makes it easier to find the files you need. In addition, it's a best practice to store completed files separately from in-progress files so the files you need are easy to find. Archive older files in a separate folder or in an external storage location.

Key takeaways

Use consistent, meaningful file-naming conventions throughout your project to save you and your team time by making data easy to find and use. File-naming conventions should be agreed upon by all team members before starting a project and should describe the project by including its name, the date, and the revision version. Document this information in a location that team members can access.