DATABASE MANAGEMENT SYSTEM

Chapter 11
Data Analysis

Bikash Khadka Shah

MCSE, MSDA, OCP, RHCE, RHCVA, CCNA, CEH

9841766620 | 9801076620

Data analysis

Data analysis is the process of cleaning, changing, and processing raw data, and extracting actionable, relevant information that helps businesses make informed decisions.

A simple example of data analysis can be seen whenever we take a decision in our daily lives by evaluating what has happened in the past or what will happen if we make that decision.

Basically, this is the process of analyzing the past or future and making a decision based on that analysis.

Why is Data Analysis Important?

Better Customer Targeting:

- You don't want to waste your business's precious time, resources, and money putting together advertising campaigns targeted at demographic groups that have little to no interest in the goods and services you offer.
- Data analysis helps you see where you should be focusing your advertising efforts.

You Will Know Your Target Customers Better:

- Data analysis tracks how well your products and campaigns are performing within your target demographic.
- Through data analysis, your business can get a better idea of your target audience's spending habits, disposable income, and most likely areas of interest

Reduce Operational Costs:

- Data analysis shows you which areas in your business need more resources and money, and which areas are not producing and thus should be scaled back or eliminated outright.

Better Problem-Solving Methods:

- Informed decisions are more likely to be successful decisions. Data provides businesses with information. You can see where this progression is leading.
- Data analysis helps businesses make the right choices and avoid costly pitfalls.

You Get More Accurate Data:

- If you want to make informed decisions, you need data, but there's more to it. The data in question must be accurate.
- Data analysis helps businesses acquire relevant, accurate information, suitable for developing future marketing strategies, business plans, and realigning the company's vision or mission.

The Data Analysis Process

1. Defining the question

The first step in any data analysis process is to define your objective. In data analytics jargon, this is sometimes called the 'problem statement'.

Defining your objective means coming up with a hypothesis and figuring how to test it. Start by asking: What business problem am I trying to solve?

For instance, your organization's senior management might pose an issue, such as: "Why are we losing customers?"

A data analyst's job is to understand the business and its goals in enough depth that they can frame the problem the right way.

Data Requirement Gathering:

Ask yourself why you're doing this analysis, what type of data analysis you want to use, and what data you are planning on analyzing.

2. Step two: Collecting the data

Once you've established your objective, you'll need to create a strategy for collecting and aggregating the appropriate data. A key part of this is determining which data you need.

This might be quantitative (numeric) data, e.g. sales figures, or qualitative (descriptive) data, such as customer reviews. All data fit into one of three categories: first-party, second-party, and third-party data.

Data Collection:

Guided by the requirements you've identified, it's time to collect the data from your sources.

3. Step three: Cleaning the data

- Once you've collected your data, the next step is to get it ready for analysis. This means cleaning, or 'scrubbing' it, and is crucial in making sure that you're working with high-quality data. Key data cleaning tasks include:

- Removing major errors, duplicates, and outliers—all of which are inevitable problems when aggregating data from numerous sources.

- Removing unwanted data points—extracting irrelevant observations that have no bearing on your intended analysis.

- Bringing structure to your data—general 'housekeeping', i.e. fixing typos or layout issues, which will help you map and manipulate your data more easily.

- Filling in major gaps—as you're tidying up, you might notice that important data are missing. Once you've identified gaps, you can go about filling them.

- Data Cleaning:

- Not all of the data you collect will be useful, so it's time to clean it up. This process is where you remove white spaces, duplicate records, and basic errors.

Step four: Analyzing the data

- Finally, you've cleaned your data.

- The type of data analysis you carry out largely depends on what your goal is. But there are many techniques available.

- Univariate or bivariate analysis, time-series analysis, and regression analysis are just a few you might have heard of. More important than the different types, though, is how you apply them.

- Data Analysis: Here is where you use data analysis software and other tools to help you interpret and understand the data and arrive at conclusions.

5. Step five: Sharing your results

This is more complex than simply sharing the raw results of your work—it involves interpreting the outcomes, and presenting them.

Data analysts commonly use reports, dashboards, and interactive visualizations to support their findings.

There are lot of data visualization tools available, suited to different experience levels. Popular tools requiring little or no coding skills include Google Charts, Tableau, Data wrapper, and Infogram.

