Because learning changes everything.®

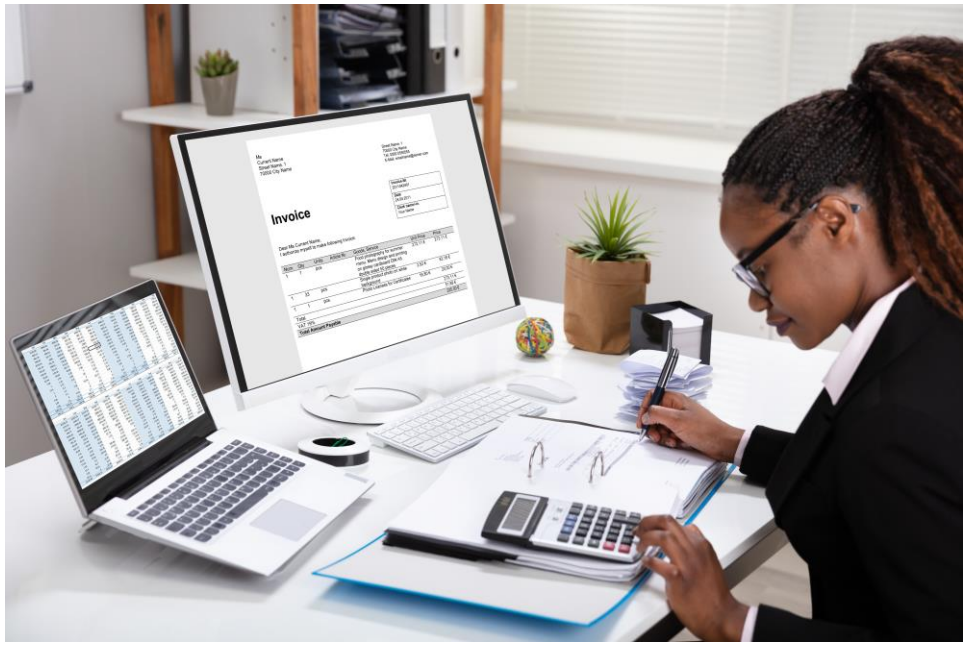
MIS

Data Analytics and Data Ecosystems

Overview of Data Analytics and Data Ecosystems

Data analytics.

- The process of investigating raw data of various types to:
  - Uncover trends and correlations.
  - To answer specifically crafted questions.
- The process of data analytics includes.
  - Descriptive analytics.
  - Predictive analytics.
  - Prescriptive analytics.



andreypopov/123RF

Descriptive Analytics

Uncovers historical trends in data sets.

Can be thought of as trying to answer the questions:

- “what happened?” or
- “what is occurring?”

Examples:

- Return on Investment (ROI).
- Summaries of past events such as sales, operational efficiency, impact of sales and marketing.

Predictive Analytics

Focuses on understanding, predicting, and planning for future events and business outcomes.

Utilize probability analysis techniques as well as data mining, statistical modeling, machine learning, and deep learning to generate possible future outcomes given certain conditions.

Can be used in a variety of business areas including:

- E-commerce.
- Cybersecurity.
- Information technology.
- Healthcare.

Prescriptive Analytics

Used to help determine best courses of action.

Considered to be the most advanced form of data analytics.

Seek to predict what, when, and why a given scenario might occur.

Examples:

- Tracking fuel prices by airlines to determine possible increases and decreases.
- Monitoring flu strains and activity to determine possible outbreak areas.
- Analysis of cyberthreats and activity to identify possible threats and breaches.

Trends in Data Analytics

Smarter, scalable artificial intelligence (AI).

- Allows for better learning algorithms.
- Allows for shorter development times.

