The open data debate

Just like data privacy, open data is a widely debated topic in today's world. Data analysts think a lot about open data, and as a future data analyst, you need to understand the basics to be successful in your new role.

What is open data?

In data analytics, **open data** is part of **data ethics,** which has to do with using data ethically. **Openness** refers to free access, usage, and sharing of data. But for data to be considered open, it has to:

- Be available and accessible to the public as a complete dataset
- Be provided under terms that allow it to be reused and redistributed
- Allow universal participation so that anyone can use, reuse, and redistribute the data

Data can only be considered open when it meets all three of these standards.

The open data debate: What data should be publicly available?

One of the biggest benefits of open data is that credible databases can be used more widely. Basically, this means that all of that good data can be leveraged, shared, and combined with other data. This could have a huge impact on scientific collaboration, research advances, analytical capacity, and decision-making. But it is important to think about the individuals being represented by the public, open data, too.

Third-party data is collected by an entity that doesn't have a direct relationship with the data. You might remember learning about this type of data earlier. For example, third parties might collect information about visitors to a certain website. Doing this lets these third parties create audience profiles, which helps them better understand user behavior and target them with more effective advertising.

Personal identifiable information (PII) is data that is reasonably likely to identify a person and make information known about them. It is important to keep this data safe**.** PII can include a person's address, credit card information, social security number, medical records, and more.

Everyone wants to keep personal information about themselves private. Because third-party data is readily available, it is important to balance the openness of data with the privacy of individuals.

Resources for open data

Luckily for data analysts, there are lots of trustworthy resources available for open data. It is important to remember that even reputable data needs to be constantly evaluated, but these websites are a useful starting point:

1. **U.S. government data site**: Data.gov is one of the most comprehensive data sources in the US. This resource gives users the data and tools that they need to do research, and even helps them develop web and mobile applications and design data visualizations.
2. **U.S. Census Bureau**: This open data source offers demographic information from federal, state, and local governments, and commercial entities in the U.S. too.
3. **Open Data Network**: This data source has a really powerful search engine and advanced filters. Here, you can find data on topics like finance, public safety, infrastructure, and housing and development.
4. **Google Cloud Public Datasets**: There are a selection of public datasets available through the Google Cloud Public Dataset Program that you can find already loaded into BigQuery.
5. **Dataset Search:** The Dataset Search is a search engine designed specifically for data sets; you can use this to search for specific data sets.