Data Advanced Analytics Google Certification

**Course 1: Foundations of Data Science**

A logo of a company

Description automatically generated

**Module 1:**

**Google Advanced Data Analytics Certificate overview**

Hello, and welcome to the **Google Advanced Data Analytics Certificate**! The seven courses in this certificate program will advance your knowledge and help you grow in your career as a data analytics professional. A foundational background in data analytics is strongly encouraged.

**Enter a growing field**

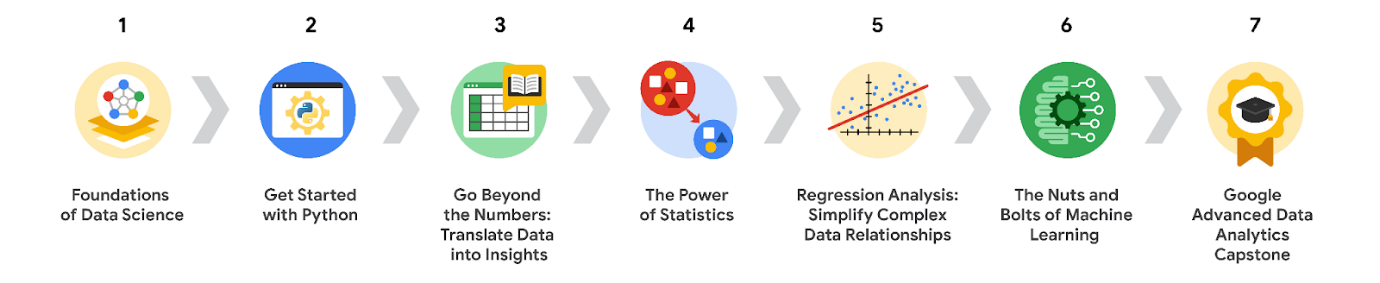
Why are data analytics skills in such high demand? Companies report the [largest skills gap](https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/beyond-hiring-how-companies-are-reskilling-to-address-talent-gaps) in data analytics, and the demand for data analytics skills is growing over [15 times faster](https://www.weforum.org/agenda/2022/11/digital-skills-labour-market-future/) than the demand for the average skill in the US. As computing power has accelerated, industries have sought to leverage the increasing amount of data available to them. Companies are incorporating data analysis to better understand their customers and improve their products and services. Increasingly, companies are employing data professionals to help guide future business decisions. As a result, data professionals need the ability to better understand the challenges and possibilities of machine learning, build accurate predictive models, and clearly articulate their findings.

Throughout this program, you will explore concepts and scenarios that will help you grow your knowledge and skills as a data professional.

**Google Advanced Data Analytics Certificate courses**

The Google Advanced Data Analytics has seven courses. During each course, you will complete hands-on activities that are similar to work performed by data analytics professionals every day. In doing so, you’ll gain experience that will help you effectively reveal the stories within the data and build models that can provide insight into trends. By the end of this program, you’ll understand how to program using Python, prepare data for analysis, utilize statistics with data, build predictive models, optimize machine learning capabilities, and much more. The courses of the program are as follows:

1. [**Foundations of Data Science**](https://www.coursera.org/learn/foundations-of-data-science/home/week/1)
2. [**Get Started with Python**](https://www.coursera.org/learn/get-started-with-python/home/week/1)
3. [**Go Beyond the Numbers: Translate Data into Insights**](https://www.coursera.org/learn/go-beyond-the-numbers-translate-data-into-insight/home/week/1)
4. [**The Power of Statistics**](https://www.coursera.org/learn/the-power-of-statistics/home/week/1)
5. [**Regression Analysis: Simplify Complex Data Relationships**](https://www.coursera.org/learn/regression-analysis-simplify-complex-data-relationships/home/week/1)
6. [**The Nuts and Bolts of Machine Learning**](https://www.coursera.org/learn/the-nuts-and-bolts-of-machine-learning/home/week/1)
7. [**Google Advanced Data Analytics Capstone**](https://www.coursera.org/learn/google-advanced-data-analytics-capstone/home/week/1)



**Google Data Analytics Certificate content**

These courses build on many foundational concepts examined in the [Google Data Analytics Certificate](https://www.coursera.org/professional-certificates/google-data-analytics?utm_source=google&utm_medium=institutions&utm_campaign=gwgsite-gDigital-paidha-sem-bk-data-exa-glp-br-null&_ga=2.122602571.1926911371.1663777781-1123481676.1663777781), and you will encounter links to content from the Data Analytics Certificate courses. This material is meant to provide an optional review and act as a useful resource.

**Benefits for job seekers**

After completing all seven courses, Google Advanced Data Analytics Certificate graduates get access to exclusive job search resources, courtesy of Google. You’ll have the opportunity to:

* Build your resume, participate in mock interviews, and receive job search tips through Big Interview, a job-training platform that’s free for program graduates.
* Improve your interview technique with Interview Warmup, a tool built by Google with certificate graduates in mind. Access data analytics-specific practice questions, transcripts of your responses, and automatic insights that help you grow your skills and confidence.
* Access thousands of job postings and free one-on-one career coaching with Career Circle. (You must be eligible to work in the U.S. to join.)
* Claim your Google Advanced Data Analytics Certificate badge, and share your achievement on LinkedIn to stand out among other candidates to potential employers.

Congratulations on taking this next step in your journey to a fulfilling career in data analytics!

**Course 1 overview**

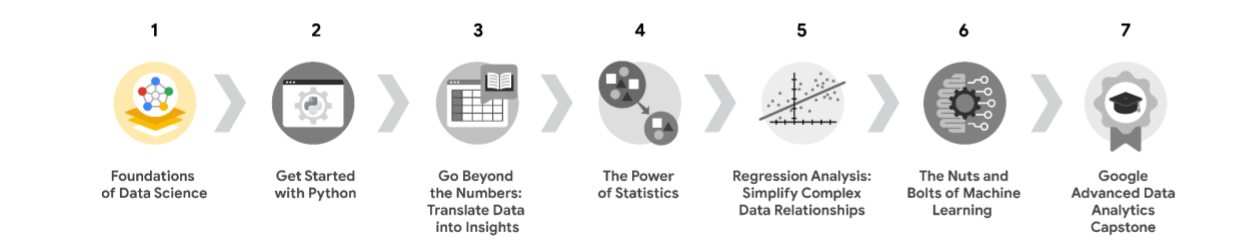


Hello, and welcome to **Foundations of Data Science**, the first course in the Google Advanced Data Analytics Certificate. You’re starting an exciting journey!

In this course, you will be introduced to the basic concepts of data science. You will discover how data professionals operate in the workplace and how different roles contribute to an organization’s vision of the future. You will also begin to explore the data workflow PACE, which stands for plan, analyze, construct, and execute. The PACE workflow will help you better understand how to navigate the technical and workplace expectations of this career space.

**Course descriptions**

The Google Advanced Data Analytics Certificate has seven courses. **Foundations of Data Science** is the first course.



