Data Science for Managerial Decisions (MB 511)

Instructor
Anant Prakash Awasthi

Data Science for Managerial Decisions (MB 511)

References/Literature

- Jojo Moolayil, "Smarter Decisions The Intersection of IoT and Data Science", PACKT, 2016.
- Cathy O'Neil and Rachel Schutt , "Doing Data Science", O'Reilly, 2015.
- David Dietrich, Barry Heller, Beibei Yang, "Data Science and Big data Analytics", EMC 2013
- Raj, Pethuru, "Handbook of Research on Cloud Infrastructures for Big Data Analytics", IGI Global
- Management Information System, W.S Jawadekar, Tata Mc Graw Hill Publication.
- Management Information System, David Kroenke, Tata Mc Graw Hill Publication.
- MIS Management Perspective, D.P. Goyal, Macmillan Business Books.

Data Science for Managerial Decisions (MB 511)

EMBA Program
MB-511

Online Resources

Software Resources

Data Science for Managerial Decisions (MB 511)

Program Overview

- Introduction to Data Science
- Information Technology An Overview
- Applications of Data Science in various fields
- MIS and Control Systems
- Data Collection and Data Pre-Processing
- Building Information Systems
- Support Systems for Management Decisions

EMBA Program

MB-511

MIS and Control Systems

EMBA Program
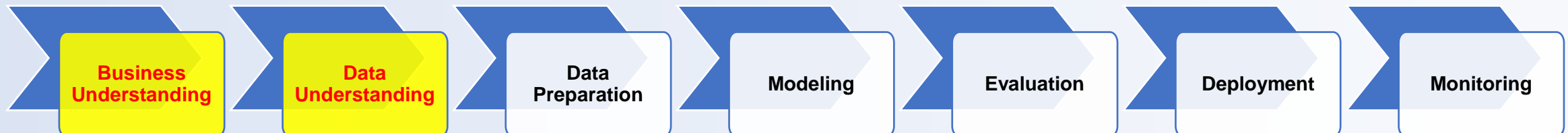MB-511

Building Data Science Solutions

Frequency Distribution

A frequency distribution is a tabular representation of data that shows the number of occurrences of different values or ranges of values within a dataset. It provides a summary of the distribution of values, making it easier to understand the overall pattern or shape of the data.

Grade	# of Students
A+	36
A	19
B+	33
B	36
C+	21
C	33
F	24

Score Range	# of Students
0–10	3
11–20	5
21–30	8
31–40	10
41–50	6

Business Understanding → Data Understanding → Data Preparation → Modeling → Evaluation → Deployment → Monitoring

Building Data Science Solutions

Frequency Distribution – Representation
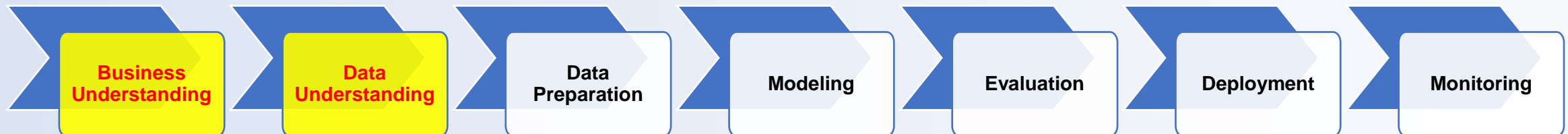
A frequency distribution is a tabular representation of data that shows the number of occurrences of different values or ranges of values within a dataset. It provides a summary of the distribution of values, making it easier to understand the overall pattern or shape of the data.

