

Types of automation for the data lifecycle

You've learned that automation is the use of software, scripting, or machine learning to perform data analysis processes without human work. Data analytics automation enables you to develop processes to automatically run some phases of the data lifecycle without manually performing tasks. This can be critical to performing continuous data analysis, or to creating time for human analysis of business problems. In this reading, you'll learn more about data analytics automation, and explore some examples of how automation works at each phase of the data lifecycle.

Data analytics automation

When working as a data analyst, you may find that you're performing the same tasks on a regular basis. For example, you might upload data sets every day, or update reports and dashboards regularly. Automating this work can give you more time to analyze the data itself.

There are several types of automation you might use, including software-based, scripted, and machine learning.

Software-based automation

In data analysis, software staples, from Excel and Google Sheets, to BigQuery, offer opportunities to automate your work. In Excel and Google Sheets, you can auto-sort or remove duplicate data using menu options. BigQuery offers the same options to filter your data, or extract outliers.

Scripted automation

If you use SQL or Python, or another programming language to manage your data, you can write scripts to automate data lifecycle phases. For example, you could write a Python script that automatically creates a summary report of your data. Then, you can set the Python script to run at a set interval to generate the report at a frequency of your choice.

Machine learning

Training machine learning models can also speed up some of your data analysis tasks. You can use machine learning to predict patterns in datasets, or classify data. Once your model is trained, you can have these tasks occur on a regular basis. You can also use machine learning to perform error analysis.

Key takeaways

While some aspects of data analysis require human interpretation, there are aspects of every stage of the data lifecycle that can be automated. Whether you automate with software, script, or machine learning, you can use technology to avoid repetitive tasks, and focus on the human-driven insights that will benefit your organization.

Resources for more information

These links provide more information about data automation strategies and processes:

- Google Cloud gives options for how data lifecycles can be controlled:
<https://cloud.google.com/storage/docs/control-data-lifecycles>
- Indeed defines data automation and gives examples of its benefits:
<https://www.indeed.com/career-advice/career-development/what-is-data-analytics-automation>