1. [**Foundations of Data Science**](https://www.coursera.org/learn/foundations-of-data-science/home/week/1) — *(current course)* Learnhow data professionals operate in the workplace and how different roles in the field of data science contribute to an organization’s vision of the future. Then, explore data science roles, communication skills, and data ethics.
2. [**Get Started with Python**](https://www.coursera.org/learn/get-started-with-python/home/week/1) —Discover how the programming language Python can power your data analysis. Learn core Python concepts, such as data types, functions, conditional statements, loops, and data structures.
3. [**Go Beyond the Numbers: Translate Data into Insights**](https://www.coursera.org/learn/go-beyond-the-numbers-translate-data-into-insight/home/week/1) — Learn the fundamentals of data cleaning and visualizations and how to reveal the important stories that live within data.
4. [**The Power of Statistics**](https://www.coursera.org/learn/the-power-of-statistics/home/week/1) — Explore descriptive and inferential statistics, basic probability and probability distributions, sampling, confidence intervals, and hypothesis testing.
5. [**Regression Analysis: Simplify Complex Data Relationships**](https://www.coursera.org/learn/regression-analysis-simplify-complex-data-relationships/home/week/1) — Learn to model variable relationships, focusing on linear and logistic regression.
6. [**The Nuts and Bolts of Machine Learning**](https://www.coursera.org/learn/the-nuts-and-bolts-of-machine-learning/home/week/1) — Learn unsupervised machine learning techniques and how to apply them to organizational data.
7. [**Google Advanced Data Analytics Capstone**](https://www.coursera.org/learn/google-advanced-data-analytics-capstone/home/week/1) — Complete a hands-on project designed to demonstrate the skills and competencies you acquire in the program.

**Course 1 content**

Each course of this certificate program is broken into modules. You can complete courses at your own pace, but the module breakdowns are designed to help you finish the entire Google Advanced Data Analytics Certificate in about six months.

What’s to come? Here’s a quick overview of the skills you’ll learn in each module of this course.



**Module 1: Introduction of data science concepts**

You will find out what’s in store for Course 1 and the whole certificate program. You’ll explore the Coursera platform, procedures, and content types, and meet other learners in the program. Then, you’ll learn about data science fundamentals, and investigate the question “What is data science?” by examining  elements of its history, and tracing its current defining characteristics. Finally, you’ll learn some tips for embarking on careers in this field.

**Module 2: The impact of data today**

In this part of the course, you will explore examples of the ways data-driven insights impact different industries. You’ll learn how data analytics professionals are transforming organizations and discover how data professionals are approaching data stewardship. Then, you’ll consider the trajectory of data analytics and forecast the future of data fields.

**Module 3: Your career as a data professional**

You will learn what it means to have a career as a data professional. In this section, you’ll examine the required skills of a data analyst and compare it to many other data professional roles. Then you’ll discover how data professionals operate within cross-functional teams and how they serve their organizations.

**Module 4: Data applications and workflow**

You will find out how to structure your work as a data analyst and learn about the PACE workflow, which will guide you as you learn how to think like a data professional. You’ll also investigate what it means to be a successful communicator and how communication can help drive data analysis through the PACE workflow.

**Module 5: Course 1 end-of-course project**

You will work on a project that will help develop your resume and refine your ability to think like a data professional. You’ll end the course by reflecting on your learning journey, and then you will get a preview of the next course.

**What to expect**

Each course offers many types of learning opportunities:

* **Videos** led by Google instructors teach new concepts, introduce the use of relevant tools, offer career support, and provide inspirational personal stories.
* **Readings** build on the topics discussed in the videos, introduce related concepts, share useful resources, and describe case studies.
* **Discussion prompts** explore course topics for better understanding and allow you to chat and exchange ideas with other learners in the [**discussion forums**](https://www.coursera.org/learn/foundations-of-data-science/discussions).
* **Self-review activities** and **labs** give you hands-on practice in applying the skills you are learning and allow you to assess your own work by comparing it to a completed example.
* **Interactive plug-ins** encourage you to practice specific tasks and help you integrate the knowledge you have gained in the course.
* **In-video quizzes** help you check your comprehension as you progress through each video.
* **Practice quizzes** allow you to check your understanding of key concepts and provide valuable feedback.
* **Graded quizzes** demonstrate your understanding of the main concepts of a course. You must score 80% or higher on each graded quiz to obtain a certificate, and you can take a graded quiz multiple times to achieve a passing score.

**Tips for success**

* It is strongly recommended that you go through the items in each lesson in the order they appear because new information and concepts build on previous knowledge.
* Participate in all learning opportunities to gain as much knowledge and experience as possible.
* If something is confusing, don’t hesitate to replay a video, review a reading, or repeat a self-review activity.
* Use the additional resources that are referenced in this course. They are designed to support your learning. You can find all of these resources in the [**Resources**](https://www.coursera.org/learn/foundations-of-data-science/resources/6o2SH) tab.
* When you encounter useful links in this course, bookmark them so you can refer to the information later for study or review.
* Understand and follow the [Coursera Code of Conduct](https://www.coursera.support/s/article/208280036-Coursera-Code-of-Conduct?) to ensure that the learning community remains a welcoming, friendly, and supportive place for all members.

**Course 1 overview**

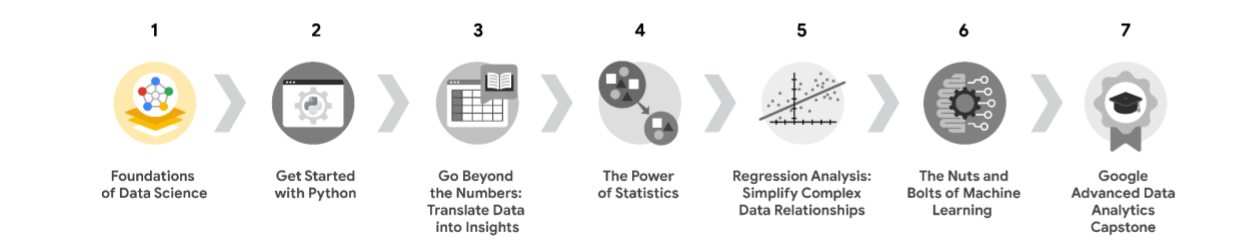


Hello, and welcome to **Foundations of Data Science**, the first course in the Google Advanced Data Analytics Certificate. You’re starting an exciting journey!

In this course, you will be introduced to the basic concepts of data science. You will discover how data professionals operate in the workplace and how different roles contribute to an organization’s vision of the future. You will also begin to explore the data workflow PACE, which stands for plan, analyze, construct, and execute. The PACE workflow will help you better understand how to navigate the technical and workplace expectations of this career space.

**Course descriptions**

The Google Advanced Data Analytics Certificate has seven courses. **Foundations of Data Science** is the first course.



1. [**Foundations of Data Science**](https://www.coursera.org/learn/foundations-of-data-science/home/week/1) — *(current course)* Learnhow data professionals operate in the workplace and how different roles in the field of data science contribute to an organization’s vision of the future. Then, explore data science roles, communication skills, and data ethics.
2. [**Get Started with Python**](https://www.coursera.org/learn/get-started-with-python/home/week/1) —Discover how the programming language Python can power your data analysis. Learn core Python concepts, such as data types, functions, conditional statements, loops, and data structures.
3. [**Go Beyond the Numbers: Translate Data into Insights**](https://www.coursera.org/learn/go-beyond-the-numbers-translate-data-into-insight/home/week/1) — Learn the fundamentals of data cleaning and visualizations and how to reveal the important stories that live within data.
4. [**The Power of Statistics**](https://www.coursera.org/learn/the-power-of-statistics/home/week/1) — Explore descriptive and inferential statistics, basic probability and probability distributions, sampling, confidence intervals, and hypothesis testing.
5. [**Regression Analysis: Simplify Complex Data Relationships**](https://www.coursera.org/learn/regression-analysis-simplify-complex-data-relationships/home/week/1) — Learn to model variable relationships, focusing on linear and logistic regression.
6. [**The Nuts and Bolts of Machine Learning**](https://www.coursera.org/learn/the-nuts-and-bolts-of-machine-learning/home/week/1) — Learn unsupervised machine learning techniques and how to apply them to organizational data.
7. [**Google Advanced Data Analytics Capstone**](https://www.coursera.org/learn/google-advanced-data-analytics-capstone/home/week/1) — Complete a hands-on project designed to demonstrate the skills and competencies you acquire in the program.