Grade	# of Students
A+	36
A	19
B+	33
B	36
C+	21
C	33
F	24

Score Range	# of Students
0–10	3
11–20	5
21–30	8
31–40	10
41–50	6

Business Understanding → Data Understanding → Data Preparation → Modeling → Evaluation → Deployment → Monitoring

EMBA Program
MB-511

Building Data Science Solutions

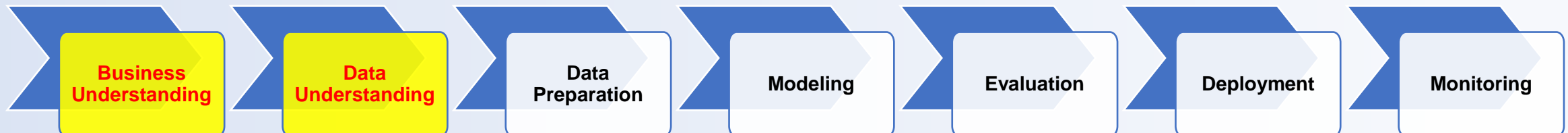Frequency Distribution – Construction of Frequency Table

how a frequency distribution is typically constructed:

- **Identify the Data:** Determine the dataset for which you want to create a frequency distribution.

- **Determine the Categories:** Decide how to group the data into categories or intervals. For numerical data, this often involves dividing the range of values into equal-sized intervals, known as bins. For categorical data, each unique category becomes its own group.

- **Count the Frequencies:** Count the number of occurrences of each value or within each category/bin.

- **Tabulate the Data:** Organize the counts into a table, with the categories or intervals listed in one column and their respective frequencies listed in another column.

- **Display the Distribution:** Optionally, you can create visual representations of the frequency distribution, such as histograms, bar charts, or frequency polygons, to provide a more intuitive understanding of the data.

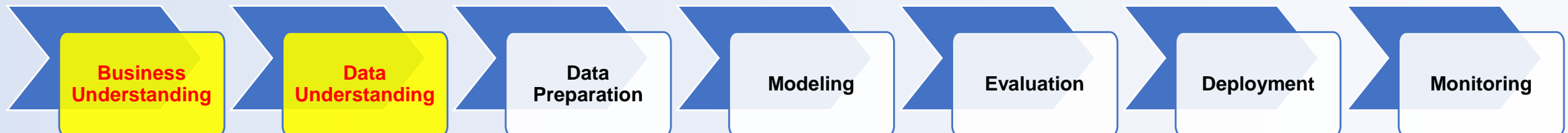| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

Building Data Science Solutions

A one-variable frequency table, also known as a univariate frequency table, is a type of frequency table that summarizes the distribution of a single variable. It displays the counts or frequencies of each unique value or category within the variable.

One-variable frequency tables are useful for gaining insight into the distribution and patterns of a single variable. They are often used in exploratory data analysis to understand the central tendency, variability, and shape of the data. Additionally, they serve as a basis for creating visual representations like histograms or bar charts to visualize the distribution more effectively.

Grade	# of Students
A+	36
A	19
B+	33
B	36
C+	21
C	33
F	24

Score Range	# of Students
0-10	3
11-20	5
21-30	8
31-40	10
41-50	6

EMBA Program
MB-511

Business Understanding → Data Understanding → Data Preparation → Modeling → Evaluation → Deployment → Monitoring

Building Data Science Solutions

Frequency Distribution – Types of Frequency Tables

Cross tabulation, also known as a contingency table or a two-way frequency table, is a statistical method used to summarize and analyze the relationship between two categorical variables. It displays the frequency counts of observations that fall into various categories of both variables.

EMBA Program
MB-511

	Science	Art	Commerce
A	24	37	23
B	30	30	42
C	32	24	36

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

EMBA Program
MB-511

Activity

Building Frequency Table
- **Discreate Data**
- **Continuous Data**

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

Building Data Science Solutions
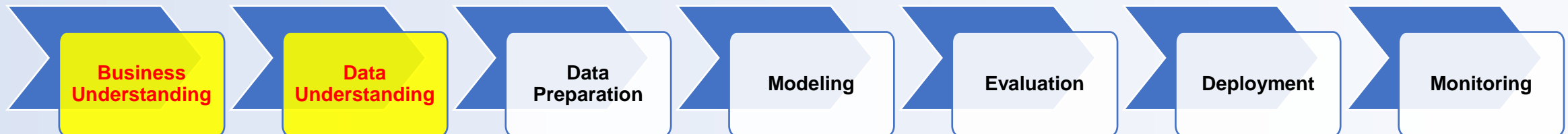
EMBA Program
MB-511

Measures of central tendency are statistical measures that describe the center or average of a dataset. These measures provide insights into the typical or central value around which the data tend to cluster. The three main measures of central tendency are the mean, median, and mode.