Data Visualization: Data visualization is a fancy way of saying, "graphically show your information in a way that people can read and understand it."

6. Summary

- **Define the question—What business problem are you trying to solve? Frame it as a question to help you focus on finding a clear answer.**

- **Collect data—Create a strategy for collecting data. Which data sources are most likely to help you solve your business problem?**

- **Clean the data—Explore, scrub, tidy, de-dupe, and structure your data as needed. Do whatever you have to! But don't rush...take your time!**

- **Analyze the data—Carry out various analyses to obtain insights. Focus on the four types of data analysis: descriptive, diagnostic, predictive, and prescriptive.**

- **Share your results—How best can you share your insights and recommendations? A combination of visualization tools and communication is key.**

- **Data Interpretation: Now that you have your results, you need to interpret them and come up with the best courses of action, based on your findings.**

Types of Data Analysis

Diagnostic Analysis:

- Diagnostic analysis answers the question, "Why did this happen?" Using insights gained from statistical analysis (more on that later!), analysts use diagnostic analysis to identify patterns in data. Ideally, the analysts find similar patterns that existed in the past, and consequently, use those solutions to resolve the present challenges hopefully.

●●●

Bikash Khadka Shah - MCSE, MSDA, OCP, RHCE, RHCVA, CCNA, CEH   9841766620 | 9801076620

Types of Data Analysis

Predictive Analysis:

- Predictive analysis answers the question, "What is most likely to happen?" By using patterns found in older data as well as current events, analysts predict future events.
- While there's no such thing as 100 percent accurate forecasting, the odds improve if the analysts have plenty of detailed information and the discipline to research it thoroughly.

Types of Data Analysis

Prescriptive Analysis:

- Mix all the insights gained from the other data analysis types, and you have prescriptive analysis.
- Sometimes, an issue can't be solved solely with one analysis type, and instead requires multiple insights.

•••

Types of Data Analysis

Statistical Analysis:

- Statistical analysis answers the question, "What happened?" This analysis covers data collection, analysis, modeling, interpretation, and presentation using dashboards.

Types of Data Analysis

Qualitative Data Analysis: The qualitative data analysis method derives data via words, symbols, pictures, and observations. This method doesn't use statistics. The most common qualitative methods include:

- **Content Analysis**, for analyzing behavioral and verbal data.
- **Narrative Analysis,** for working with data culled from interviews, diaries, surveys.

● ● ●

Steps involved in Data Analysis

Before Data Collection

- Determine the method of data analysis
- Determine how to process the data
- Consult a statistician
- Prepare dummy tables

After Data Collection

- Process the data
- Prepare tables and graphs
- Analyze and interpret findings
- Consult again the statistician
- Prepare for editing
- Prepare for presentation

Data Analysis Methods

- **Data analysis Research Methods**
  - **Qualitative data analysis**
  - **Quantitative data analysis**
- **Statistical Methods used for Analysis**
  - **Descriptive Statistics**
  - **Inferential Statistics**

Qualitative Data Analysis

- **Qualitative Data Analysis (QDA)**
  - **Range of processes**
  - **Procedures**
  - **Move from the qualitative data**
- **QDA is usually based on an interpretative philosophy.**
- **The idea is**
  - **To examine**
    - **The meaningful content**
    - **Symbolic content**