**Course 1 content**

Each course of this certificate program is broken into modules. You can complete courses at your own pace, but the module breakdowns are designed to help you finish the entire Google Advanced Data Analytics Certificate in about six months.

What’s to come? Here’s a quick overview of the skills you’ll learn in each module of this course.



**Module 1: Introduction of data science concepts**

You will find out what’s in store for Course 1 and the whole certificate program. You’ll explore the Coursera platform, procedures, and content types, and meet other learners in the program. Then, you’ll learn about data science fundamentals, and investigate the question “What is data science?” by examining  elements of its history, and tracing its current defining characteristics. Finally, you’ll learn some tips for embarking on careers in this field.

**Module 2: The impact of data today**

In this part of the course, you will explore examples of the ways data-driven insights impact different industries. You’ll learn how data analytics professionals are transforming organizations and discover how data professionals are approaching data stewardship. Then, you’ll consider the trajectory of data analytics and forecast the future of data fields.

**Module 3: Your career as a data professional**

You will learn what it means to have a career as a data professional. In this section, you’ll examine the required skills of a data analyst and compare it to many other data professional roles. Then you’ll discover how data professionals operate within cross-functional teams and how they serve their organizations.

**Module 4: Data applications and workflow**

You will find out how to structure your work as a data analyst and learn about the PACE workflow, which will guide you as you learn how to think like a data professional. You’ll also investigate what it means to be a successful communicator and how communication can help drive data analysis through the PACE workflow.

**Module 5: Course 1 end-of-course project**

You will work on a project that will help develop your resume and refine your ability to think like a data professional. You’ll end the course by reflecting on your learning journey, and then you will get a preview of the next course.

**What to expect**

Each course offers many types of learning opportunities:

* **Videos** led by Google instructors teach new concepts, introduce the use of relevant tools, offer career support, and provide inspirational personal stories.
* **Readings** build on the topics discussed in the videos, introduce related concepts, share useful resources, and describe case studies.
* **Discussion prompts** explore course topics for better understanding and allow you to chat and exchange ideas with other learners in the [**discussion forums**](https://www.coursera.org/learn/foundations-of-data-science/discussions).
* **Self-review activities** and **labs** give you hands-on practice in applying the skills you are learning and allow you to assess your own work by comparing it to a completed example.
* **Interactive plug-ins** encourage you to practice specific tasks and help you integrate the knowledge you have gained in the course.
* **In-video quizzes** help you check your comprehension as you progress through each video.
* **Practice quizzes** allow you to check your understanding of key concepts and provide valuable feedback.
* **Graded quizzes** demonstrate your understanding of the main concepts of a course. You must score 80% or higher on each graded quiz to obtain a certificate, and you can take a graded quiz multiple times to achieve a passing score.

**Tips for success**

* It is strongly recommended that you go through the items in each lesson in the order they appear because new information and concepts build on previous knowledge.
* Participate in all learning opportunities to gain as much knowledge and experience as possible.
* If something is confusing, don’t hesitate to replay a video, review a reading, or repeat a self-review activity.
* Use the additional resources that are referenced in this course. They are designed to support your learning. You can find all of these resources in the [**Resources**](https://www.coursera.org/learn/foundations-of-data-science/resources/6o2SH) tab.
* When you encounter useful links in this course, bookmark them so you can refer to the information later for study or review.
* Understand and follow the [Coursera Code of Conduct](https://www.coursera.support/s/article/208280036-Coursera-Code-of-Conduct?) to ensure that the learning community remains a welcoming, friendly, and supportive place for all members.

# Helpful resources and tips

As a learner, you can choose to complete one or multiple courses in this program. However, to obtain the Google Advanced Data Analytics Certificate, you must complete all of the courses. This reading describes what is required to obtain a certificate and best practices for you to have a good learning experience on Coursera.

## Obtain the Google Advanced Data Analytics Certificate

To receive your official Google Advanced Data Analytics Certificate, you must:

* Pass all graded assignments in all 7 courses of the certificate program. Each graded assignment is part of a cumulative graded score for the course, and the passing grade for each course is 80%.

AND **one**of the following:

* Pay the [course certificate fee](https://www.coursera.support/s/article/209818963-Payments-on-Coursera?language=en_US),
* Be approved for [Coursera Financial Aid](https://www.coursera.support/s/article/209819033-Apply-for-Financial-Aid-or-a-Scholarship?language=en_US), **or**
* Complete the certificate through an educational institution, employer, or agency that's sponsoring your participation.

## Healthy habits for course completion

Here is a list of best practices that will help you complete the courses in the program in a timely manner:

* **Plan your time:** Setting regular study times and following them each week can help you make learning a part of your routine. Use a calendar or timetable to create a schedule, and list what you plan to do each day in order to set achievable goals. Find a space that allows you to focus when you watch the videos, review the readings, and complete the activities.
* **Work at your own pace:** Everyone learns differently, so this program has been designed to let you work at your own pace. Although your personalized deadlines start when you enroll, feel free to progress through the program at the speed that works best for you. There is no penalty for late assignments; to earn your certificate, all you have to do is complete all of the work. You can extend your deadlines at any time by going to **Overview** in the navigation panel and selecting **Switch Sessions**. If you have already missed previous deadlines, select **Reset my deadlines** instead.
* **Be curious:** If you find an idea that gets you excited, act on it! Ask questions, search for more details online, explore the links that interest you, and take notes on your discoveries. The steps you take to support your learning along the way will advance your knowledge, create more opportunities in this high-growth field, and help you qualify for jobs.
* **Take notes:** Notes will help you remember important information in the future, especially as you’re preparing to enter a new job field. In addition, taking notes is an effective way to make connections between topics and gain a better understanding of those topics.
* **Review exemplars:** Exemplars are completed assignments that fully meet an activity's criteria. Many activities in this program have exemplars for you to compare to your own work. Although there are often many ways to complete an assignment, exemplars offer you guidance and inspiration about how to complete the activity.
* **Build your career identity**: Your career identity is the unique value you bring to the workforce. [Watch this video](https://www.youtube.com/watch?v=_xbT4qMrot4) to learn about the key components of career identity and write your own career identity statement. Discovering and defining your own career identity makes you better equipped to choose a career path that aligns with your strengths, values, and goals and is more effective in your chosen profession.
* **Study together:** It can be hard to find the time to progress through a Google Career Certificate with all life’s distractions. [Register for Study Group](https://lu.ma/GoogleCareerCertificatesStudyGroup?utm_source=course) – live study sessions hosted Monday through Thursday that offer you dedicated time and space to focus on your coursework alongside a community of your peers. Participation in Study Group is optional for our learners.
* **Connect with other learners:** If you have a question, chances are, you’re not alone. Reach out in the discussion forum to ask for help from other learners in this program. You can also visit Coursera’s private [Google Advanced Data Analytics Community](http://www.coursera.support/s/group-invite?id=MEY5VkgwMDAwMDAwMWdyMEFB) to expand your network, discuss career journeys, and share experiences. Check out the [quick start guide](https://www.coursera.support/s/article/Community-Quick-Start-Guide).
* **Update your profile:** Consider [updating your profile](https://www.coursera.org/account/profile) on Coursera to include your photo, career goals, and more. When other learners find you in the discussion forums, they can click on your name to access your profile and get to know you better.