These measures of central tendency serve different purposes and may be used in different contexts depending on the characteristics of the data and the objectives of the analysis. The choice of which measure to use depends on factors such as the distribution of the data, the presence of outliers, and the research question being addressed.

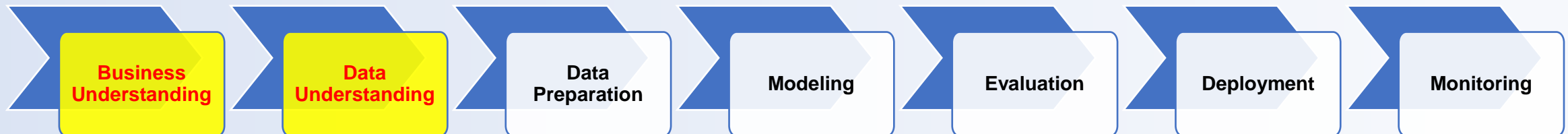| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

Building Data Science Solutions

EMBA Program
MB-511

Selecting the right measure of central tendency depends on various factors, including the characteristics of the data and the objectives of the analysis.

- **Consider the Data Type:** Determine whether your data is numerical or categorical. The mean and median are suitable for numerical data, while the mode can be used for both numerical and categorical data.

- **Assess the Data Distribution:** Examine the distribution of your data. If the data are symmetrically distributed without significant outliers, the mean may be an appropriate measure. If the data are skewed or contain outliers, consider using the median, which is more robust to extreme values.
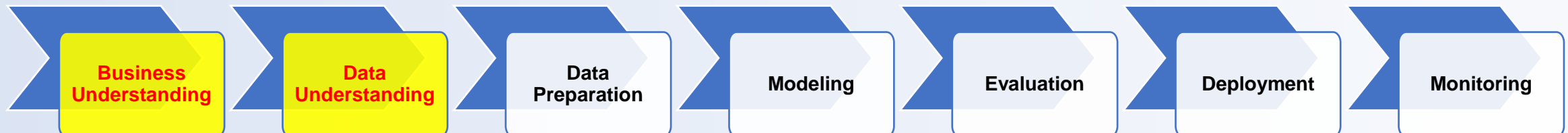
| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

Building Data Science Solutions

Measure of Central Tendency – Selecting the right one

- **Identify the Level of Measurement:** Consider the level of measurement of your data. The mean is suitable for interval or ratio data, while the median can be used for ordinal, interval, or ratio data. The mode can be applied to nominal, ordinal, or discrete data.

- **Understand the Research Question:** Consider the research question or objectives of your analysis. Determine which measure of central tendency best aligns with the information you want to convey or the insights you want to gain from the data.

- **Account for Data Variability:** Assess the variability and dispersion of your data. If your data exhibit high variability or if extreme values are present, the median may provide a better representation of the central tendency compared to the mean.

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

Activity

Computing Measure of Central tendency
- Discreate Data
- Continuous Data

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |
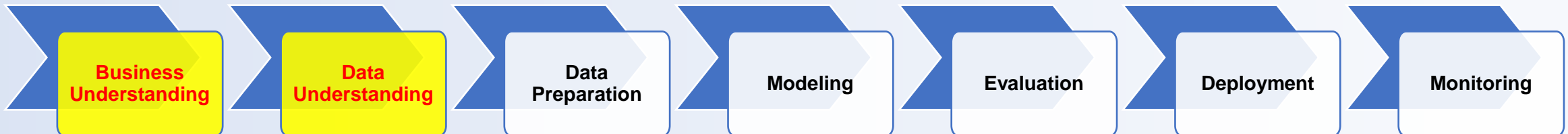
Data is not just a piece of information, but a powerful tool that can transform businesses and