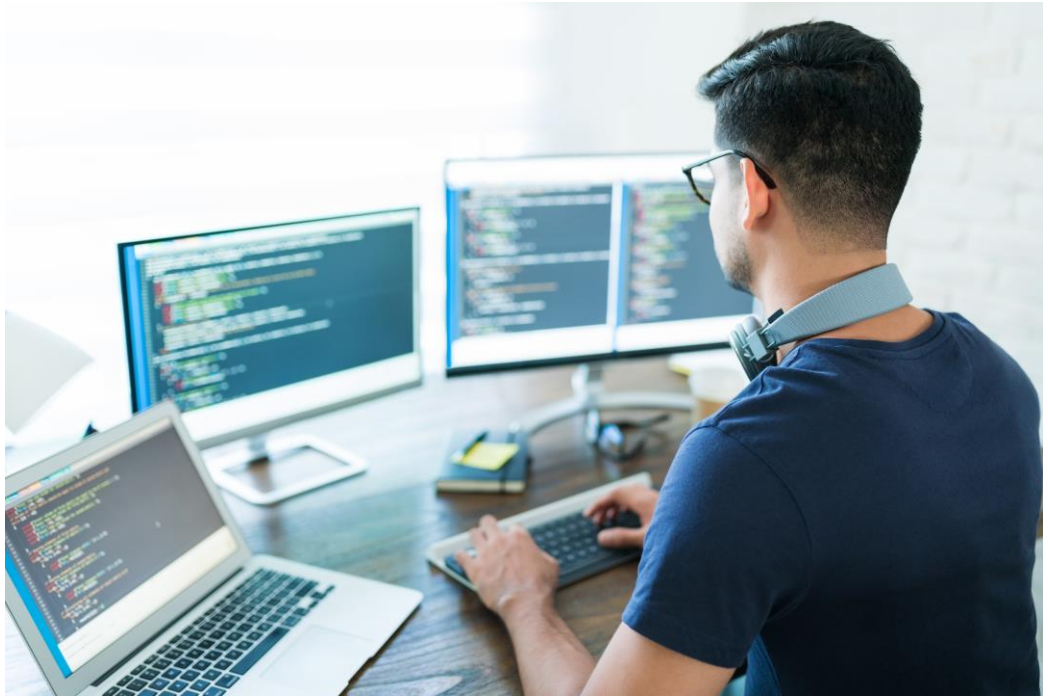
Composable data and analytics.

- Creates a more user-friendly experience.
- Makes finding critical insights easier.
- Enables collaboration and insights.

Data fabric as a foundation.

- A set of data services that spans on-premise and Cloud environments.

Data analytics as a core business function.



antoniodiaz/123RF

Uses of Data Analytics in Business

- Improved and effective decision making.
- Improved customer service.
- Increased efficiency of production and operation processes.
- Improved patient experiences and internal procedures in healthcare.

Data Ecosystems

- The information technology (IT) architecture and infrastructure, software applications, programming languages, and storage technologies used for the collection, storage, analysis, and interpretation of meaningful data.
- Based on the unique circumstances of each organization, a specialized ecosystem is created for every business entity.

5 Steps of Data Project Life Cycle

- Sensing - Meaningful data that should be collected by the organization.
- Collecting - Gathering of data from determined and validated data sources.
- Wrangling - Conversion of raw data into a user-friendly format.
- Analysis - Examination of data.
- Storage - Involves hardware, software, and procedures for securing, maintaining, and accessing data.

Skills and Characteristics of Data Analysts

Analytical thinking and skills

Include characteristics and capabilities that allow for.

- Observation.
- Research.
- Interpretation of a problem.

Can be applied for the investigation of:

- Supply chain issues.
- Customer relationship management.
- Product and service improvements.

Five Aspects of Analytical Skills

- Curiosity – involves investigation into learning more about a problem or change of a desired state as well as knowing the right questions to ask to uncover issues.
- Understanding context – involves understanding where data and information fit into the plan and approach.
- Having a technical mindset – involves the ability to break down processes and information into smaller digestible/analytical steps.
- Data design – involves the ability to conceptualize of how data and information should be organized.
- Data strategy – involves the ability to analyze the people, processes, hardware, and software used in data analysis.

