

1. In the context of data science methodology, select the answer that describes what happens during the Data Requirements stage.
- ☐ Selecting the programming language for analysis
 - ☒ Identifying the necessary data content, formats, and sources
 - ☐ Defining the ingredients needed for a meal
 - ☐ Organizing data into columns and rows

1 / 1 point

✔ **Correct**
Correct! The Data Requirements stage in data science involves identifying the necessary data content, formats, and sources for the analysis. This stage is comparable to identifying the ingredients needed for a meal before cooking.

2. What technique can you apply to the data set during the data collection stage to assess data content, quality, and initial insights?
- ☐ Data extraction and merging
 - ☐ Data preparation
 - ☒ Data visualization and descriptive statistics
 - ☐ Predictive model

1 / 1 point

✔ **Correct**
Correct! You can apply data visualization and descriptive statistics to the data set during the data collection stage to assess content, quality, and initial insights about the data, similar to examining the ingredients on a cutting board.

3. Imagine you are a data scientist working on a healthcare project. Your task is to define the data requirements for a decision tree classification approach to predict patient readmission rates.
- Select the answer that describes the purpose of defining data requirements in the context of the healthcare project.
- ☐ To identify the best programming language for analysis
 - ☐ To select the appropriate patients for the cohort
 - ☒ To identify the necessary data content, formats, and source
 - ☐ To determine the total budget for the project

1 / 1 point

✔ **Correct**
Correct! Defining data requirements involves identifying the necessary data content, formats, and sources for the analysis, similar to identifying the ingredients needed for a meal.

4. You are a data scientist tasked with collecting and preparing data for a predictive modelling project related to patient outcomes. You have encountered some data sources, mainly drug information, that are missing. How would you handle this situation?
- What is a suitable approach when encountering missing data sources in the data collection process context?
- ☒ Attempt to acquire the missing data after obtaining intermediate results.
 - ☐ Proceed with incomplete data and hope for accurate results.
 - ☐ Exclude the entire patient cohort from the analysis.
 - ☐ Discard the project and start over with a different methodology.

1 / 1 point

✔ **Correct**
Correct! In the provided scenario, it is acceptable to defer decisions about unavailable data and attempt to acquire the data later, such as after obtaining intermediate results from predictive modeling.

5. Which stage of the methodology involves collaborating with DBAs and programmers to extract and merge data from various sources?
- ☐ Data Understanding
 - ☐ Data Preparation
 - ☐ Data Requirements
 - ☒ Data Collection

1 / 1 point

✔ **Correct**
Correct! Data collection involves extracting and merging data from various sources, and collaboration with DBAs and programmers is essential to ensuring data quality and efficiency.