Primera

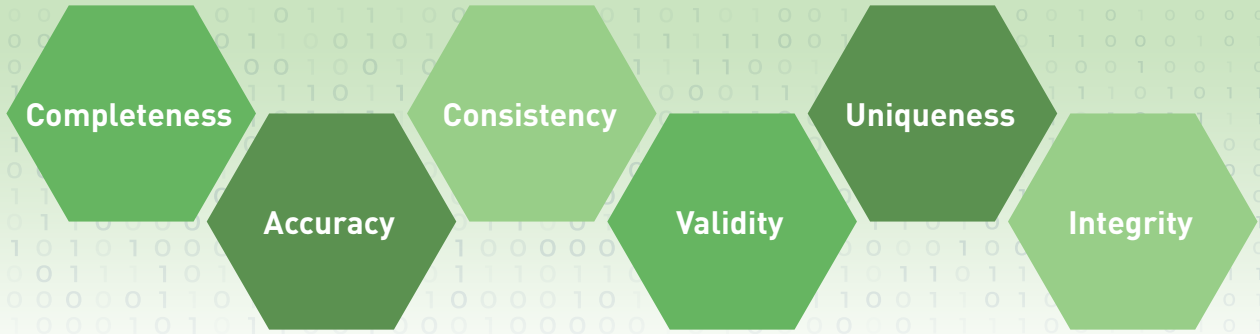
DATA QUALITY 101

6 CHARACTERISTICS *THAT YIELD USEFUL OUTPUT*

BY: ALEXANDRIA MELTON

As our world technologically advances, more and more data is being collected, stored, and used. Data is powerful. It's the foundation of all information. Harnessing the power of data helps to improve our quality of life, create targeted messages, predict trends, improve business processes, and much more.

However, to do any of this well, your data quality should be inspected and maintained. There are two key questions to ask before gathering data that will ensure what you collect is of the highest quality. First ask, is the data we're collecting fit for our intended purpose? Second, does the data represent the real-world scenario? When you have high quality data from the start, the information derived from that data will also be of high quality. This in turn leads to wiser decision making and better business outcomes. How else can you ensure that your data is high quality? By incorporating the following six quality check points as part of your best practices. We've also included a few Excel spreadsheet tips that may benefit your data efforts as well.



6 Data Quality Checks

1. Completeness

Your The data should you begin with should be as complete as possible to start withhave the minimum amount of information in order toto provide the necessary targeted information you're looking for. For example, if you would like to upgrade streetlights to LEDs, then then the you would need the lolocation for of each streetlight;, the current lamp type;, wattage;, lamp style;, overhead or underground fed; etc. would be necessary information to have from the start. These are just a few examples of attributes that are vital to efficiently upgrade your streetlights. If any one of these attributes is missing data, it the data would be considered incomplete.

Excel Tips: You can check for completeness within your data by using the aggregate function in Eexcel. The following This formula will generate a percentage of blank values within your data set:

$$\frac{\text{COUNTBLANK(Data Array)}}{\text{COUNTBLANK(Data Array)+COUNTA(Data Array)}}$$

2. Accuracy

Data accuracy is the level in which your data represents the real-world scenario. Measuring data accuracy requires validation with another

reference. In our streetlight example, our other reference would be existential what is infactors within the field. Does the latitude and longitude take you to the correct location? Is Does the light a operate with the use of high-pressure sodium? If your answer is no, then your data may not be accurate.

3. Consistency

Consistency can be difficult to measure if there is your data is contained to only one system that contains your data. If your data exists in multiple systems, then the information for one streetlight should be the same across all systems. Data consistency and data accuracy often go hand- in-hand.

4. Validity

The validity of your data checks if the attributes have use the correct formatting. Should your data be in all capsCAPS? If the attribute is a date, does every data point have the same date format? If there are anyany invalid data exists, it will affect the completeness, and therefore, the validity of your data.

5. Uniqueness

Uniqueness indicates if there is a single-recorded instances within your dataset. There should not be any duplicates within your data.

Excel Tips: The best way to check uniqueness is by using the duplicate values function in the conditional formatting section. Be sure to apply this function to your unique identifier column.

6. Integrity

Data integrity indicates if your data has been maintained properlyProperly maintained data will ensure data integrity. If your data exists in multiple systems, then those systems should be connected. When one system is updated, that update will should be pushed to all other systems.

Additional Tips to Ensure Ongoing Data Quality

To achieve ongoing accuracy, consistency, and validity of your data, consider utilizing a centralized database and keeping all data sources in sync. When you have multiple data sources it is important to keep them connected. Real-time syncing allows users to access the most up-to-date information.

Additionally, your data entry process should be standardized. When a defined process for entering data exists, consistency and data quality is near guaranteed from the moment it is entered into the system. The best process is usually the simplest. With digital entry, it's easy to make fields required, thus ensuring minimal gaps within your dataset. Another way to ensure completeness and accuracy is to implement drop-down options for each field.

Data management and quality control is vital to the health - and therefore growth - of any business. High quality data allows for better decision making across organizations, more prosperous marketing campaigns and improved customer relationships.

References

- The 6 Dimensions of Data Quality
- 10 Tips on How to Improve Data Quality
- Data Quality - What, Why, How, 10 Best Practices & More

About The Author

ALEXANDRA MELTON

BUSINESS ANALYST III

Alexandria Melton is a Business Analyst in Primera's Consulting division. As an experienced research assistant with a background in higher education, Alexandria is a key member of our data and analytics team and helps develop performance metrics for executive reporting and assists with automating work processes. She is skilled in mathematics, Matlab, data analysis, and science, and has a master's degree in statistics from Loyola University Chicago.