Responsibilities of a Data Analyst

Data Analyst

Uses the knowledge or processing software, business strategy, and analytical skill to deliver data and reports that guide management as they create well-informed decision making

Duties include:

- Working on data analytics teams to extract data from large data sets.
- Creating reports that outline key findings.
- Monitoring KPIs to identify success or failure.
- Analyzing data to identify trends.

Syda Productions/Shutterstock

Additional Responsibilities

- Collaboration with executives and other stakeholders to uncover areas for improvement.
- Data visualization to aid in the interpretation of data.
- Structuring large data sets to ensure data is accessible and usable.
- Creating reports and presentations for management and executives that outline key findings and recommendations.

metamorworks/Shutterstock

Key Skills of Data Analysts

- Technical writing skills.
- Experience with computer code including SQL, Python, and Oracle.
- Strong analytical and problem-solving skills.
- Experience with data visualization software including Tableau and Power BI.
- Microsoft Excel and spreadsheet experience.
- Effective time management and the ability to multitask and to meet deadlines.
- Oral communication and presentation software skills.

Dragon Images/Shutterstock

Employment

Data analysts are employed in:

- Finance.
- E-Commerce.
- Healthcare.
- Government.
- Science.

Andriy Popov/123RF

U.S. Bureau of Labor Statistics for Data Analysts

Entry-level education bachelor's degree in:

- Management information systems.
- Computer science.
- Mathematics or statistics.
- Economics or finance.

Median salary.

- Across different disciplines is $88,770.

Job outlook.

- 15% growth (much faster than average).

How Data Analysts Define Success for a Project

Formula to define success:

Data

+

Organizational and business knowledge

=

problem solved

NicoElNino/Shutterstock

Solve the Problem

Data must be available.

Data must be accessible.

Data analysts must be knowledgeable about:

- The organization.
- The problem/question they want to solve.

adam Sternin/Cavan Images/Image Source

Questions Data Analysts Can Ask

- What is the overall outcome and results that are needed?
- Who is the receiver of the information and analysis?
- What is the question that is being asked? Am I answering the question?
- What are the time constraints for the project? How quickly must a decision be made or information produced?

Data-Driven Decision Making

- What is data driven decision making?
- What are the steps involved in decision making?
- What are the tools used to make decisions?

What is Data-Driven Decision Making?

According to Tableau, **data-driven decision making** (DDDM) is the following:

- The use of facts, metrics, and data to guide strategic business decisions that align with organizational goals, objectives, and initiatives.

Organizations are investing in the development of three data-driven competencies:

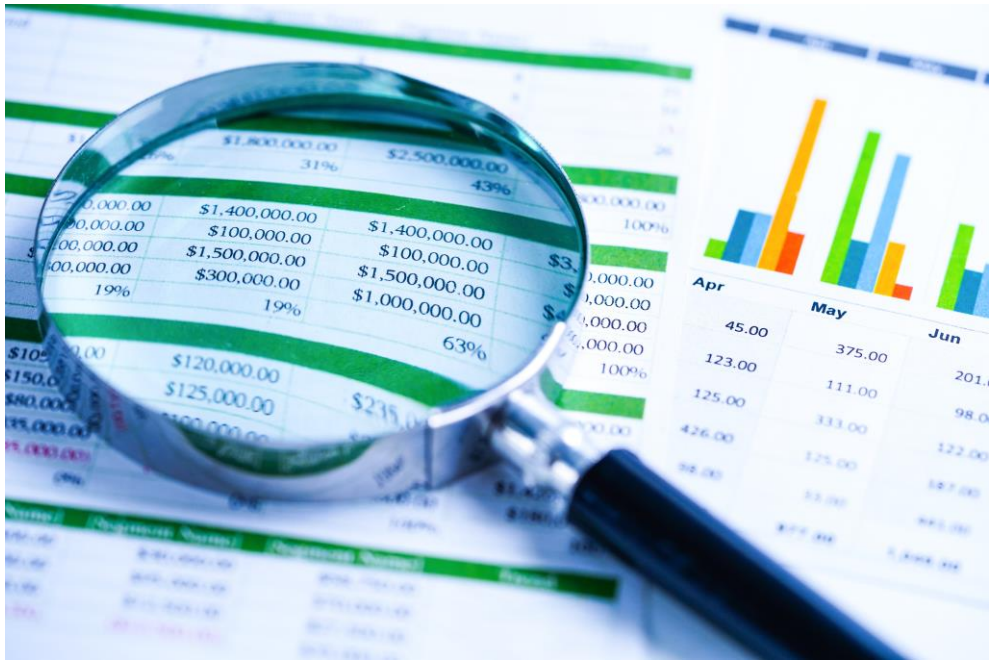
- Data proficiency.
- Agility in analytics and data analysis.
- Building of a data-driven culture and community.

