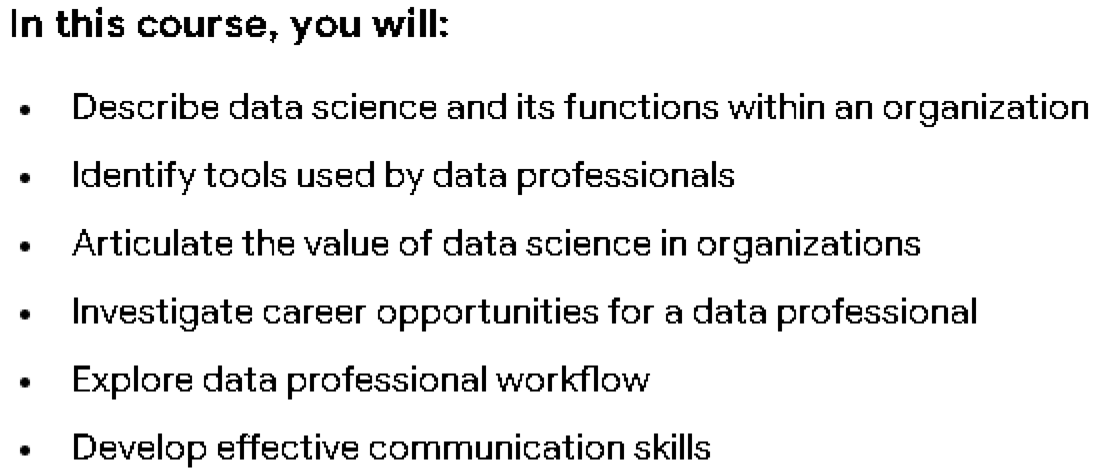
Module 1



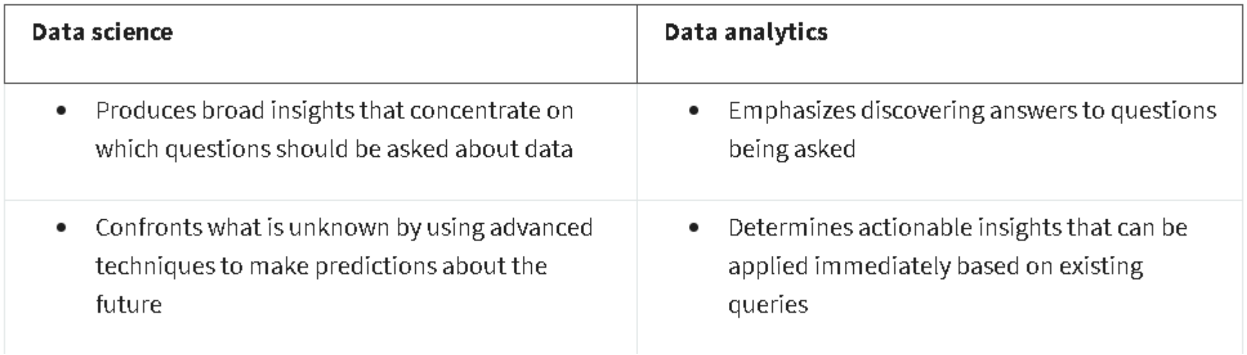
In this course, you will:

- Describe data science and its functions within an organization
- Identify tools used by data professionals
- Articulate the value of data science in organizations
- Investigate career opportunities for a data professional
- Explore data professional workflow
- Develop effective communication skills

Machine learning
- Automated task using data instead of instructions

Data science
- making data more useful
- field of study that uses raw data → model data and understand unknown

Data analytics
- capture, process, organize data



Data science	**Data analytics**
• Produces broad insights that concentrate on which questions should be asked about data	• Emphasizes discovering answers to questions being asked
• Confronts what is unknown by using advanced techniques to make predictions about the future	• Determines actionable insights that can be applied immediately based on existing queries

Data professional: data scientists and analysts
Data analytics professional: focused on data analytical processes
Data career space: spectrum of jobs in data science

