Lesson summary

Module 1 Lesson 2: From Requirements to Collection



Congratulations! You have completed this lesson. At this point in the course, you know:

- Data Requirements stage tasks include identifying the correct and necessary data content, data formats, and data sources for the specific analytical
 approach.
- During the Data Collection stage, expert data scientists meticulously revise data requirements and make critical decisions regarding the quantity and quality of data.



Data Requirement and Collection

Data is an ingredient; Ensuring the Right Ingredients

This stage of the data science methodology emphasizes the criticality of identifying, sourcing, understanding, and preparing the necessary data for analysis.



DATA REQUIREMENTS



DATA COLLECTION



DATA INTEGRATION



DATA UNDERSTANDING



Initial data collection assessed for requirement fulfillment. Revise requirements based on availability, quality, and content 2

Requires knowing the source or where to find needed data elements, Data acquisition strategies and deferred decisions ont unavailable data.

8

DBAs / programmers extract, merge data from diverse sources. Eliminate redundant data, prep for data understanding. Automate database processes for efficiency and speed.



Use statistics and visualization for data assessment. Identify data gaps, plan missing data to be filled or substituted. Prepare data for next analysis stages.

IBM

DATA SCIENCE METHODOLOGY



• Data scientists apply descriptive statistics and visualization techniques to thoroughly assess the content, quality, and initial insights gained from the collected data, identify gaps, and determine if new data is needed or to substitute existing data.

Author(s)

<u>Dr. Pooja</u> <u>Patsy R. Kravitz</u>

Changelog

Date	Version	Changed by	Change Description
2023-08-09	0.1	Patsy R. Kravitz	Initial version created
2023-08-13	0.2	Dr. Pooja	Included infographic

© IBM Corporation 2023. All rights reserved.