drive innovation." – Unknown

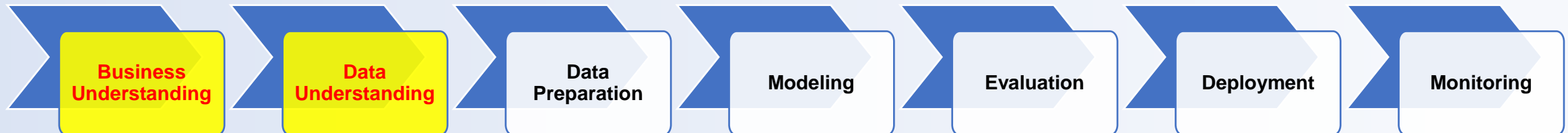| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |
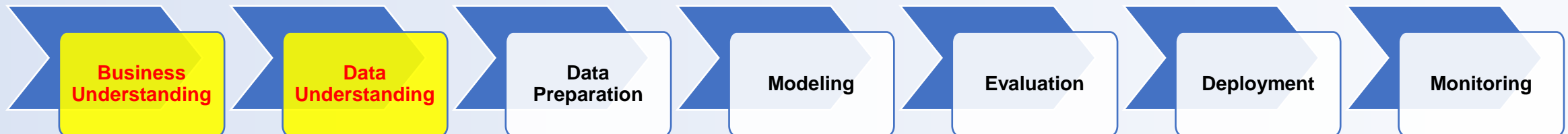
Building Data Science Solutions

Data Management in Information System

Agenda:

- Role of Data Management in Information System
- Terminologies
- Hands on Sessions
  - Data Import
  - Data Export
  - Sub-setting data (selecting rows and columns)
  - Data types conversion in Pandas
  - Operators and New Variable Creation
  - Functions with Pandas
  - User defined functions in Pandas
  - Control structure in Pandas
  - Sorting data
  - Merging data
  - Reshaping data
  - Data Aggregation and Pivoting

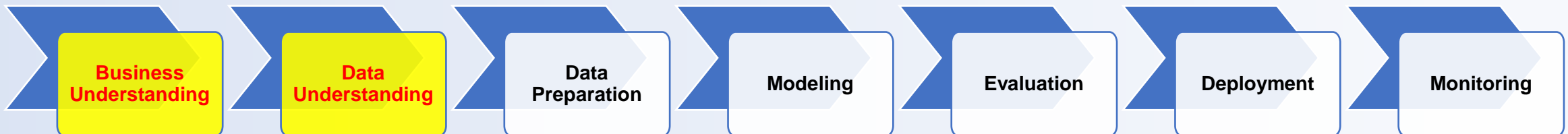| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

EMBA Program
MB-511

Data Management and Manager – Objective

Right Data Points are available to right person at the right time

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

Building Data Science Solutions

The role of data management in an information system is crucial, as it encompasses the processes, policies, technologies, and practices used to acquire, store, organize, secure, and manage data throughout its lifecycle within the system.
Effective data management is essential for ensuring the reliability, security, integrity, and usability of data within an information system, thereby supporting the organization's objectives and enabling informed decision-making and innovation.
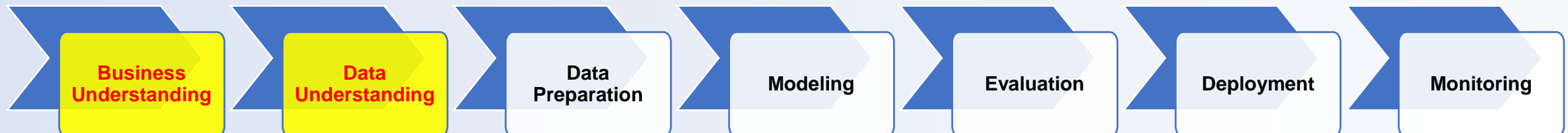Here's an overview of the key roles data management plays in an information system:

1. Data Acquisition:
- Data management involves the acquisition of data from various sources, both internal and external to the organization.
- It ensures that data is collected in a structured and consistent manner to maintain its quality and integrity.

2. Data Storage and Organization:
- Data management oversees the storage and organization of data within the information system.
- It determines the appropriate storage mechanisms and structures (databases, data warehouses, data lakes) to efficiently store and retrieve data.

