

Nulls and aggregate data

Previously, you learned about one of the most common data transformations that analysts perform: aggregation. Data aggregation is the process of gathering data and expressing it in a summary form. Aggregation is useful for managing data, making data more accessible, and observing data trends. In this reading, you'll review data aggregation and explore one of the most common challenges with data aggregation: **NULLs**.

Benefits of data aggregation

Data aggregation has three main benefits:

- **Managing data:** The increased volume and velocity of incoming data can create storage issues for organizations that want to maintain historical data, but have limited space. Aggregation enables organizations to just store relevant and meaningful data, and eliminate storage clutter by removing old and irrelevant data.
- **Promoting accessibility:** It can be problematic for users to have to sort through large amounts of data to find the answers they need. This also makes the data less accessible and useful to your organization. Using aggregations to reduce the data your team has to sift through to get what they need makes it more effective for analysis.
- **Observing data trends:** Recognizing trends in the data is what allows data teams to provide insights that guide an organization's decision making. Aggregating data creates summarized data that's easier to understand and recognize patterns in.

As a cloud data professional, you'll probably use aggregations to transform data often. So, understanding the benefits of aggregation can help you better implement aggregations to data in the future.

The trouble with NULLs

Generally, most aggregate functions ignore **NULL** values in columns. A **NULL** value means there's no available value for a field; it's not the same as a zero. This is usually the desired outcome of an aggregate function. Ignoring **NULL** values prevents unknown or inapplicable values from affecting calculations performed using the data. For example, if you're trying to calculate the average number of years employees worked at a company, and the **NULL** values were treated as zeros, this would create issues with your results. As a cloud data professional, understanding how aggregate functions address **NULLs** means you can better understand the results of your aggregations, and also apply them more intentionally if you want to address **NULLs** in a specific way.

Key takeaways

As a cloud data professional, transforming data will likely be an important part of most of your future projects. Aggregations are one of the most common types of transformations, so understanding the benefits of aggregating data is useful for understanding the work you'll be doing in the future.