

Data Science Tools

olubayodeeben

ECON 5253

Measurement in Data Science

Measurement plays a critical role in data science, serving as the foundation for insights and policies. It's crucial for data scientists to understand what their dataset captures and does not capture.

R Programming

Python

Julia

Web Scraping

Data scientists should be adept at web scraping, leveraging publicly available data from webpages either through APIs or HTML parsing for their analyses.

Handling Large Data Sets With Hadoop or Spark

For data too large for a single machine, solutions like Resilient Distributed Datasets (RDDs) using Hadoop or Spark are essential for distributed data processing.

SQL

SQL is highlighted as a fundamental tool for data transformation, enabling easy data subsetting, merging, and other transformations.