Six Steps to Create Data-Driven Decisions

1. Ask – Determine what questions you need to ask and establish a clear definition of the problem
2. Prepare – Data should be collected, stored, and secured in preparation for data processing
3. Process – Data cleansing and checking should occur to ensure high-quality data is ready for analysis
4. Analysis – Data is analyzed to find patterns, trends, and relationships within the data set
5. Share – Data is shared with the appropriate audience. This can include visualization and presentations to key stakeholders
6. Act – Once data and results have been shared, decisions must be made.

Three Key Data Analyst Tools [1]

- Spreadsheets.
- Databases and query languages.
- Data visualization tools.



chormail/123RF

Three Key Data Analyst Tools 2

Spreadsheets.

- The two most popular spreadsheet programs are Microsoft Excel and Google Sheets.
- Cloud-based and desktop versions of Excel, Sheets, and Numbers are widely used today.
- Allow for the collection, storage, organization, and manipulation of data.
- Users can identify patterns in data as well as create data visualization tools with charts and graphs.

Three Key Data Analyst Tools 3

Databases and query languages.

- Allow for the collection, storage, organization, and manipulation of data.
- Users can extract and query data which cannot be done using spreadsheet software.
- Databases use structured query language (SQL) to ask questions (or query) the data.
- SQL allows users to isolate specific information contained in a database.
- SQL helps in the selection, creation, addition, and extraction of data from a database.
- Popular SQL programs include MYSQL, Microsoft SQL, and Oracle SQL.

Three Key Data Analyst Tools 4

Data visualization tools.

- The graphical and structured representation of data.
- Visualization makes it easier to see the results and conclusions ascertained from data analysis.
- Allow for the creation of charts, tables, graphs, and maps.
- Popular data visualization tools include Tableau, Microsoft Power BI, and Looker.

Spreadsheets and Databases 1

The type of job and problem to be solved leads to the choice of tool to be used for data analysis:

- If complex and attention-grabbing graphics are needed, a data visualization tool might be the best choice.
- If data needs to be organized, cleansed, and analyzed, then spreadsheet or database software might be the best choice.

Spreadsheets and databases are compared and contrasted in the following table:

Spreadsheets and Databases 2

Spreadsheets	Databases
Software applications that rely on hardware to run and store information	Software applications that rely on hardware to run and store information
Data is structured in rows and columns	Data is structured using rules and relationships defined by the DBMS
Information is organized in cells	Data is organized in complex and meaningful collections specified by the database administrator. These are referred to as fields.
Often requires manual data entry that is not constrained	Data entry is often automated and constrained by defined parameters
Frequently restricted to one user at a time, although Cloud-based spreadsheet programs allow for multiple users	Allows for multiple users to be working at the same time

Using SQL to Communicate with a Database 1

Characters.

- Includes letters, numbers, and symbols that compose a field.

Fields.

- A set of data values made up of characters.
- Also referred to as data attributes.
- Are commonly found in columns.
- Examples in a student information table:
  - first name, last name, student ID number, address, email, phone.

Wright Studio/Shutterstock

Using SQL to Communicate with a Database 2

Records.

- Collections of related fields.
- Are often structured in rows of data.
- Example in a student information table:
  - The structured information on a specific student who is enrolled at the college.

Using SQL to Communicate with a Database 3

Tables.