## Documents, spreadsheets, presentations, and labs for course activities

To complete certain activities in the program, you will need to use digital documents, spreadsheets, presentations, and/or labs. Data analytics professionals use these software applications to collaborate within their teams and organizations. If you need more information about using a particular tool, refer to these resources:

* [Microsoft Word: Help and learning](https://support.microsoft.com/en-us/word): Microsoft Support page for Word
* [Google Docs](https://support.google.com/docs/topic/9046002?hl=en&ref_topic=1382883): Help Center page for Google Docs
* [Microsoft Excel: Help and learning](https://support.microsoft.com/en-us/excel): Microsoft Support page for Excel
* [Google Sheets](https://support.google.com/docs/topic/9054603?hl=en&ref_topic=1382883): Help Center page for Google Sheets
* [Microsoft PowerPoint: Help and learning](https://support.microsoft.com/en-us/powerpoint): Microsoft Support page for PowerPoint
* [How to use Google Slides](https://support.google.com/docs/answer/2763168?hl=en&co=GENIE.Platform%3DDesktop): Help Center page for Google Slides
* [Common problems with labs](https://support.google.com/qwiklabs/answer/9133560?hl=en&ref_topic=9134804): Troubleshooting help for Qwiklabs activities

## Module, course, and certificate glossaries

This program covers a lot of terms and concepts, some of which you may already know and some of which may be unfamiliar to you. To review terms and help you prepare for graded quizzes, refer to the following glossaries:

* **Module glossaries**: At the end of each module’s content, you can review a glossary of terms from that module. Each module’s glossary builds upon the terms from the previous modules in that course. The module glossaries are not downloadable; however, all of the terms and definitions are included in the course and certificate glossaries, which are downloadable.
* **Course glossaries**: At the end of each course, you can access and download a glossary that covers all of the terms in that course.
* **Certificate glossary**: The certificate glossary includes all of the terms in the entire certificate program and is a helpful resource that you can reference throughout the program or at any time in the future.

You can access and download the certificate glossaries and save them on your computer. You can always find the course glossary using the course’s [Resources](https://www.coursera.org/learn/foundations-of-data-science/resources/cPuiu) tab. To access the **Advanced Data Analytics Certificate glossary**, click the following link and select Use Template.

* Link to the glossary: [Advanced Data Analytics Certificate glossary](https://docs.google.com/document/d/193-AtS7MlB2w4buwiCyPjBoOhIbbByKgHWPpYnSR9VI/template/preview)

OR

* If you don’t have a Google account, you can download the glossary directly from the following attachment.

[Advanced Data Analytics Certificate glossary](https://d3c33hcgiwev3.cloudfront.net/PFymNGYQQ5Cf1XbjyxwNOg_fe8a91120d2244988c658b5a363087f1_Advanced-Data-Analytics-Certificate-glossary.docx?Expires=1719878400&Signature=EoSRlxpbO-RnIEkLCYTPRH1YAFey1eQvZ3GdZGx5MLUtCIfIw1EwuuvDSN4TEoUicQXryORSuZH0-jQ-aXDsY5gdh57~f6vplqYwd~ZtSiZgZUNQuFbe06Fvtrd89M~npYgX2tJjT96O7YgZsZHQPI1gWA0jwRbogWKEIb-cn-I_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[DOCX File](https://d3c33hcgiwev3.cloudfront.net/PFymNGYQQ5Cf1XbjyxwNOg_fe8a91120d2244988c658b5a363087f1_Advanced-Data-Analytics-Certificate-glossary.docx?Expires=1719878400&Signature=EoSRlxpbO-RnIEkLCYTPRH1YAFey1eQvZ3GdZGx5MLUtCIfIw1EwuuvDSN4TEoUicQXryORSuZH0-jQ-aXDsY5gdh57~f6vplqYwd~ZtSiZgZUNQuFbe06Fvtrd89M~npYgX2tJjT96O7YgZsZHQPI1gWA0jwRbogWKEIb-cn-I_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

## Data Analytics Certificate glossary

If you completed the original [Google Data Analytics Certificate](https://www.coursera.org/professional-certificates/google-data-analytics?utm_source=google&utm_medium=institutions&utm_campaign=gwgsite-gDigital-paidha-sem-bk-gen-exa-glp-br-null&_ga=2.170664992.1625030801.1661901112-1742325342.1661901112), you may recognize some overlap with several of the glossary terms in this program. Refer to the Data Analytics Certificate glossary, linked in the [Resources](https://www.coursera.org/learn/foundations-of-data-science/resources/CWv8t) tab, to review these foundational terms and concepts. The definitions of some terms in the Data Analytics Certificate glossary differ from the definitions of the same terms in this program since the Advanced Data Analytics Certificate builds upon the concepts taught in the previous program.

## Course feedback

Providing feedback on videos, readings, and other materials is easy. With the resource open in your browser, you can find the thumbs-up and thumbs-down symbols.

* Click **thumbs-up** for materials you find helpful.
* Click **thumbs-down** for materials that you do not find helpful.

If you want to flag a specific issue with an item, click the flag icon, select a category, and enter an explanation in the text box. This feedback goes back to the course development team and isn’t visible to other learners. All feedback received helps to create even better certificate programs in the future.

For technical help, visit the [Learner Help Center](https://learner.coursera.help/hc/en-us).

## Question

Fill in the blank: Data professionals use \_\_\_\_\_ to work efficiently with large datasets.

programming languages

# Prepare to assess your readiness for the Google Advanced Data Analytics Certificate

The Google Advanced Data Analytics Certificate is a program for everyone interested in advanced concepts in data analysis. However, a background in data analytics is a useful foundation for many of the concepts that will be taught in these courses. We have created a quiz that will enable you to assess your readiness and to decide if you would like to proceed with the certificate or review some foundations first.

The quiz will ask questions about how comfortable you are with the following data analytics concepts:

* General data analytics principles and data types
* Data strategy
* Data integrity and data cleaning
* Programming languages
* Data visualization, dashboards, and sharing insights with others

If you are coming to this program from the Google Data Analytics Certificate or you have worked in a data analytics field, you may already be familiar with these concepts. That’s great! This quiz is just a way to double-check that you’re prepared to succeed. Plus, it will provide you with some resources to give you a refresher before moving on.

To help you decide if you’re ready for this course:

* Take the optional readiness quiz.
* Refer to the scoring guide to determine if you’re prepared for the Advanced Data Analytics content. A score of 80% or higher is the target goal for readiness for these courses.
* Based on your individual score, follow the recommendations in the scoring guide for your next steps.

If you are confident that you have the prerequisite knowledge and experience for this certificate, you can skip the readiness quiz after this reading and move directly to the next activity, which is the weekly wrap-up video located at the beginning of the next lesson.

# Data discourse over the years

You will encounter the terms “data science” and “data analytics” throughout this program. Although these disciplines can be quite different in their scope, industry professionals often use them interchangeably. It’s important to examine the similarities and differences between data science and data analytics to better understand the data career field and where your skills fit. In this reading, you’ll learn more about data science, data analytics, and how the data profession has evolved over time.

## What is data science vs. data analytics

Data science is an entire field dedicated to making data more useful. A data scientist is a professional that uses raw data to develop new ways to model data and understand the unknown. Often, their job responsibilities incorporate various components of computer science, predictive analytics, statistics, and machine learning. The collections of information that data scientists work with can be quite large, requiring expertise to organize and navigate.