EMBA Program
MB-511

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

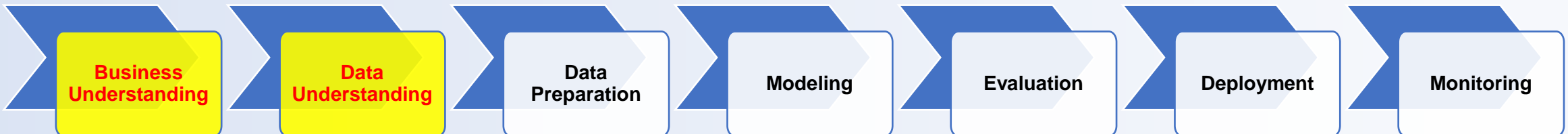**3. Data Quality and Integrity:**
- Data management is responsible for maintaining data quality and integrity throughout its lifecycle.
- It involves processes such as data cleansing, deduplication, validation, and normalization to ensure that data is accurate, consistent, and reliable.

4. Data Security and Privacy:
- Data management ensures the security and privacy of data within the information system.
- It implements measures such as access controls, encryption, authentication, and auditing to protect data from unauthorized access, breaches, and misuse.

5. Data Integration and Interoperability:
- Data management facilitates the integration of data from disparate sources and systems within the information system.
- It ensures that data is standardized and harmonized to enable seamless interoperability and data exchange between different components of the system.

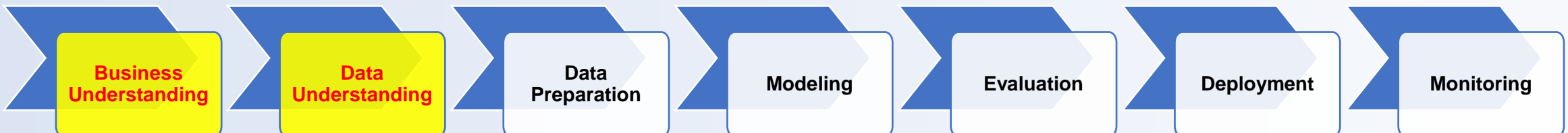| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

6. Data Governance and Compliance:
- Data management establishes policies, standards, and procedures for governing data within the information system.
- It ensures compliance with regulatory requirements, industry standards, and organizational policies related to data usage, privacy, security, and ethical considerations.

7. Data Analysis and Decision Making:
- Data management enables data analysis and decision-making processes within the information system.
- It provides access to quality data and tools for analyzing, querying, and visualizing data to derive insights, make informed decisions, and drive business outcomes.

8. Data Lifecycle Management:
- Data management oversees the entire lifecycle of data within the information system, from creation to deletion.
- It involves managing data storage, usage, archiving, and disposal in accordance with organizational policies and legal requirements.

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

EMBA Program
MB-511

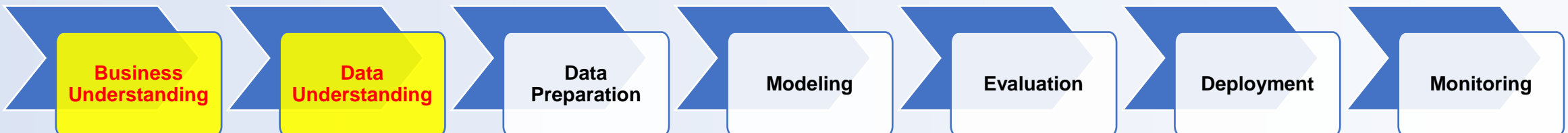**6. Data Governance and Compliance:**
- Data management establishes policies, standards, and procedures for governing data within the information system.
- It ensures compliance with regulatory requirements, industry standards, and organizational policies related to data usage, privacy, security, and ethical considerations.

7. Data Analysis and Decision Making:
- Data management enables data analysis and decision-making processes within the information system.
- It provides access to quality data and tools for analyzing, querying, and visualizing data to derive insights, make informed decisions, and drive business outcomes.

8. Data Lifecycle Management:
- Data management oversees the entire lifecycle of data within the information system, from creation to deletion.
- It involves managing data storage, usage, archiving, and disposal in accordance with organizational policies and legal requirements.

| Business Understanding | Data Understanding | Data Preparation | Modeling | Evaluation | Deployment | Monitoring |

Have a question?
Feel Free to Reach out at
- **+91-88846-52929** (WhatsApp)
- **anant.awasthi@outlook.com** (E-Mail)