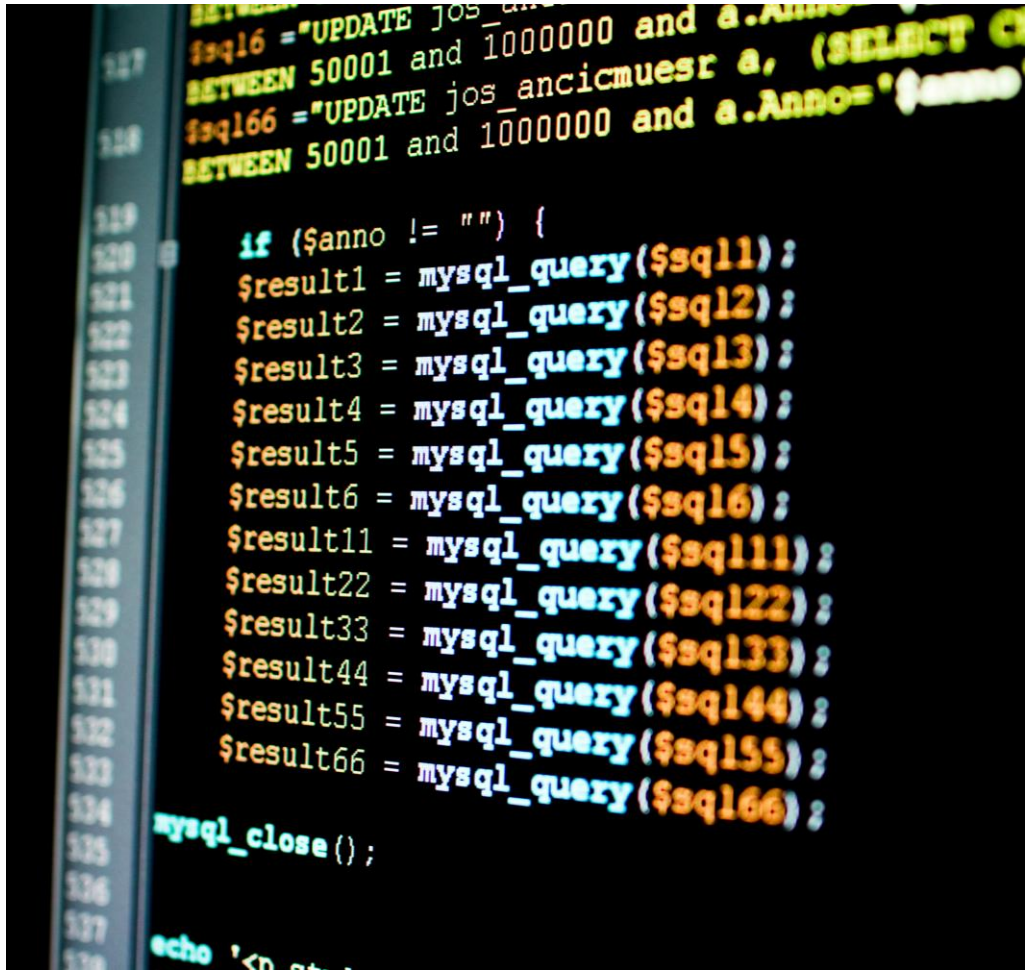
- Groups of assigned rows and columns that contain related data.
- Most databases contain multiple tables.
- Examples are tables for:
  - Student data.
  - Faculty data.
  - Financial aid data.
  - Staff data.
  - Course offerings.

SQL (Structured Query Language)

- A popular programming language used to communicate with databases.
- Pronounced sequel or SQL.
- Useful when working with large data sets.
- Is widely used in today's business environment.
- Allows user to filter specific data, and to track correlated pieces of data.

The Basic Syntax of SQL

- Syntax refers to a set of rules and guidelines that define a specific computer language.
- Includes the structure of words, symbols, numbers, and punctuation used to extract data and information.
- SQL database queries are executed using precise syntax of SQL statements.



Paolo De Gasperis/Shutterstock

SQL Query

A query in a database is a question or a request for specific information contained in a database.

