10/8/24, 4:53 PM about:blank

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.



Data Understanding and Preparation

Data Understanding and Iterative Assessment

The significant role of data assessment and effective preparation techniques for achieving successful analytical outcomes









IMPORTANCE OF UNDERSTANDING ASSESSMENT AND **ITERATION**

DATA **PREPARATION** **EFFICIENCY AND OUALITY**



Data understanding assesses the quality and significance of data components. Uses visualizations and descriptive statistics.

Iterations refine both problem definition and collection methods.

It involves transforming data by addressing issues, such as missing values, formatting, duplicates, and feature engineering. This sets the stage for effective model building and analysis.



This involves automating processes and prioritizing modeling and problemsolving. Feature engineering and text analysis enhance quality and performance. The focus should be on ensuring precise and reliable outcomes.



TEM DATA SCIENCE METHODOLOGY



- 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)

1/2 about:blank

10/8/24, 4:53 PM about:blank

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

 $\ensuremath{\text{@}}$ IBM Corporation 2023. All rights reserved.

about:blank 2/2