



Gateway Higher Institute of Science and Technology (GHIST)

Motto: Citadel of Learning and Innovation

Mile 3 Nkwen, Northwest Region, Cameroon

Tel: +237 673643469

Email: ghist.edu@gmail.com, Website: <https://ghist.org>



DATA ANALYSIS CERTIFICATION PROGRAM

Module 1: Introduction to Data Analysis

Lecture Notes Outline

1. Understanding the Data Analytic Process

Objective:

Understand the full life-cycle of data analysis, from data collection to making informed decisions based on data-driven insights.

A. What is Data Analytic?

- ✓ Definition: Data Analytic refers to the process of examining raw data to find trends, patterns, correlations, and insights that can help in making informed decisions.
- ✓ Importance: Data analytic is used to optimize processes, make predictions, and support decision-making across various industries (e.g., finance, healthcare, marketing, retail).

B. Key Phases of the Data Analytic Process

1. Data Collection:

- ✓ Gathering raw data from various sources like databases, sensors, surveys, logs, etc.
- ✓ Example: Sales data from an e-commerce site or patient records from a hospital system.

2. Data Cleaning:

- ✓ Handling missing, inconsistent, or inaccurate data to prepare it for analysis.
- ✓ Techniques include removing duplicates, filling missing values, correcting typos, etc.
- ✓ Example: Cleaning a dataset where some product sales entries have missing prices.

3. Data Exploration:

- ✓ Conducting exploratory data analysis (EDA) to understand the dataset.
- ✓ Involves summary statistics (mean, median, mode) and visualizing data with plots (e.g., histograms, scatter plots).
- ✓ Example: Plotting a histogram to observe the distribution of sales amounts.

4. Data Modeling:

- ✓ Applying statistical techniques, machine learning models, or algorithms to analyze and extract insights from data.
- ✓ Example: Predicting future sales based on historical data using regression analysis.

5. Data Interpretation:

- ✓ Analyzing the results to draw meaningful conclusions.
- ✓ Understanding what the data reveals about the business problem.
- ✓ Example: Determining that a specific product performs better during certain months.

6. Communicating Insights:

- ✓ Presenting the insights in a clear, understandable manner using visualizations, dashboards, or reports.
- ✓ Example: Creating a report with charts showing trends in product sales.

7. Decision Making:

- ✓ Using the insights derived from data to make informed business decisions.
- ✓ Example: Recommending changes in the product pricing strategy based on sales data.

C. Tools for Data Analytic:

- ✓ Excel, Google Sheets: Good for smaller datasets and basic analysis.
- ✓ Python, R: Programming languages for data manipulation and analysis.
- ✓ SQL: For querying databases.
- ✓ Tableau, Power BI: Data visualization and dashboard tools.

2. Identifying Business Problems and Framing Data Analysis Questions

Objective:

Learn how to translate a business problem into specific, actionable data analysis questions.

A. Importance of Business Context:

- ✓ Data analysis is most effective when aligned with specific business objectives.
- ✓ Understanding the problem helps define what data is needed and what techniques to apply.

B. Steps to Identify Business Problems:

1. Understand Business Goals:

- ✓ Start by identifying the main goals or objectives of the business.
- ✓ Example: A retail company wants to increase its revenue by 15% this year.

2. Identify Key Challenges or Opportunities:

- ✓ What are the roadblocks or areas for improvement?
- ✓ Example: The company is seeing declining sales in specific product categories.

3. Frame the Problem as a Question:

- ✓ Frame the business problem as a question that data analysis can address.
- ✓ Example: “Which product categories contribute most to the declining sales, and what are the underlying factors?”

4. Define the Scope:

- ✓ Be specific about what will be analyzed and the metrics involved.
- ✓ Example: “Analyze sales data from the past 12 months to determine the top underperforming categories.”

C. Types of Data Analysis Questions:

1. Descriptive Analysis:

- ✓ What happened?
- ✓ Example: How did sales fluctuate over the last year?

2. Diagnostic Analysis:

- ✓ Why did it happen?
- ✓ Example: What factors contributed to the decrease in sales for Product A?

3. Predictive Analysis:

- ✓ What is likely to happen?
- ✓ Example: Based on past data, what will the sales trend look like next quarter?

4. Prescriptive Analysis:

- ✓ What actions can be taken?
- ✓ Example: What pricing strategy can maximize future sales in underperforming categories?

D. Case Study: Framing a Data Analysis Question

- ✓ Business Problem: A company's online customer base is growing, but conversion rates (users becoming paying customers) are decreasing.
- ✓ Framed Analysis Question: “Which customer segments have the lowest conversion rates, and what factors (e.g., browsing behavior, time on site) are influencing this trend?”

3. Data Types and Structures

Objective:

Understand different types of data and how they are structured for analysis.

A. Types of Data

1. Qualitative Data (Categorical):

- ✓ Definition: Data that can be grouped into categories but not quantified.
- ✓ Examples:
 - ✧ Gender (Male/Female),
 - ✧ Product Type (Electronics, Clothing).
- ✓ Subtypes:
 - ✧ Nominal Data: No specific order (e.g., colors: red, blue, green).
 - ✧ Ordinal Data: Ordered categories (e.g., survey ratings: poor, average, good, excellent).

2. Quantitative Data (Numerical):

- ✓ Definition: Data that can be measured and expressed as numbers.
- ✧ Examples:
 - ❖ Sales amounts,
 - ❖ Temperature,
 - ❖ Age.
- ✧ Subtypes:
 - ❖ Discrete Data: Countable, whole numbers (e.g., number of products sold).
 - ❖ Continuous Data: Measured on a continuous scale (e.g., temperature, revenue).

B. Common Data Structures

1. Tabular Data:

- ✓ Organized in rows and columns, like a spreadsheet.
 - Examples:
 - ✧ Excel sheet of customer transactions,
 - ✧ SQL database table.
- ✓ Each row represents an individual record (e.g., a transaction), and each column represents an attribute (e.g., customer ID, amount spent).

2. Time Series Data:

- ✓ Data points collected or recorded at regular time intervals.
- ✓ Examples:
 - ✧ Stock prices over time,
 - ✧ Daily website traffic.

3. Hierarchical Data:

- ✓ Data organized in a tree-like structure with parent-child relationships.
- ✓ Example:
 - ✧ Organizational structure,
 - ✧ Product categories and subcategories.

4. Unstructured Data:

- ✓ Data that doesn't fit neatly into rows and columns.
- ✓ Examples:
 - ✧ Text data (e.g., customer reviews),
 - ✧ Images,
 - ✧ Videos.

C. How Data Types and Structures Affect Analysis:

- ✓ Different data types require different analysis techniques. For example:
 - ✧ Categorical data: Usually analyzed using counts, proportions, or frequencies.
 - ✧ Numerical data: Analyzed using statistical measures like mean, median, standard deviation.
 - ✧ Unstructured data: Requires specialized tools like text analysis or natural language processing (NLP) for text, or image recognition algorithms for image data.

Conclusion:

Understanding the data analytic process, framing business questions effectively, and recognizing data types and structures are the foundation skills for data analysts. These concepts guide analysts in collecting the right data, applying appropriate methods, and deriving actionable insights to solve business problems.

End of Lecture Notes: Module 1

Next Steps:

- ✓ Assignments/Practice: Frame a business problem in your industry of interest and develop data analysis questions. Identify what types of data would be required for analysis.