- Example:
  - “How many people purchased our product in May 2021?“
  - “How many customers do we have in Idaho and Washington State?”

SQL statements are used to ask a question or communicate a request of the database management system.

Components of SQL 1

- SELECT – Allows the user to choose the precise fields to be returned.
- FROM – Allows the user to choose the tables where the fields needed for the query are located.
- WHERE – Allows the user to filter for desired and specific information.

Components of SQL 2

- In most SQL databases you can write SQL queries using lowercase letters.
- You may not have to worry about using the correct spacing.
- It is good practice to use proper syntax rules relating to case (upper or lower) and spacing in your queries.
- Punctuation rules must be followed exactly.

Components of SQL 3

SELECT * **FROM** Customers	**SELECT** followed by the asterisk (*) indicates that the user wants to extract all the columns of data **FROM** the Customers table. Use **SELECT** without context sparingly, because in a large database a huge volume of information may be returned.
SELECT City **FROM** Customers	**SELECT** identifies that the user wants to select the City column **FROM** the Customers table.
SELECT first_name **FROM** customer_information.customer_name **WHERE** first_name = 'david'	**SELECT** identifies the field to extract data from; in this case it is the first_name field. **FROM** indicates the table in which the field is located, and **WHERE** helps to narrow the query so that the only customers with the first name of 'David' are returned. You can use = **LIKE 'da%'** to return results from all first names that begin with the letters "da". The (%) is treated as a wildcard to match one or more characters. In some databases the asterisk (*) is used as a wildcard.

Components of SQL: Extracting Data From Multiple Fields

- The structure of a SQL query that extracts from multiple files is:

SELECT student_id, first_name, last_name **FROM** student_information.student_data **WHERE** first_name = 'david'	**SELECT** the columns named ‘student_id’, ‘first_name’, and ‘last_name’ **FROM** the field named 'student_data' in a table named ‘student_information’ **WHERE** only records from the first_name column are ‘david’.

Data Visualization

- The graphical and structured representation of data.
- Makes it easier to see the results and conclusions ascertained from data analysis.

ra2studio/Shutterstock

Three Step Process of Data Visualization

- **Step 1** – Explore data sets for patterns.
- **Step 2** – Planning for visuals.
- **Step 3** – Create your visuals.

ra2studio/Shutterstock

Tableau

Analytics software that offers business intelligence and data visualization across many areas as well as for organizational processes.

Allows users to explore and analyze data in seconds using a drag and drop interface as well as natural language processing which allows for the presentation of questions.

Can connect external data from:

- Spreadsheets.
- Databases.
- Big data.
- Data warehouses.
- Cloud data.

Tableau Licensing Options

Three role-based licensing options provided by Tableau:

- Creators – can build analytical content including data design, cleansing, curation of data sources, and the creation of visualizations and dashboards.
- Explorers – can access and analyze data published by Creators and can create their own dashboard.
- Viewers – can view and interact with visualizations and dashboards.

Microsoft Power BI

Analytics software that allows for the processing, manipulation, and visualization of data.

Users can quickly connect to data, prep it, model and visualize it using a variety of built-in tools and then securely share insights both internally and externally.

Users can connect to thousands of data sources located on-premise or in the Cloud including:

- Microsoft Excel, Salesforce, Google Analytics, a variety of social networks, and Internet of Things (IoT) devices.

Users can create data visualizations of the data as well as dashboards that give a 360-degree view of the organization.

Three Tiers of Microsoft Power BI

- Power BI desktop (free to download).
- Power BI Pro.
- Power BI Premium.

Sergei Bachlakov/Shutterstock

Selecting and Using the Right Data: Data Anonymization

- Aims to protect private or sensitive information by eliminating or encrypting this type of information.
- "personally identifiable information" (PII) is often used in data anonymization.
- Organizations have a legal and ethical responsibility to protect their data and the data of stakeholders and customers.

Personally Identifiable Information (PII)

- Name.
- Social security numbers.
- Medical records.
- Email addresses.
- Account numbers.
- Phone numbers.
- IP addresses.

maxkabakov/Getty Images

Open Data 1

Data openness (or open data) refers to free access, distribution, and usage of data.

