

Data Analysis Course 1: Foundations: Data, Data Everywhere

Course 1, Module 1

Six steps of data analysis process:

1. Ask: business challenge, objective, or question
2. Prepare: data generation, collection, storage, and data management
3. Process: data cleaning and data integrity
4. Analyze: data exploration, visualization, and analysis
5. Share: communicating and interpreting results
6. Act: putting insights to work to solve the problem

data analysis: the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data Science: discipline of making data useful is umbrella term that encompasses three disciplines:

- Machine learning
- Statistics
- Analytics

There are four key types of business analytics: descriptive, predictive, diagnostic, and prescriptive. Descriptive analytics is the interpretation of historical data to identify trends and patterns, while predictive analytics centers on taking that information and using it to forecast future outcomes. Diagnostic analytics can be used to identify the root cause of a problem. In the case of prescriptive analytics, testing and other techniques are employed to determine which outcome will yield the best result in a given scenario.

data ecosystems: various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data.

data science: creating new ways of modeling and understanding the unknown by using raw data.

Data scientists create new questions using data, while analysts find answers to existing questions by creating insights from data sources.

data analytics: the science of data.

data driven decision-making: using facts to guide business strategy.

Ask often: "How do I define success for this project?"

Analytical skills are qualities and characteristics associated with solving problems using facts.

1. Curiosity: what kinds of questions would you ask based on the data and how it relates to the objectives of the exploratory data analysis (EDA)?
2. Understanding context: start to understand why the data shows what it does.
3. Having a technical mindset: approaching problems (and datasets) in a systematic and logical manner.
4. Data design: deals with how information is organized.
5. Data strategy: the management of the people, processes, and tools used in data analysis

Analytical thinking involves identifying and defining a problem and then solving it by using data in an organized, step-by-step manner.

5 Key Aspects to Analytical Thinking

1. Visualization: graphical representation of information.
2. Strategy: knowing what you want to achieve with the data and how to get there.
3. Problem-orientation: to identify, describe, and solve problems. Ask a lot of questions...
4. Correlation: relationship between two or more pieces of data. Correlation does not equal causation.
5. Big-picture and detail-oriented thinking: seeing big picture and details. Like a jigsaw puzzle.

Need analytical, creative, and critical thinking.

A few example questions:

- What is the root cause of a problem? (reason why a problem occurs.
 - Ask "why?" five times to reveal the root cause.
- Where are the gaps in our process?
 - Do gap analysis: a method for examining and evaluating how a process works currently in order to get where you want to be in the future.
 - Where you are now vs. where you want to be.
- What did we not consider before?

A quartile divides data points into four equal parts or quarters.

The analysts take their thinking a step further by using context to make

predictions, research answers, and eventually draw conclusions about what they've discovered.

Terms and definitions for Course 1, Module 1

Analytical skills: Qualities and characteristics associated with using facts to solve problems

Analytical thinking: The process of identifying and defining a problem, then solving it by using data in an organized, step-by-step manner

Context: The condition in which something exists or happens

Data: A collection of facts

Data analysis: The collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making

Data analyst: Someone who collects, transforms, and organizes data in order to draw conclusions, make predictions, and drive informed decision-making

Data analytics: The science of data

Data design: How information is organized

Data-driven decision-making: Using facts to guide business strategy

Data ecosystem: The various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data

Data science: A field of study that uses raw data to create new ways of modeling and understanding the unknown

Data strategy: The management of the people, processes, and tools used in data analysis

Data visualization: The graphical representation of data

Dataset: A collection of data that can be manipulated or analyzed as one unit

Gap analysis: A method for examining and evaluating the current state of a process in order to identify opportunities for improvement in the future

Root cause: The reason why a problem occurs

Technical mindset: The ability to break things down into smaller steps or pieces and work with them in an orderly and logical way

Visualization: (Refer to data visualization)

Course 1, Module 2

Stages of the data life-cycle

1. Plan: Decide what kind of data is needed, how it will be managed, and who will be responsible for it.
2. Capture: Collect or bring in data from a variety of different sources.
3. Manage: Care for and maintain the data. This includes determining how and where it is stored and the tools used to do so.
4. Analyze: Use the data to solve problems, make decisions, and support business goals.
5. Archive: Keep relevant data stored for long-term and future reference.
6. Destroy: Remove data from storage and delete any shared copies of the data.

Phases of data analysis:

- Ask: define the problem to be solved and make sure you fully understand stakeholder expectations.
- Prepare: where analysts collect and store data to use for upcoming analysis.
- Process: find and eliminate any errors and inaccuracies that can get in the way of results. This usually means cleaning data, transforming it into a more useful format, combining two or more datasets to make information more complete and removing outliers, which are any data points that could skew the information.
- Analyze: using tools to transform and organize data so that you can draw useful conclusions, make predictions, and drive informed decision-making.
- Share: how data analysts interpret results and share them with others to help stakeholders make effective data-driven decisions. Visualization is your best friend in this phase.
- Act: when business takes insights and puts them to work in order to solve the original business problem.