Qualitative Data Analysis

Data Collection

Observation
Interviews
Document
Recording

Data Reduction

Selection
Focusing
Simplifying
Abstracting
Transforming

Data Display

Matrices
Graphs
Networks
Charts

Conclusion Drawing

Give Meanings
Confirming
Verifying
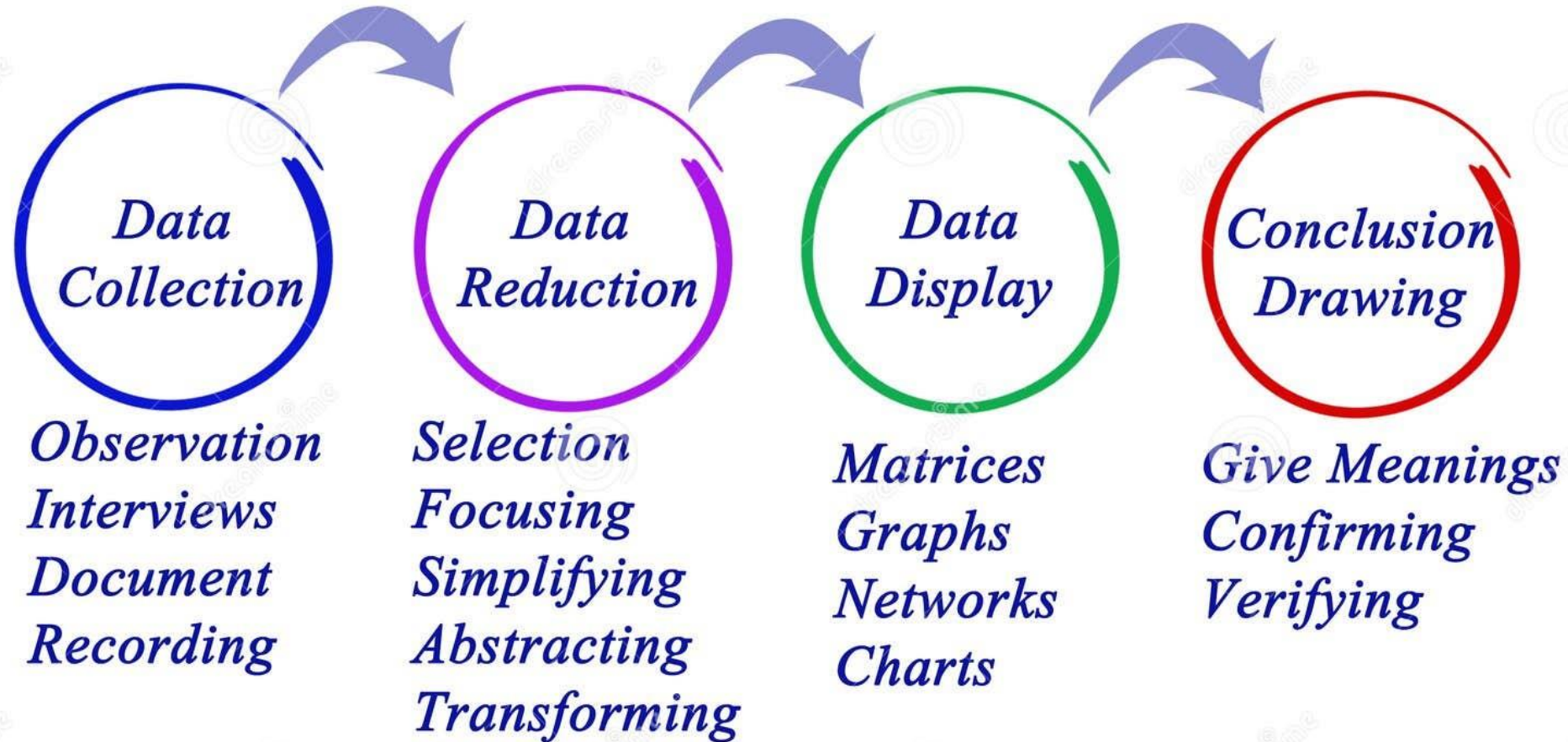
Quantitative Data Analysis
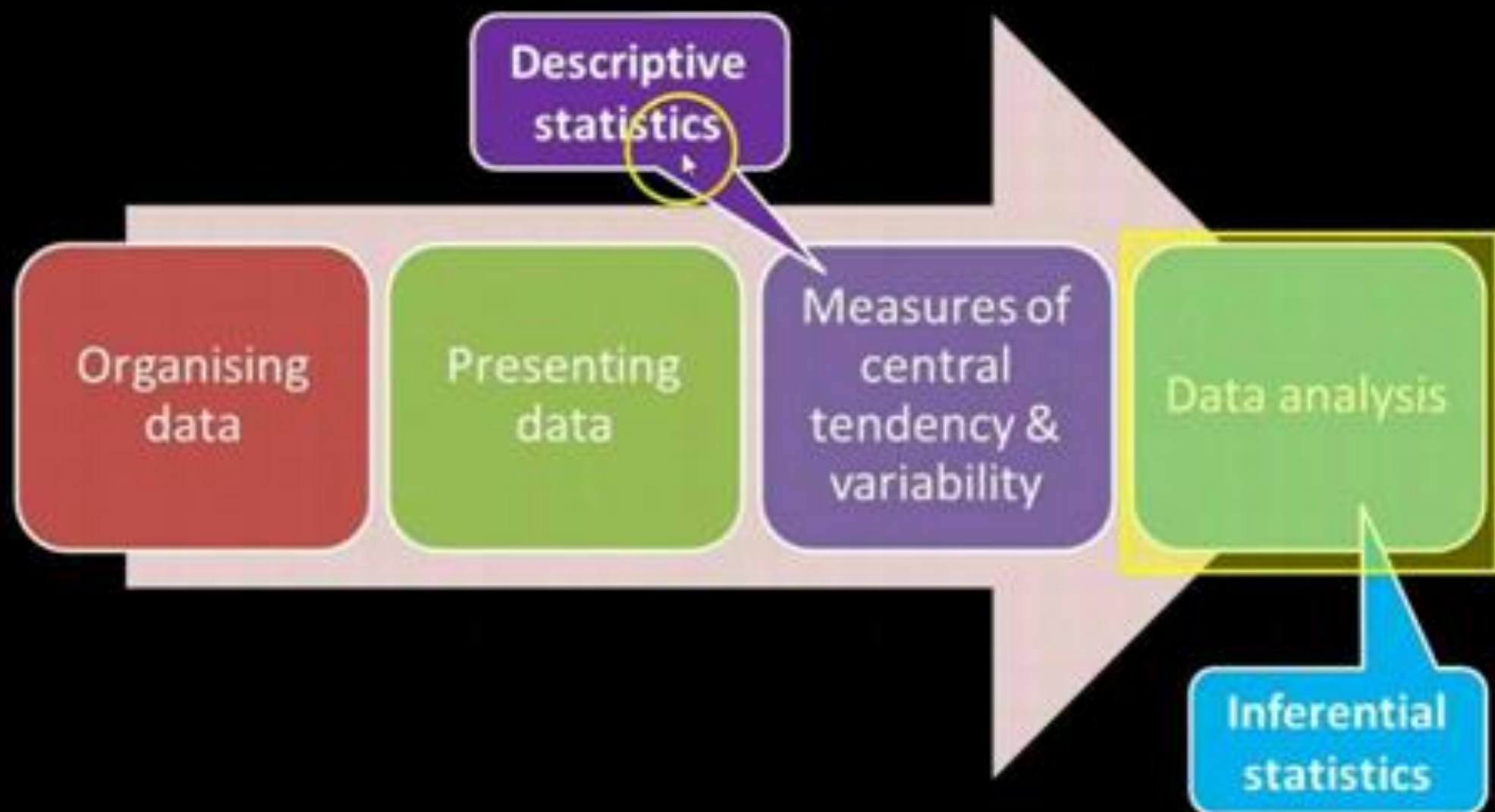
- **Quantitative data analysis**
  - **You are expected to turn raw numbers into meaningful data**
    - **Through the application of rational and critical thinking.**

- **Quantitative data analysis may include**
  - **Calculation of frequencies of variables**
  - **Differences between variables**

- **Quantitative approach is usually associated**
  - **Finding evidence**
  - **Hypotheses**

LIFE CYCLE OF A DATA ANALYSIS PROJECT

Based on CRISP-DM Methodology

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6

Business Issue Understanding

Define business objectives

Gather required information

Determine appropriate analysis method

Clarify scope of work

Identify deliverables

Data Understanding

Collect initial data

Identify data requirements

Determine data availability

Explore data and characteristics

Data Preparation

Gather data from multiple sources

Cleanse

Format

Blend

Sample

Exploratory Analysis and Modeling

Develop methodology

Determine important variables

Build model

Assess model

Validation

Evaluate results

Review process

Determine next steps

Results are valid proceed to step 6 →

Results are invalid revisit steps 1-4 ←

Visualization and Presentation

Communicate results

Determine best method to present insights based on analysis and audience

Craft a compelling story

Make recommendations

Bikash Khadka Shah - MCSE, MSDA, OCP, RHCE, RHCVA, CCNA, CEH   9841766620 | 9801076620