**Data and Analysis**

11.0101.DA

*Students can demonstrate the concepts, practices and process of data collection, resource management, applying statistical and graphical techniques to different types of data in order to discover useful information that can communicate storytelling and to inform decision-making.*

| **Collection**  **11.0101.DA-01** | Data is constantly collected or generated through automated processes that are not always evident, raising privacy concerns. The different collection methods and tools that are used influence the amount and quality of the data that is observed and recorded. | | | |
| --- | --- | --- | --- | --- |
| **Level 1** | **Level 2** | **Level 3** | **Level 4** |
| Identify privacy and ethical implications of sourcing internal and external data  Define key business indicators and target improvement metrics  Determine what data is needed to test a hypothesis | Collect and present the same data in various visual formats.  Demonstrate SQL skills for querying databases and joining tables  Collaborate with cross-functional stakeholders to identify  quantifiable improvements  Ensure hypotheses to be tested are aligned with business value | Organize and present collected data visually to highlight relationships and support a claim.  Demonstrate ability to work with data from multiple data sources: SQL Databases, NoSQL Databases, csv and json files  Leverage business acumen to understand how to take a business problem and put it into quantifiable form.  Formulate business problem as a research question with associated hypotheses | Build a report with findings  Deliver a presentation sharing insights |
| **Storage**  **11.0101.DA-02** | Data can be composed of multiple data elements that relate to one another. For example, population data may contain information about age, gender, and height. People make choices about how data elements are organized and where data is stored. These choices affect cost, speed, reliability, accessibility, privacy, and integrity. | | | |
| **Level 1** | **Level 2** | **Level 3** | **Level 4** |
| Store, copy, search, retrieve, modify, and delete information using a computing device and define the information stored as data. | Represent data using multiple encoding schemes. | Compare data audit techniques and approaches.  Translate between different bit representations of real-world phenomena, such as characters, numbers, and images.  Evaluate the tradeoffs in how data elements are organized and where data is stored. | Design, review and monitor optimal approach for data quality assessment  Design and implement data  cleaning approach for complex data  and projects. |
| **Visualization and Transformation**  **11.0101.DA-03** | People transform, generalize, simplify, and present large data sets in different ways to influence how other people interpret and understand the underlying information. Examples include visualization, aggregation, rearrangement, and application of mathematical operations. | | | |
| **Level 1** | **Level 2** | **Level 3** | **Level 4** |
| Identify when/why data transformations are necessary  Identify the necessity of data audit.  Define the necessity of data cleaning and potential data quality issues.  Identify in which situations data may need to be scaled  Identify outliers and understand options  to handle outliers  Understand the impact of working  with unbalanced data | Distinguish basic data audit techniques and approaches.  Infer Data quality using relevant tools and programming languages for potential issues such as missing values, duplicate data.  Summarize some basic data cleaning techniques and approaches  Clean noisy and incomplete data using relevant tools and programming languages. | Select data collection tools and techniques to generate data sets that support a claim or communicate information.  Construct a fully usable dataset  Utilize quality assessments  Conduct data cleaning of  noisy, incomplete data or data with  established data quality issues.  Justify how the interaction of multiple data issues, such as missing data, outliers, multiple values and meaning of data, impacts analysis and identifies an appropriate cleaning approach. | Create interactive data visualizations using software tools to help others better understand real-world phenomena.  Build a report documenting decisions made on the data  Deliver a presentation sharing process and results |
| **Inference and Models**  **11.0101.DA-04** | The accuracy of predictions or inferences depends upon the limitations of the computer model and the data the model is built upon. The amount, quality, and diversity of data and the features chosen can affect the quality of a model and ability to understand a system. Predictions or inferences are tested to validate models. | | | |
| **Level 1** | **Level 2** | **Level 3** | **Level 4** |
| Identify and describe patterns in data visualizations, such as charts or graphs, to make predictions. | Refine computational models based on the data they have generated.  Use data to highlight or propose cause-and-effect relationships, predict outcomes, or communicate an idea. | Evaluate the ability of models and simulations to test and support the refinement of hypotheses.  Demonstrate how to deploy a model  Demonstrate the ability to monitor model performance and to define thresholds for model re-training | Create computational models that represent the relationships among different elements of data collected from a phenomenon or process.  Apply different sampling techniques  and ways to avoid bias  Design a small project building a simple application that will use a machine learning deployed model to predict results |