

Data Analytics

1. Data Basics

- 1.1 Define the concept of data**
- 1.2 Describe basic data variable types**
 - Boolean, numeric, string
- 1.3 Describe basic structures used in data analytics**
 - Tables, rows, columns, lists
- 1.4 Describe data categories**
 - Qualitative, quantitative, metadata, big data

2. Data Manipulation

- 2.1 Import, store, and export data**
 - ETL (extract, transform and load) processes, data manipulation tools (SQL, R, Python), common data storage file formats (delimited data files, XML, JSON)
- 2.2 Clean data**
 - Purpose and common practices (handling NULL, special characters, trimming spaces, inconsistent formatting, removing duplicates, etc.); validating data
- 2.3 Organize data**
 - Purpose and common practices (sorting, filtering, slicing, transposing, appending, truncating, etc.)
- 2.4 Aggregate data**
 - Purpose and common practices (grouping, merging, summarizing, pivoting, etc.)

3. Data Analysis

- 3.1 Define and identify practices of descriptive analytics**
 - Metrics (aggregate functions such as Sum, Max, Min, Count, Avg/Mean, Mode, Median, Std Dev, Unique values), searching, filtering, interpreting results (identifying patterns and trends)
- 3.2 Define and identify practices of diagnostic analytics**
 - Data drilling, data mining (anomalies, correlation analysis, patterns, outliers, etc.), data relationships, calculating trends, interpreting results
- 3.3 Define and identify practices of hypothesis testing**
 - t-Test (student's t), p-value (test of significance)
- 3.4 Define and identify practices of predictive analytics**
 - Data relationships, using calculated trends, interpreting regression analyses, interpret results of predictive models; role of Artificial Intelligence (AI) and machine learning

IT SPECIALIST EXAM OBJECTIVES

3.5 Define and identify practices of prescriptive analytics

- Determining expected values; interpreting decision tree output; role of Artificial Intelligence (AI) and machine learning

4. Data Visualization and Communication

4.1 Report data

- Effectively display information in tables and charts; explain when and why to disaggregate data

4.2 Create visualizations from data

- Identify data visualization practices that minimize the potential for misinterpretation; identify visualization types that represent the underlying data structure and analysis questions (including comparison, time/trend, part-to-whole, relationship, distribution, correlation graphs, box and whisker diagram, scatter plot, scatter plot, bar chart, Sankey diagram, histogram, pie chart, column chart, etc.)

4.3 Derive conclusions from a data visualization

- Translate a visual representation of data into words; identify differences between claims based on an analysis and its graphical representation

5. Responsible Analytics Practices

5.1 Describe data privacy laws and best practices

- GDPR, FERPA, HIPAA, IRB, PCI, etc.

5.2 Describe best practices for responsible data handling

- Methods of handling PII, securing data, and protecting anonymity within small data sets; importance of anonymizing data; trade-offs when balancing interpretability and accuracy; shortcomings of making population-level generalizations with limited sample data

5.3 Given a scenario, describe types of bias that affect collection and interpretation of data

- Confirmation bias, human cognitive bias, motivational bias, sampling bias; selecting visualizations/data representations to avoid bias