Data analyst tools

- Spreadsheets: Excel and Google Sheets
 - Store, organize, and sort data.

- Formula: a set of instructions that performs a specific calculation using the data in a spreadsheet.
- Function: a preset command that automatically performs a specific process or task using the data in a spreadsheet.
- Query language: a computer programming language that allows you to retrieve and manipulate data from a database.
 - SQL: structured query language.
- Visualization tools: helps data be understood and shared. Tableau and Looker.

A database is a collection of structured data stored in a computer system.

Terms and definitions for Course 1, Module 2

Database: A collection of data stored in a computer system

Formula: A set of instructions used to perform a calculation using the data in a spreadsheet

Function: A preset command that automatically performs a specified process or task using the data in a spreadsheet

Query: A request for data or information from a database

Query language: A computer programming language used to communicate with a database

Stakeholders: People who invest time and resources into a project and are interested in its outcome

Structured Query Language: A computer programming language used to communicate with a database

Spreadsheet: A digital worksheet

SQL: (Refer to Structured Query Language)

Course 1, Module 3

Spreadsheets

Attribute: a characteristic or quality of data used to label a column in a table.

Observation: all of the attributes for something contained in a row of a data table.

Query: request for data or information from a database.

SQL: store, organize, and analyze your data.

- Structure of a basic SQL query: select, from, where

Basic structure of a SQL query

SELECT	<i>[choose the column(s) you want]</i>	#2
FROM	<i>[from the appropriate table]</i>	#1
WHERE	<i>[a certain condition is met]</i>	#3

This is the suggested order in which you write your SQL queries. Start big (data table) and go small (specific conditions).

Use SELECT to choose the columns you want to return.

Use FROM to choose the tables where the columns you want are located.

Use WHERE to filter for certain information.

Syntax is the predetermined structure of a language that includes all required words, symbols, and punctuation, as well as their proper placement.

Course 1, Module 4

Blend personal and technical sides in analytics role.

Issue: topic or subject to investigate.

Question: designed to discover information.

Problem: an obstacle or complication that needs to be worked out.

Business task: question or problem data analysis answers for a business.

Fairness means ensuring that your analysis doesn't create or reinforce bias.




Fairness best practices:

- Consider all of the available data.
- Identify surrounding factors.

- Include self-reported data.
- Use oversampling effectively.
- Think about fairness from beginning to end.

Oversampling is the process of increasing the sample size of nondominant groups in a population. This can help you better represent them and address imbalanced datasets.

- Business analyst—analyzes data to help businesses improve processes, products, or services
- Data analytics consultant—analyzes the systems and models for using data
- Data engineer—prepares and integrates data from different sources for analytical use
- Data scientist—uses expert skills in technology and social science to find trends through data analysis
- Data specialist—organizes or converts data for use in databases or software systems
- Operations analyst—analyzes data to assess the performance of business operations and workflows

Decoding the job description			
	 Data Analysts	 Data Scientists	 Data Specialists
Problem solving	Use existing tools and methods to solve problems with existing types of data	Invent new tools and models, ask open-ended questions, and collect new types of data	Use in-depth knowledge of databases as a tool to solve problems and manage data
Analysis	Analyze collected data to help stakeholders make better decisions	Analyze and interpret complex data to make business predictions	Organize large volumes of data for use in data analytics or business operations
Other relevant skills	<ul style="list-style-type: none"> • Database queries • Data visualization • Dashboards • Reports • Spreadsheets 	<ul style="list-style-type: none"> • Advanced statistics • Machine learning • Deep learning • Data optimization • Programming 	<ul style="list-style-type: none"> • Data manipulation • Information security • Data models • Scalability of data • Disaster recovery

- Marketing analyst—analyzes market conditions to assess the potential sales of products and services
- HR/payroll analyst—analyzes payroll data for inefficiencies and errors

- Financial analyst—analyzes financial status by collecting, monitoring, and reviewing data
- Risk analyst—analyzes financial documents, economic conditions, and client data to help companies determine the level of risk involved in making a particular business decision
- Healthcare analyst—analyzes medical data to improve the business aspect of hospitals and medical facilities

Tips for interviews

Think about a time that you've used data to solve a problem.

Prepare specific questions for the interviewer

They want to see your thought process

Look for the recruiter

- Reach out directly

Showcase that you're open to learning

Terms and definitions for Course 1, Module 4

Business task: The question or problem data analysis resolves for a business

Fairness: A quality of data analysis that does not create or reinforce bias

Oversampling: The process of increasing the sample size of nondominant groups in a population. This can help you better represent them and address imbalanced datasets

Self-reporting: A data collection technique where participants provide information about themselves