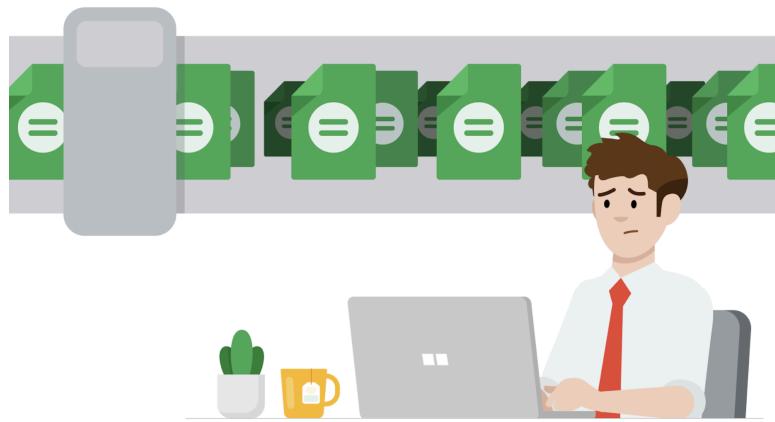# Google

Reading: Anticipate data limitations

We live in a world where data is constantly being generated. There is so much information out there to learn from. But we also live in a world that is constantly changing, and often the data that we encounter has certain limitations we need to consider as we analyze data and draw insights from it.

Factors of data availability

Previously, you learned about the importance of data availability, which is the degree or extent to which timely and relevant information is readily accessible and able to be put to use. The factors that influence data availability are:

- Data integrity: The accuracy, completeness, consistency, and trustworthiness of data throughout its life cycle.
- Data visibility: The degree or extent to which information can be identified, monitored, and integrated from disparate internal and external sources.
- Update frequency: How often disparate data sources are being refreshed with new information.
- Change: The process of altering data, either through internal processes or external influence.

Next, you are going to consider the limitations of data that might change the availability and how you can anticipate those limitations as a BI professional.

Missing data

If you have incomplete or nonexistent data, you might not have enough data to reach a conclusion. Or, you might even be exploring data about a totally different business problem! Understanding what data is available, identifying potential other sources, and filling in the gaps is an important part of the BI process.

Misaligned data

As a BI professional, you will often use data from different sources. Some of these might be internal sources to the business you're working with, but they might also include external sources. These sources might define and measure things in completely different ways. In cases like these, establishing how to measure things early on standardizes the data across the board for greater reliability and accuracy. This will make sure comparisons between sources are meaningful and insightful.

Dirty data

Dirty data refers to data that contains errors. Dirty data can cause errors in your system, inaccurate reports, and poor decision-making. Implementing processes for cleaning data by fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset is one way you can prepare for this limitation.

Conclusion

As a BI professional, you'll need to understand that sometimes the data you work with will have limitations. This could mean that it doesn't fit within a certain time range, or it only applies to specific situations, or there are challenges identifying the data you need. Being able to anticipate those issues and consider them when you build tools and systems for your business will allow you to ensure that those limitations don't stop your stakeholders from getting the data they need to make great decisions and ensure project success!