<u>Practice Quiz: Assess readiness</u>

1. what is data science
2. diff b/w qual and quan data

3. what are wide and long data
4. structured data in which formats
5. boolean data type has 2 possible values
6. stakeholder have invested time and resources in project, interested in outcome
7. reasons to consider sample size
    a. collect data that represent diverse perspectives
    b. make sure unusual responses don't skew results
8. SMART questions that lead to change
    a. action-oriented
9. leading questions direct to particular answer, suggesting answer within question
    a. how satisfied
    b. in what ways
10. characteristics of metric
    a. quantifiable
    b. eval performance
    c. used for measurement
11. Interpretation bias can construe ambiguous situations
12. consent presumes indiv's right to know how and why data will be used
13. conditional formatting changes how cells appear based on condition
14. delimiter is specified character separating each item
15. programming languages write instructions for computers
16. benefit of programming language to work w/ data
    a. save time
    b. easily reproduce and share work
    c. clarify steps of analysis
17. syntax is predetermined structure of coding language
18. open-source freely avail, modified, and shared
19. programming language
    a. predictive modeling
    b. data transformation
    c. data visualization
    d. data cleaning
20. histogram: how often data values fall into ranges
21. line chart demonstrate over time
22. dynamic visualization automatically update over time
23. more effective to label data visualization instead of legend
    a. placed near data
    b. visualization more accessible
    c. allow text explanation
24. correlation measure degree to which two var change in relationship to each other
25. filters in data visualization tools
    a. limiting # row or columns
    b. providing data to diff users based on needs
    c. highlight indiv data points

Terms and definitions from Course 1, Module 1

Data professional: Any individual who works with data and/or has data skills

Data science: The discipline of making data useful

Data stewardship: The practices of an organization that ensure that data is accessible, usable, and safe

Edge computing: A way of distributing computational tasks over a bunch of nearby processors (i.e., computers) that is good for speed and resiliency and does not depend on a single source of computational power

Jupyter Notebook: An open-source web application used to create and share documents that contain live code, equations, visualizations, and narrative text

Machine learning: The use and development of algorithms and statistical models to teach computer systems to analyze patterns in data

Metrics: Methods and criteria used to evaluate data

Python: A general-purpose programming language

Module 1 Quiz

1. data professional explores, cleans, analyzes, and visualizes data
2. machine learning uses data instead of explicit instructions
3. evaluate data using metrics before creating predictive models to identify trends
4. Python is flexible, online community, easiest
5. Jupyter: web-based computing platform w/ Python
6. data visualization: sharing complex data, enriching data stores with visual, simplifying data using graphical interface
7. edge computing distributes computational tasks

Module 2

technical data professional

- build models and make predictions
- explore datasets
- data analysts, machine learning engineers, statisticians

strategic data professional

- interpret info
- align w/ business strategies
- business intelligence professionals and technical project managers

Data scientist and data analysts
- uncover trends, patterns, insights
- modeling and statistical analytic techniques
- explore vast and complex datasets

data management and infrastructure
- manage data sources and overall infrastructure
- databases

business intelligence
- perform predictive analysis
- create tables, reports, dashboards

product development teams
- manage analytical strategy

C-suite (chief)
- responsible for data and data professionals
- decision makers

Data cleaning
- formatting and removing data

Ideal qualities for data analytics professionals
- being coachable
- passion for data analysis
- lifelong learning
- strong interpersonal skills
- communication
- problem solver

Personally identifiable information (PII)
- sensitive → identify theft and security concerns

Aggregate information
- data from sig # users that has eliminated PII

bias
- preference in favour or against person, group of people, or thing

sample

- segment represent population

data privacy

- preserving data subject’s info and activity any time a data transaction occurs
    - data protection or information privacy
        - access, use, and collection of personal data
- protection from unauthorized access, freedom from inappropriate use, right to inspect/update data, give consent, legal right

data anonymization

- protecting private or sensitive by eliminating PII
- blanking, hashing, masking, hiding

data aggregation

- dataset is shown in groups

data stewardship

- practice ensuring data is accessible, usable, and safe

General Data Protection Regulation (GDPR)