Data analytics is a subfield of the larger data science discipline. The aim of data analytics is to create methods to capture, process, and organize data to uncover actionable insights for current problems. Analysts focus on processing the information stored in existing datasets and establishing the best way to present this data. Data analysts rely on statistics and data modeling to solve problems and offer recommendations that can lead to immediate improvements.

The following table presents a side-by-side comparison:

| **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 |

## The connections between data science and data analytics

Data science and data analytics share a fundamental goal: discover insights that can be used to lead an organization to improve and grow. They are closely connected with information gathered through interactions within the measurable world. As data projects become more complicated, organizations are discovering the advantages of assembling data teams, bringing data analysts and data scientists together. Within these highly-collaborative working environments, data scientists and data analysts work on common tasks using similar software packages, tools, and programming languages. Additionally, the responsibilities of data scientists and data analysts are impacted by a company’s resources, budgetary constraints, and other factors which vary depending on the organization and industry. The commonalities between data science and data analytics within the workplace explain why practitioners and industry professionals often use the two titles interchangeably. This has also extended to job postings.

## The data professional in the data career space

This program uses a few terms when addressing data science and data analytics professionals. The term **data professional** is used when referring to data scientists and analysts collectively. **Data analytics professional** is used to highlight positions where the majority of tasks focus on data analytical processes. The **data career space** functions as an inclusive term referencing the spectrum of jobs and careers in data science and related areas. The intention for these descriptors is to serve as inclusive labels—enabling the broadest possible grouping of data

| **Field** | **Summary** |
| --- | --- |
| **Technology** | Phone lines become the superhighways of communication. By the 1980s, the scanning fax machine became a necessity for businesses, allowing them to quickly send and receive documents. Desktops were now small enough for users to have a personal computer. Windows 1.0 was released. |
| **Data storage** | The floppy disk format replaced data punch cards, allowing 360 kilobytes of data storage. The average internal hard drive capacity was 10 megabytes. |
| **Data analysis** | Data organization was revolutionized with the development of electronic spreadsheets. These programs first emulated traditional financial ledgers. |
| **Communication** | Internal electronic messages or email were being used by government agencies, universities, and defense/military industries to communicate. |

science and data analytics related jobs and titles.

Career opportunities for data professionals span a wide range of industries. Each with their individualized needs, demands, and responsibilities. It’s possible for two employees with the same job title at two different organizations to have different roles and tasks. To help clarify how careers across data science and data analytics are described in this program, the data career space is used to collectively refer to the variety of careers and job opportunities available to those with data science skills.

## The data career space over time

So far, you’ve been considering how data science and data analytics work together in the data career space. This field has also been evolving and changing over time as technology, data storage, data analysis, and communication has changed. Progress in the field means that data professionals need to learn to adapt and evolve to build and define careers within this space—and understanding how the data career space has evolved so far can help you do just that.

### 1965

In 1965, the world’s population was 3.3 billion. The Mariner 4 spacecraft took the first close-up pictures of Mars. The first commercial communications satellite allowed nearly instantaneous contact between Europe and North America.

| **Field** | **Summary** |
| --- | --- |
| **Technology** | Computations were still done by hand; companies were just starting to  invest in new computing technologies to assist analysts. The first desktop programmable calculators replaced room-sized computers. Users could backup complex mathematical operations onto magnetic cards and print the results. |
| **Data storage** | Simple punch cards stored programming information. These cards would be loaded in a specific order to give commands. Businesses allocated physical space to store the thousands of data punch cards regular computer usage would generate. Larger organizations upgraded to magnetic tape reels that could hold the equivalent of 50,000 punch cards, or about 3 megabytes. But these reels still required a lot of storage space within an office. |
| **Data analysis** | As computer technology reduced in size and cost, more companies began to pivot away from doing computation work by hand. |
| **Communication** | Documentation and correspondence were transformed by the electric typewriter.Documents and messages were hand delivered within offices. New telephone systems allowed multiple incoming and outgoing lines, programmable buttons for saving phone numbers and intercom abilities for up to 30 extensions. |

### 1985

In 1985, the world’s population was 4.8 billion. The internet’s domain name system was created. A hole in the ozone layer over Antarctica was discovered. DNA is first used in a criminal case.

### 2005

In 2005, the world’s population was 6.5 billion. The first video was uploaded to YouTube. NASA’s Deep Impact space probe impacted a comet in motion.

| **Field** | **Summary** |
| --- | --- |
| **Technology** | Dial-up connections are replaced with high-speed connectivity. Smart devices give an increasing number of people access to the internet, especially within urban areas. Businesses become aware of the usefulness of their collected user data. |
| **Data storage** | Large-scale companies began digitizing and storing data within cloud services. Filing cabinets full of documentation were slowly being converted into data archives. Flash drives reached the market, offering a gigabyte of portable storage that could easily fit within a pocket. |
| **Data analysis** | Companies employed data analysis and predictive modeling as primary business elements. Customer interaction yields datasets and data analysis opportunities, creating large repositories of unstructured data. Social media significantly elevated the need for tools, technologies, and analytics techniques to gather useful information from unstructured data. |
| **Communication** | Inter-department collaboration, worldwide email, and instant messaging became available on smart devices. |

### Today

Today the world’s population is around 8 billion. The James Webb Space Telescope is expanding our knowledge of the universe’s earliest moments.

| **Field** | **Summary** |
| --- | --- |
| **Technology** | Augmented and virtual reality offer interactive experiences, enriched by computer-generated information. The Internet of Things (IoT) connects a growing number of devices and systems. Edge computing is becoming a common practice; it is a way of distributing computational tasks over many nearby processors (i.e., computers) that is good for speed and resiliency and does not depend on a single source of computational power. |
| **Data storage** | Quantum computing, a type of computing that employs the collective properties of quantum mechanics, is being developed to increase processing and storage and solve problems that are too complex for classical computers. International efforts will continue to build quantum computing systems with the necessary cooling systems, user interfaces, networks, and storage capabilities. |
| **Data analysis** | Artificial intelligence is helping to speed up the rate of innovation by augmenting human capabilities. This technology resides at the intersection between social and technological issues. Technological advances continue to be inspired by biological and psychological knowledge about how humans process and learn information. |
| **Communication** | Virtual meetings and online messaging have become commonplace. Task management software is helping organizations facilitate the completion of projects, allowing workers across the globe to work asynchronously. |

## Key takeaways

Data science and data analytics are distinct but connected fields that have developed over time in the evolving data career space. Organizations often use these terms interchangeably when hiring, so being able to recognize the similarities and differences between the two professions will help you identify opportunities that are aligned with your particular skill set. That’s why, throughout this course, you will learn more about the data career space as a whole and how data analytics professionals work within that space.

# Assess your readiness for the Advanced Analytics Data Certificate

The following questions are designed to help you assess your comfort level with general data analytics principles and data types. Use them to self-reflect and make your own decisions about the foundational knowledge you need to be successful in this course.

### 1.

Question 1

What is data science?

1 / 1 point

A tool for organizing data elements and how they relate to one another

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

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

A process used to solve complex problems in a user-centric way

Correct

Data science is a field of study that uses raw data to create new ways of modeling and understanding the unknown.

### 2.

Question 2

What is the key difference between qualitative and quantitative data?

1 / 1 point

Qualitative data measures qualities and characteristics; quantitative data measures numerical facts.

Qualitative data is subjective; quantitative data is specific.

Qualitative data is about the quality of a product or service; quantitative data is about how much of that product or service is available in the marketplace.

