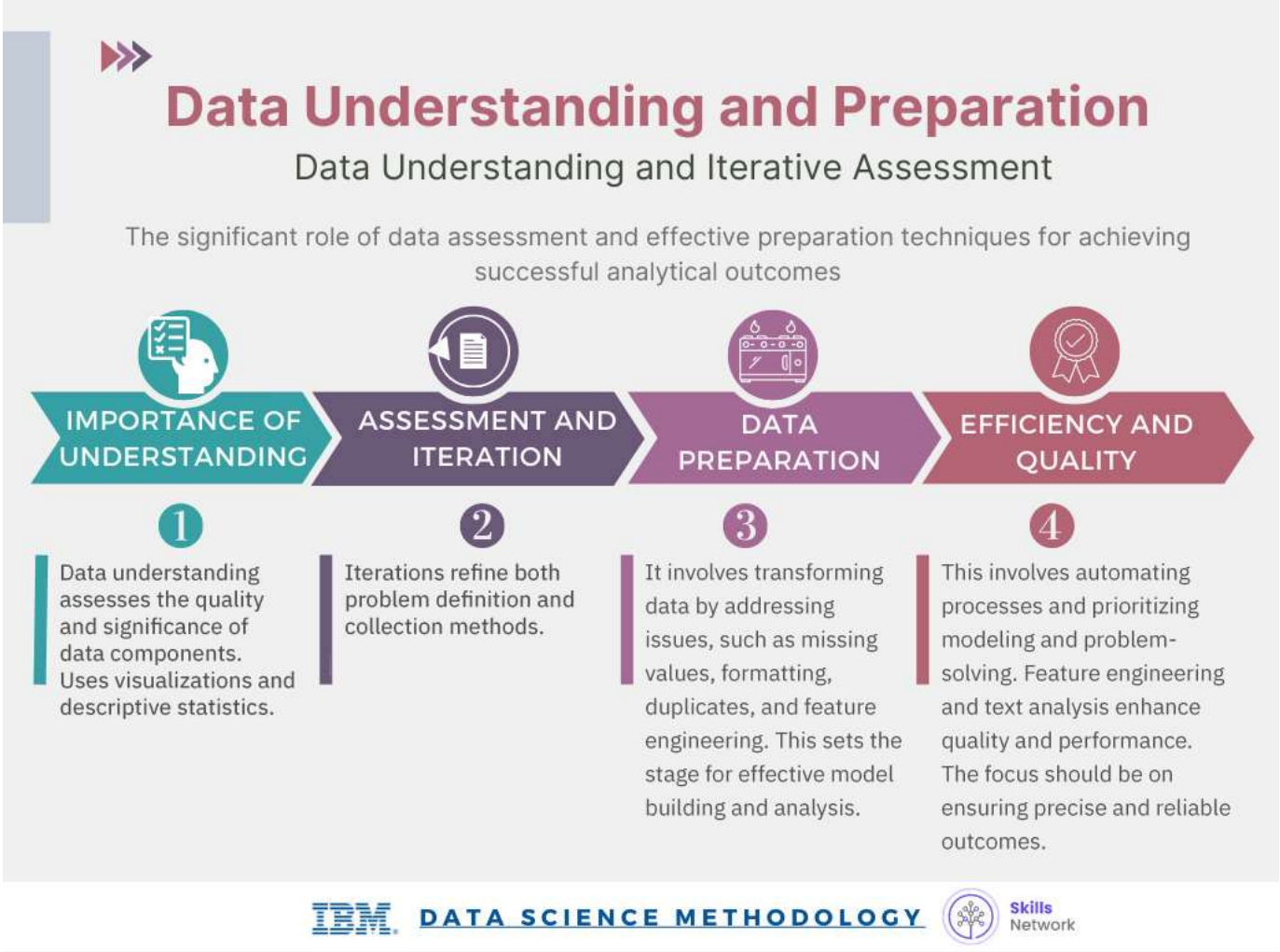
Lesson summary

Module 2 Lesson 1: From Understanding to Preparation

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

- The Data Understanding stage encompasses all activities related to constructing the data set and answers the question as to whether the data you collected represents the problem to be solved.
- During the Data Understanding stage, scientists might use descriptive statistics, predictive statistics, or both.
- Data scientists commonly apply Hurst, univariates, and other statistics on each variable, such as mean, median, minimum, maximum, standard deviation, pairwise correlation, and histograms.
- Data scientists also use univariates, statistics, and histograms to assess data quality.



- During the Data Preparation stage, data scientists must address missing or invalid values, remove duplicates, and validate that the data is properly formatted.
- Feature engineering, also part of the Data Preparation stage, uses domain knowledge of the data to create features that make the machine learning algorithms work.
- Text analysis during the Data Preparation stage is critical for validating that the proper groupings are set and that the programming is not overlooking hidden data.

Author(s)

Dr. Pooja
Patsy R. Kravitz

Skills Network