- European Union law
- toughest privacy and security law
- other e.g. LGPD, CCPA

5 principles for data team building

1. adaptability: desire for learning and growing
2. activation: data literate, good relationship
3. standardization: best practices and transferability of info
4. accountability: transparency, explainability, and security to data
5. business impact: difficulty of integration commitment of resources, changes to project timeline

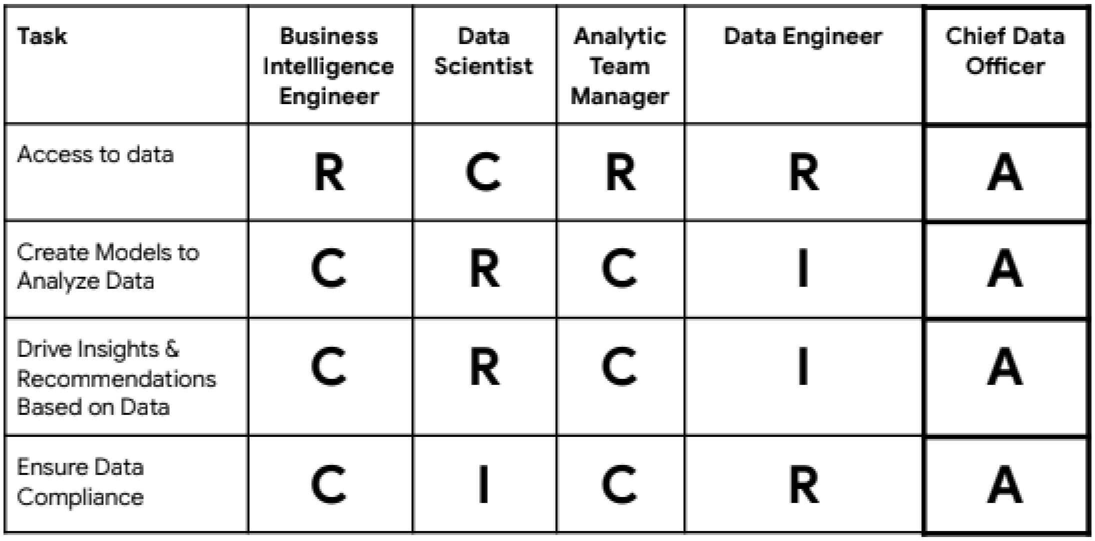
RACI (Responsible, Accountable, Consulted, Informed)

- organize roles and responsibilities
- who to contact & helps structure communication w/ team members
- levels of engagement
    - responsible: performing work or make decision to complete task
    - accountable: approving (i.e. manager or project lead
    - consulted: offer input (i.e. SME)
    - informed: aware of progress, senior leadership

Common data professional roles

- data scientist

    - work w/ analytics, provide meaningful insights
    - consults access to data
    - responsible for creating models to analyze data
- data engineer
    - infrastructure, responsible for developing and managing databases
    - work w/ data scientists to build custom pipelines to analyze raw data
    - responsible for ensuring data compliance
- analytics or insights teams manager
    - support team, lead, supervise
- business intelligence (BI) engineer
    - use business trends and databases to organize data and make accessible
    - provide reports
- Chief data officer
    - accountable for consistency, accuracy, relevancy, interpretability, and reliability of data



Task	Business Intelligence Engineer	Data Scientist	Analytic Team Manager	Data Engineer	Chief Data Officer
Access to data	R	C	R	R	A
Create Models to Analyze Data	C	R	C	I	A
Drive Insights & Recommendations Based on Data	C	R	C	I	A
Ensure Data Compliance	C	I	C	R	A

Module 2 Quiz

1. BI professionals are strategic data professionals
2. Hackathon is an event w/ programmers
3. PII includes national identification number
4. data aggregates collect info from sign # users, represent population as a whole
5. good sample represent entire population
6. AI perform tasks that would require human intelligence
7. all data professional responsible for ensuring inclusive practices, applying ethical
8. preventing data security breaches include communicating sensitive info impartially