Qualitative data describes the kind of data being analyzed; quantitative data describes how much data is being analyzed.

Correct

Qualitative data measures qualities and characteristics; quantitative data measures numerical facts.

### 3.

Question 3

Which of the following statements accurately describe wide and long data? Select all that apply.

1 / 1 point

Long data subjects can have multiple rows that hold the values of subject attributes.

Correct

Wide data subjects can have data in multiple columns. Long data subjects can have multiple rows that hold the values of subject attributes.

Wide data subjects can have multiple rows that hold the values of subject attributes.

Wide data subjects can have data in multiple columns.

Correct

Wide data subjects can have data in multiple columns. Long data subjects can have multiple rows that hold the values of subject attributes.

Long data subjects can have data in multiple columns.

### 4.

Question 4

Structured data is likely to be found in which of the following formats? Select all that apply.

1 / 1 point

Audio file

Spreadsheet

Correct

Structured data is organized in a certain format such as rows and columns. It is likely to be found in a table or spreadsheet. To learn about structured data, enroll in the Google Data Analytics Certificate.

Digital photo

Database table

Correct

Structured data is organized in a certain format such as rows and columns. It is likely to be found in a table or spreadsheet. To learn about structured data, review course three of the Google Data Analytics Certificate.

### 5.

Question 5

Fill in the blank: A Boolean data type can have \_\_\_\_\_ possible value(s).

1 / 1 point

two

one

three

infinite

Correct

A Boolean data type can have two possible values.

|  |
| --- |
| The following questions are designed to help you assess your comfort level with data strategy. Use them to self-reflect and make your own decisions about the foundational knowledge you need to be successful in this course. |

### 6.

Question 6

What is the term for the individuals who have invested time and resources in a project and are interested in its outcome?

1 / 1 point

Project sponsors

Stakeholders

Subject-matter experts

Executives

Correct

Stakeholders are individuals who have invested time and resources in a project and are interested in its outcome.

### 7.

Question 7

When collecting data for a study, what are some reasons to consider sample size? Select all that apply.

1 / 1 point

To collect data that represents a diverse set of perspectives

Correct

Considering sample size ensures the data represents a diverse set of perspectives and helps avoid skewed results or inaccurate judgements.

To make sure a few unusual responses don’t skew results

Correct

Considering sample size ensures the data represents a diverse set of perspectives and helps avoid skewed results or inaccurate judgements.

To include as many participants as possible in the study

To eliminate certain segments of a population

### 8.

Question 8

The SMART methodology can be used to ask a question that promotes change. What type of SMART question leads to change?

1 / 1 point

Transformational

Motivational

Action-oriented

Results-focused

Correct

A SMART question that promotes change is action-oriented.

### 9.

Question 9

Which of the following inquiries are leading questions? Select all that apply.

1 / 1 point

In what ways did our product meet your needs?

Correct

Leading questions include: How satisfied were you with our customer representative? In what ways did our product meet your needs? And what do you enjoy most about our service? Leading questions direct the respondent to a particular answer, often because they suggest the answer within the question.

What do you enjoy most about our service?

Correct

Leading questions include: How satisfied were you with our customer representative? In what ways did our product meet your needs? And what do you enjoy most about our service? Leading questions direct the respondent to a particular answer, often because they suggest the answer within the question.

How satisfied were you with our customer representative?

Correct

Leading questions include: How satisfied were you with our customer representative? In what ways did our product meet your needs? And what do you enjoy most about our service? Leading questions direct the respondent to a particular answer, often because they suggest the answer within the question.

How did you learn about our company?

### 10.

Question 10

What are the key characteristics of a metric? Select all that apply.

1 / 1 point

Metrics are used for measurement.

Correct

Metrics are quantifiable data types used for measurement and performance evaluation.

Metrics are quantifiable.

Correct

Metrics are quantifiable data types used for measurement and performance evaluation.

Metrics are unorganized collections of facts.

Metrics can be used to evaluate performance.

Correct

Metrics are quantifiable data types used for measurement and performance evaluation.

|  |
| --- |
| The following questions are designed to help you assess your comfort level with data integrity and data cleaning. Use them to self-reflect and make your own decisions about the foundational knowledge you need to be successful in this course. |

### 11.

Question 11

Which type of bias is the tendency to construe ambiguous situations in a positive or negative way?

1 / 1 point

Interpretation bias

Confirmation bias

Observer bias

Cultural bias

Correct

Interpretation bias is the tendency to construe ambiguous situations in a positive or negative way.

### 12.

Question 12

Before completing a survey, an individual acknowledges reading information about how and why the data they provide will be used. What concept does this describe?

1 / 1 point

Transaction transparency

Openness

Consent

Privacy

Correct

This concept is called consent. Consent is the aspect of data ethics that presumes an individual’s right to know how and why their personal data will be used before agreeing to provide it.

### 13.

Question 13

Which spreadsheet tool changes how cells appear when values meet a specific condition?

1 / 1 point

Alternating colors

Data validation

Conditional formatting

Protected ranges

Correct

Conditional formatting is the spreadsheet tool that changes how cells appear when values meet a specific condition.

### 14.

Question 14

Fill in the blank: In a spreadsheet, the SPLIT function divides a text string around a \_\_\_, then puts each fragment into a new, separate cell.

1 / 1 point

mark

substring

delimiter

indicator

Correct

In a spreadsheet, the SPLIT function divides a text string around a delimiter, then puts each fragment into a new, separate cell.

|  |
| --- |
| The following questions are designed to help you assess your comfort level with programming languages. Use them to self-reflect and make your own decisions about the foundational knowledge you need to be successful in this course. |

### 15.

Question 15

Fill in the blank: A programming language is a system of words and symbols used to \_\_\_\_\_ for computers.

1 / 1 point

install hardware

repair infrastructure

write instructions

detect malware

Correct

A programming language is a system of words and symbols used to write instructions for computers.

### 16.

Question 16

What are the main benefits of using a programming language to work with data? Select all that apply.

1 / 1 point

Automate decision-making

Easily reproduce and share work

Correct

There are three main benefits of using a programming language to work with data: Easily reproduce and share work, save time, and clarify the steps of analysis.

Clarify the steps of analysis

Correct

There are three main benefits of using a programming language to work with data: Easily reproduce and share work, save time, and clarify the steps of analysis.

Save time

Correct

There are three main benefits of using a programming language to work with data: Easily reproduce and share work, save time, and clarify the steps of analysis.

### 17.

Question 17

In order for code to work properly, it’s necessary to follow the predetermined structure of the coding language. This includes all required words and symbols, as well as their proper placement. What is this structure called?

1 / 1 point

Syntax

Standard

Script

Symbol

Correct

In order for code to work properly, it’s necessary to follow the syntax of the coding language. This includes all required words and symbols, as well as their proper placement.

### 18.

Question 18

What is the term for programming code that is freely available and may be modified and shared by the people who use it?

1 / 1 point

Common-design

Open-source

Non-dependant

One-access

Correct

Open-source code is freely available and may be modified and shared by the people who use it.

### 19.

Question 19

Data professionals use programming languages to enable which of the following? Select all that apply.

1 / 1 point

Data transformation

Correct

Data professionals use programming languages to enable data transformation, cleaning, and visualization.

Data visualization

Correct

Data professionals use programming languages to enable data transformation, cleaning, and visualization.

Data governance

Data cleaning

Correct

Data professionals use programming languages to enable data transformation, cleaning, and visualization.

|  |
| --- |
| The following questions are designed to help you assess your comfort level with data visualization, dashboards, and sharing insights with others. Use them to self-reflect and make your own decisions about the foundational knowledge you need to be successful in this course. |

### 20.

Question 20

What type of data visualization should be used to demonstrate how often data values fall into certain ranges?