To be considered as open, data must meet the following criteria:

- The public must have access and datasets must be available for use.
- Datasets must have access rights that allow them to be reused and redistributed.
- Datasets must be universally available so that anyone can use the data.

Open Data 2

Open databases and datasets allow for more wide use of data which can lead to many benefits.

For example, data collected during the Covid-19 pandemic was openly shared and used by governments, health institutions, and municipalities across the globe.

- Allowed for scientific collaboration, research advances, and eventually vaccine development.

Third Party Data

- Data collected by an organization that often does not have a relationship with the organization collecting the data or the data being collected.
- Example: A third party might be tasked with collecting information on visitors to a social media site of an organization.
- In order to maintain privacy, it is important that third-party data and PII are anonymized to ensure personal identifiable information is not made public or misused.

Selecting the Right Data: Key Considerations 1

How will the data be collected?

- Will you use internal data that has been collected by the organization or will you use data from other sources?

Where will you get your data?

- First-party data - data that has been collected internally by the organization.
- Second-party data - data that has been collected directly by another entity and then sold for use.
- Third-party data - data that is sold by an entity that did not actually collect the data.

Selecting the Right Data: Key Considerations 2

What types of data do you need to solve a specific problem?

- It is important to choose and use data that will help to solve the business or organizational problem under investigation.

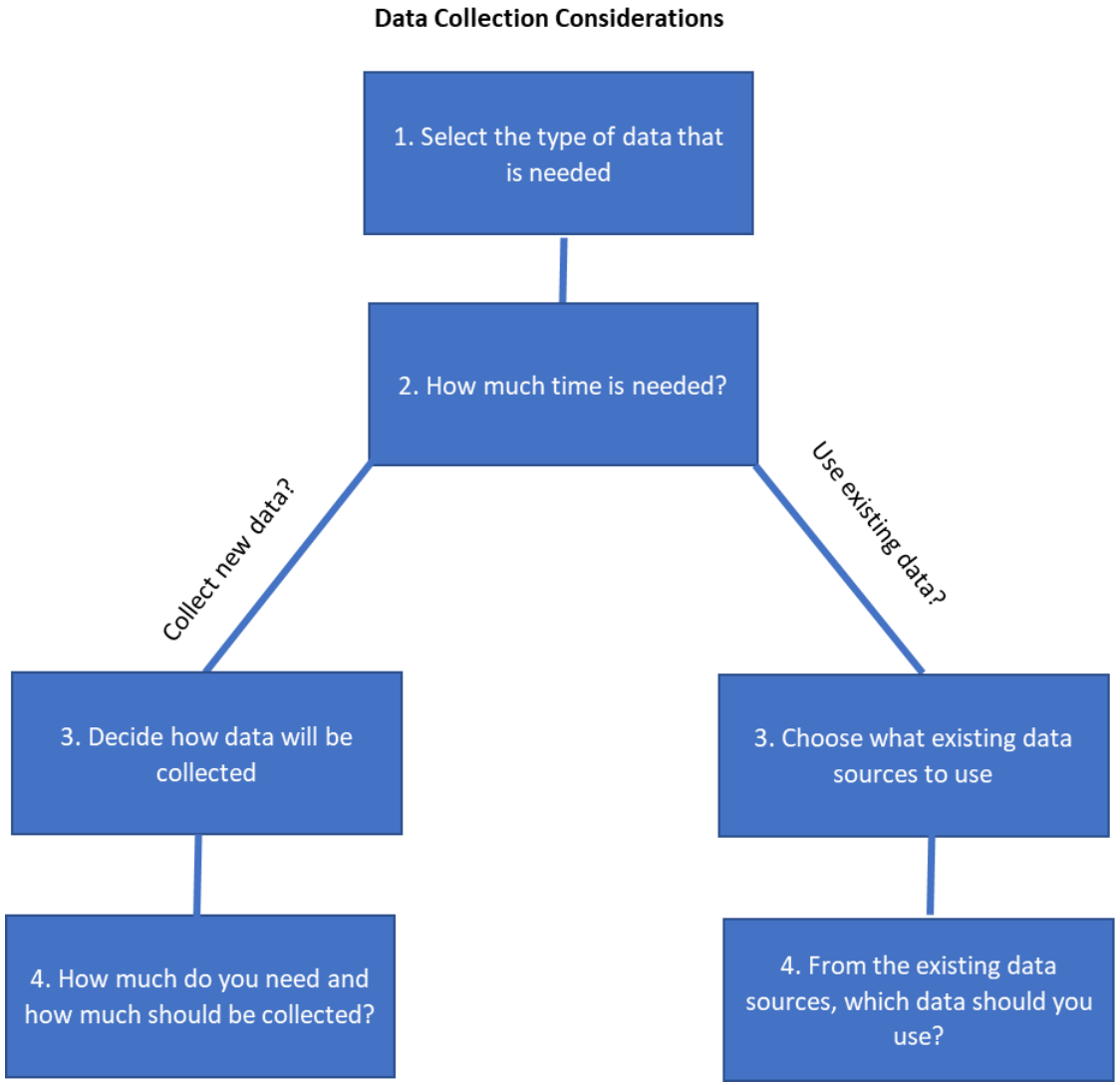
How much data should be collected?