Module 3

Tools

- spreadsheets: worksheet where data manipulated and calculated
- databases: data stored
- programming languages: write instructions
- data visualization: graphical representation
- dashboards: monitor live, incoming data

dataframe: table to organize
machine learning: use and development of algorithms and statistical models to teach computer systems to analyze and discover patterns

Large language models (LLM): type of AI algorithm trained on massive datasets

Course 1

- As a new member of a data analytics team, what steps could you take to be fully informed about a current project? Who would you like to meet with?
- How would you plan an analytics project?
- What steps would you take to translate a business question to an analytical solution?
- Why is actively managing data an important part of a data analytics team's responsibilities?
- What are some considerations you might need to be mindful of when reporting results?

Course 2

- Describe the steps you would take to clean and transform an unstructured data set.
- What specific things might you review for as part of your cleaning process?
- What are some of the outliers, anomalies, or unusual things you might consider in the data cleaning process that might impact analyses or the ability to create insights?

Course 3

- How would you explain the difference between qualitative and quantitative data sources?
- Describe the difference between structured and unstructured data.
- Why is it important to do exploratory data analysis (EDA)?
- How would you perform EDA on a given dataset?
- How do you create or alter a visualization based on different audiences?
- How do you avoid bias and ensure accessibility in a data visualization?
- How does data visualization inform your EDA?

Course 4

- How would you explain an A/B test to stakeholders who may not be familiar with analytics?

- If you had access to company performance data, what statistical tests might be useful to help understand performance?
- What considerations would you think about when presenting results to make sure they have an impact or have achieved the desired results?
- What are some effective ways to communicate statistical concepts/methods to a non-technical audience?
- In your own words, explain the factors that go into an experimental design for designs such as A/B tests.

Course 5

- Describe the steps you would take to run a regression-based analysis.
- List and describe the critical assumptions of linear regression.
- What is the primary difference between R2 and adjusted R2?
- How do you interpret a Q-Q plot in a linear regression model?
- What is the bias-variance tradeoff? How does it relate to building a multiple linear regression model? Consider variable selection and adjusted R2.

Course 6

- What kinds of business problems would be best addressed by supervised learning models?
- What requirements are needed to create effective supervised learning models?
- What does machine learning mean to you?
- How would you explain what machine learning algorithms do to a teammate who is new to the concept?
- How does gradient boosting work?

Module 3 Quiz

1. active listening: allow others to share POV before offering response
2. data cleaning: formate and remove unwanted data
3. data engineer
    a. manage infrastructure for data
    b. make data accessible
    c. ensure data ecosystem offers reliable results
4. insights or analytics team managers supervise organization’s analytical strategy
5. BI engineer organize data and make it accessible
6. RACI
    a. responsible, accountable, consulted, informed
    b. responsible: performs work and makes decisions related to task
7. adaptability: continue to learn and grow

Module 4

data workflow structure

- ask
- prepare
- process
- analyze
- share
- act

PACE

- plan
  - define scope, informational need, goal, strategy, business impact
- analyze
  - engage w/ data, acquire, clean, reorganize, transform data, engage EDA
- construct
  - build, revise ML models, uncover relationships, and apply statistical inferences
- execute
  - present findings to stakeholders, present recommendations

Key elements of communication

Message

- sender
- purpose
- receiver

elements of project proposal

Module 4 Quiz

1. In plan stage what questions
   a. strategies, goals, how business affected by plan
2. construct stage, build, interpret, revise models
3. present finding: execute
4. consider audience: receiver
5. best communicating practices: avoid unnecessary, break into shorter, keep simple
6. effect comm used in all four stages of PACE
7. active listening
   a. understand, look beyond, verify meeting notes
8. milestone include grouping of taste, which break up work in to more manageable goals

Module 5

- experiential learning
- portfolio