1 / 1 point

Correlation chart

Tree map

Bar chart

Histogram

Correct

To demonstrate how often data values fall into certain ranges, use a histogram.

### 21.

Question 21

A dashboard is designed to share insights about the housing market in a city. What type of data visualization would be most effective at demonstrating how the city’s annual home sales have risen over time?

1 / 1 point

Line chart

Scatter plot

Area chart

Pie chart

Correct

To demonstrate how the city’s annual home sales have risen over time, a line chart would be most effective.

### 22.

Question 22

What type of visualizations enable the data in a presentation to automatically update and change over time?

1 / 1 point

Static

Discrete

Customized

Dynamic

Correct

Dynamic visualizations enable the data in a presentation to automatically update and change over time.

### 23.

Question 23

Why is it more effective to label a data visualization instead of using a legend? Select all that apply.

1 / 1 point

Labels allow for text explanations to be placed directly on the visualization.

Correct

It is more effective to label a data visualization instead of using a legend for several reasons: Labels can be placed near the data, they make the data visualization more accessible, and they allow for text explanations to be placed directly on the visualization.

Labels help keep people’s attention on relevant data by redirecting their focus away from outliers.

Labels can be placed near the data, whereas legends are typically positioned away from the data.

Correct

It is more effective to label a data visualization instead of using a legend for several reasons: Labels can be placed near the data, they make the data visualization more accessible, and they allow for text explanations to be placed directly on the visualization.

Labels make the data visualization more accessible because they don’t rely on the ability to interpret color.

Correct

It is more effective to label a data visualization instead of using a legend for several reasons: Labels can be placed near the data, they make the data visualization more accessible, and they allow for text explanations to be placed directly on the visualization.

### 24.

Question 24

A data visualization reveals two variables in the data that rise and fall at the same time. When variables are related in this way, what is likely happening?

1 / 1 point

Polarity

Correlation

Divergence

Causation

Correct

When two variables in a visualization rise and fall at the same time, this is an example of correlation. Correlation is the measure of the degree to which two variables change in relationship to each other.

### 25.

Question 25

Which of the following are appropriate uses for filters in data visualization tools? Select all that apply.

1 / 1 point

Providing data to different users based on their particular needs

Correct

Filters can be used to highlight individual data points, limit the number of rows or columns in view, and provide data to different users based on their needs.

Highlighting individual data points

Correct

Filters can be used to highlight individual data points, limit the number of rows or columns in view, and provide data to different users based on their needs.

Limiting the number of rows or columns in view

Correct

Filters can be used to highlight individual data points, limit the number of rows or columns in view, and provide data to different users based on their needs.

Hiding outliers that do not support the hypothesis

**If you scored between 90% and 99% on the readiness quiz:**

* You’re somewhat familiar with the fundamental concepts involved in data analytics and can move on to the Advanced Data Analytics Certificate.
* However, we recommend that you review some material from the [Google Data Analytics Certificate](https://grow.google/certificates/data-analytics/#?modal_active=none) based on any questions you missed to refresh your memory of those topics.
* You can review videos, readings, and activities throughout the course to learn more about how this advanced program builds on your current understanding of data analytics.
* You must take each of the weekly challenges, which will count toward the 80% overall score needed to earn the certificate.
* After you complete the weekly and course challenges, proceed to Course 2.

**If you scored between 80% and 89% on the readiness quiz:**

* You have some background knowledge on fundamental concepts involved in data analytics.
* However, we recommend that you review lesson materials from the [Google Data Analytics Certificate](https://grow.google/certificates/data-analytics/#?modal_active=none) based on questions you missed in order to build your foundational knowledge before moving on.
* You must take the weekly challenges, which will count toward the 80% overall score needed to earn the certificate.

**If you scored less than 80% on the readiness quiz:**

* No problem — the [Google Data Analytics Certificate](https://grow.google/certificates/data-analytics/#?modal_active=none) has all the foundational material you need to tackle these courses!  We strongly recommend that you earn your Google Data Analytics Certificate first, then continue to the Google Advanced Data Analytics Certificate.

# Participate in program surveys

During this certificate program, you will be asked to complete a few short surveys. These are part of a research study to understand how effective the certificate is to meet your career training needs. Read further for a summary of what each survey will cover.

Your survey participation is optional but extremely helpful in making this course as effective as possible. All data is kept confidential and is aggregated for review. Your name is separated from your data when it is stored.

There are no right or wrong answers, and your responses or personal data:

* Will not affect your course experience, scores, or ability to receive a certificate or job offer
* Will not be shared outside of our research team unless you give permission to share your contact information with hiring partners

Thanks for your consideration and time!

## Entry survey

Up next, you will have an opportunity to submit a brief survey to help us understand why you enrolled in this certificate program. If you don’t fill the survey out now, you will receive an invitation to fill it out after completing your first lecture or assignment.

The survey will ask about your experiences leading up to this course and the goals you hope to achieve. This is critical information to ensure we can meet the needs of learners like you and continue offering this program in the future.

## Individual course feedback

After you complete the last graded assignment within an individual course, you may be asked to complete a survey that revisits earlier questions and asks what you have learned up to that point in the program. Again, filling out this information is voluntary but extremely beneficial to the program and future learners.

## Certificate completion survey

After you complete the last graded assignment in the final course of the certificate, you will be asked to complete a survey that revisits some earlier questions and asks what you have learned throughout the duration of the program. This survey also asks if you would like to share your contact information with prospective employers. Filling out the survey and sharing your contact information with prospective employers is completely optional and will not affect your course experience, scores, or ability to receive a certificate or job offer in any way.

# Engage with other learners

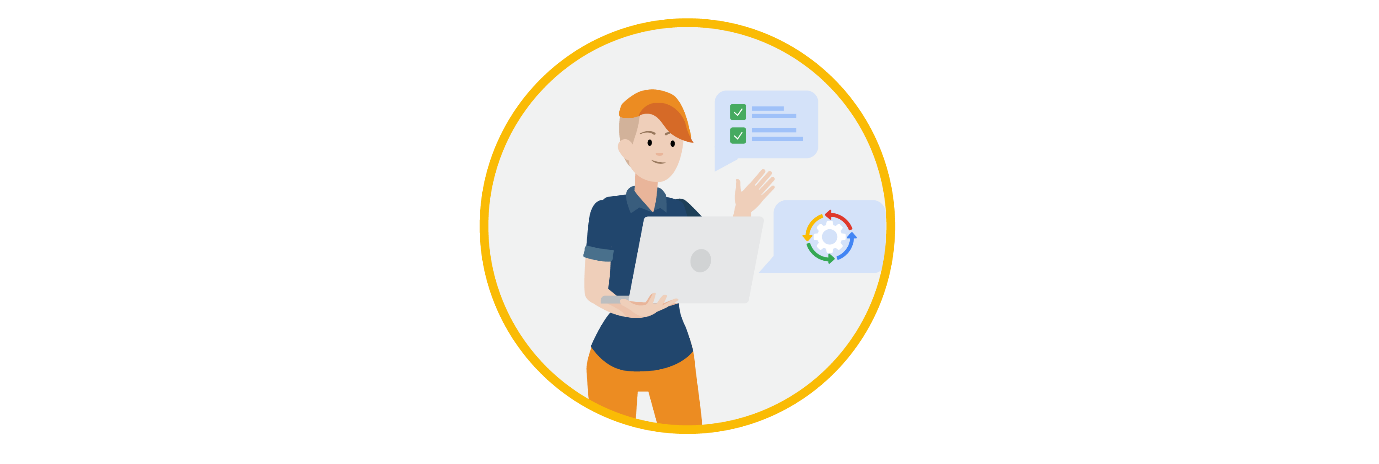
Connecting with your classmates is an essential part of any online course. Get to know each other using the discussion forums and prompts. Discussion prompts are course items that have associated threads in the discussion forums. When you answer a discussion prompt, your response goes to the associated forum, along with the responses of your peers.

