# DAT 325 Project One Template

# Data Quality Plan

Replace the bracketed text with your responses.

### Purpose Statement:

### We need data to be of high quality, otherwise the data will not be *as* insightful or meet the expectations of the company. If a data is not accurate, complete, consistent, valid, unique, or lacks integrity, then it will not be as reliable, and therefore any decisions made based on the results of studying such data will be just as unreliable.

### Organizational Goals:

### [Determine at least three goals of the organization with respect to the data.]

1. We need the data incoming from Wayne Enterprises to be cleaned and of high quality. If errors exist in the data prior to migration, then there will likely be more errors after the migration.
2. We need to ensure that the integrity of the data and the quality of the data are maintained during migration. If the migration process is interrupted in any way, suffers from interference, etc., then the data will suffer quality-wise.
3. Once the data is migrated, it needs to once again be checked for quality, transformed if needed, and then stored securely for long term use and study purposes. Encryption of the data for long term storage, and managed access to the data is recommended to help ensure higher quality data over the lifetime of the data.

### Data Quality Characteristics and Procedures:

1. **Accuracy:** the level of which data represents the “thing” being measured. Accurate data can be verified. Accuracy is critical and often the most difficult dimension to control, and without it, reporting may not be trusted.
2. **Completeness:** This dimension could have many implications for different fields and applications, but in a general sense means that you have the minimum data required to perform a study or engage with. The data needs to have all necessary attributes attached. You cannot measure data that doesn’t exist.
3. **Consistency:** The data needs to represent the same information when stored and used at multiple instances throughout its lifetime (Gupta, 2022).
4. **Integrity:** When transformed and migrated or transported across multiple systems throughout its lifetime, the attributes and relationships of data can be compromised. Both integrity and consistency must be controlled by planned testing across multiple data sets throughout the lifetime of the data (Gupta, 2022).
5. **Validity:** Data validity can in large part be controlled by *validation* processes, and ensures that the data values and attributes are of the proper type, range, color, etc. For example, you want customer birthdays to be actual dates or in a certain format (mm/dd/yyyy or dd/mm/yyyy for example).
6. **Uniqueness:** Data uniqueness is used to ensure that duplicate measurements or entries are kept to a minimum. I find that the use of composite keys can help to ensure that each measurement is unique.

### Security and Personnel Responsibility Plan:

Least privilege and authentication in combination can go a long way towards data security and quality of data, just to use one example. If anyone can go in and access data, or manipulate data, of course the integrity of data will suffer. Limiting the use of data to qualified individuals can ensure that the data is handled safely.

Gupta, A. (2022, July 14). *The 6 dimensions of data quality*. Collibra. https://www.collibra.com/us/en/blog/the-6-dimensions-of-data-quality