- Will a random sample size from existing internal data be sufficient?
- Do you need a large sample size in order to achieve statistical significance?

How much time do you have?

- Time constraints should be identified and data collection should be closely monitored to ensure deadlines are met.

Selecting the Right Data: Key Considerations 3



Access the text alternative for slide images.

Data Formats in Datasets 1

- There are a variety of data formats included in datasets.

Data Formats in Datasets 2

Classification of Data Format	Definition of Data Format	Examples of Data Format
Primary Data	Newly collected data by a researcher, also referred to as first-hand data	• Data from interviews. • Data from surveys and questionnaires.
Secondary Data	Data that has already been collected by other researchers	• Data purchased from second-party and third-party vendors. • Demographic data from census.gov.
Internal Data	Data from an organization's own systems	• Sales data. • Data from human resources. • Inventory data.
External Data	Data from outside the organization	• Customer credit reports. • Wage information from the Bureau of Labor Statistics.
Continuous Data	Measurable and counted data that can have almost any numeric value	• Temperature data. • Grade distribution in MIS courses.
Discrete Data	Measurable and counted data that has a limited number of values	• Number of people who visit a storefront daily. • Number of items sold in a month.
Qualitative Data	Subjective measures of qualities and characteristics	• How using a product makes you feel. • Favorite brands.
Quantitative Data	Specific measures of numerical facts and values	• Population of the United States. • Number of customers.
Nominal Data	Type of qualitative data that is not categorized	• New product listing, existing product listing, de-listed products.
Ordinal Data	Type of qualitative data that is categorized	• Ranked choice of product (1st, 2nd, 3rd).
Structured Data	Data that is organized according to a defined format (rows, columns)	• Product inventory. • Expenses. • Accounting information.
Unstructured Data	Data that is not organized according to any structure	• Most email inboxes. • Instagram posts.

Data Transformation 1

- The process of converting the format of data from one form to another.
- Often data must be transformed to make it easier to analyze.
- Often data needs to be transformed so that it increases usability.
- Data can be transformed by changing format, structure, or value.

Data Transformation 2

The transformation of data often involves:

- Copying or replicating data.
- The deletion of database fields or records to ensure data integrity.
- The standardization of names and variables across data sets.
- Joining of tables or datasets.
- Converting a file to a different format such as saving an Excel spreadsheet to comma separated values (CSV).

Key Reasons for Data Transformation

- Data must be organized in a way that makes it easier to use.
- Data must be compatible so that it can be used between different systems and computer applications.
- Data must be designed so that it can move from one system to another.
- Often data must be merged to enhance decision making ability.
- Data should be structured to allow for enhancement and comparison.

Because learning changes everything.®

www.mheducation.com