Establishing personal interaction with other learners will make your online learning experience much more enjoyable and engaging. And given the probability that many workplaces will have virtual meetings and hybrid or remote work environments, it's also important to begin practicing building relationships and giving and receiving feedback online. We encourage you to use the forums to deepen your learning and peer relationships.

## Share with other learners

Tell everyone a bit about yourself. Then, read some of your classmates' postings. Pick at least two postings that are interesting to you and add a friendly or encouraging response. Go to the discussion forum and click the New Thread button to begin a new thread. Then, go to the Sharing Questions and Concerns discussion prompt to introduce yourself there.

## Update your profile



Additionally, consider [updating your profile](https://www.coursera.org/account/profile), which can be accessed by clicking the Profile link. When classmates find you in the discussion forums, they can click on your name to view your complete profile and get to know you more.

## Upvote posts

When you enter the discussion forum for your course, you will find an Upvote button under each post. We encourage you to upvote posts that you find thoughtful, interesting, or helpful. This is the best way to ensure that other learners will read quality posts in the course. Upvoting also increases the likelihood that important questions get addressed and answered.

## Report abuse

A cartoon of a person sitting at a desk with a computer

Description automatically generated

Coursera's Code of Conduct prohibits:

* Bullying or threatening other users
* Posting spam or promotional content
* Posting mature content
* Posting assignment solutions (or other violations of the [Honor Code](https://learner.coursera.help/hc/en-us/articles/209818863-Coursera-Honor-Code))

Please report any posts that infringe upon copyright, are abusive or offensive, or otherwise violate Coursera's [Honor Code](https://learner.coursera.help/hc/en-us/articles/209818863-Coursera-Honor-Code). You can report posts by using the Report This option found under the menu arrow to the right of each post.

## Follow posts

If you find a particular thread interesting, click the follow button under the original post of that thread page. When you follow a post, you will receive an email notification any time a new comment is added.

## Improve your posts

Course discussion forums are your chance to interact with thousands of like-minded individuals around the world. In any social interaction, certain rules of etiquette are expected and help contribute to respectful and productive communication. The following are some best practices when posting:

1. Stay on topic in existing forums and threads. Off-topic posts make it difficult for other learners to find information they need.
2. Post in the most appropriate forum for your topic, and don’t post the same thing in multiple forums.
3. Use the filters at the top of the forum page (Latest, Top, and Unanswered) to find active, interesting content.
4. Upvote posts that you find helpful and interesting.
5. Be civil. If you disagree, explain your position with respect and refrain from any and all personal attacks. You will be communicating with learners from around the world, with different styles of engaging. Remember to be respectful when discussing concepts or asking for clarification.
6. Make sure you are understood. This is an English-language certificate, but keep in mind that many learners may speak languages other than English as their primary language. Be helpful to one another so that everyone can understand more easily by writing full sentences and avoiding text-message abbreviations or slang. Be careful when you use humor and sarcasm, as these can be easy to misinterpret.
7. If you are asking a question, provide as much information as possible before or after posing your question. For example,  you might write what you’ve already considered, what you’ve already read, etc.
8. Cite appropriate references when using someone else’s ideas, thoughts, or words.
9. Don’t use a forum to promote your product, service, or business.
10. Invite other learners to extend the discussion with an open-ended statement or question. For example, you might write something like, “I would love to understand what others think.”
11. Don’t post personal information about other posters or yourself in the forum.
12. Report spam and spammers.

For more details, please refer to Coursera's [Code of Conduct.](https://learner.coursera.help/hc/en-us/articles/208280036-Coursera-Code-of-Conduct)

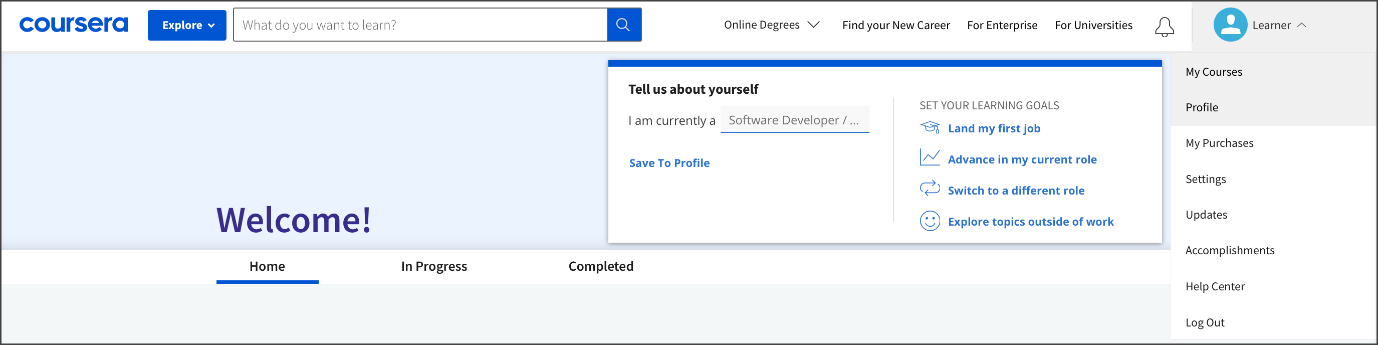
These tips and tools for interacting in this course via the forums were adapted from guidelines originally outlined by The University of Illinois. Curtis, S.. Professional Responsibility and Ethics for Accountants [MOOC]. Coursera.<https://www.coursera.org/learn/ethics>

# Connect with other learners

One of the best ways to find success in this program is to get to know other learners. Connecting with other people makes it easier to share questions and concerns, or brainstorm solutions for any challenges you might face together. Plus, if you don’t understand something when looking back over your notes, you can always reach out by posting a new thread with your question. By working together, you combine your understanding and deepen your knowledge.

## Update your profile

To connect with other learners, it helps to [update your Coursera profile](https://www.coursera.org/account-profile). You can do this by clicking on your name at the top-right corner of this screen, then click the Profile link in the menu. You will be able to add information about your past work experience and education, your career goals, and any details about yourself that you would like to share. This way, when people find you in the discussion forums, they can check out your profile and get to know you better. Only share information that you are comfortable sharing—you can change the privacy option in your settings if you would like to make your profile more private.



## Meet and greet

It also helps to use the [discussion prompt](https://www.coursera.org/learn/foundations-of-data-science/discussionPrompt/HKDxa/connect-with-your-classmates) to introduce yourself, share a brief introduction, and offer help when you’re able. Likewise, be sure to read some of the other learners’ posts. Their interesting life journeys or the things you have in common might surprise you. Leave a comment, a quick hello, or words of encouragement on their posts as you start this adventure together.

Chances are, in addition to taking this course, you are balancing work and personal responsibilities. With all that is going on, it can be easy to isolate yourself from other learners. But when you connect and make friends, learning is a lot more fun. Plus, this is a great way to prepare to work collaboratively in future jobs. Let’s explore the promising future of data analytics together!

# Glossary terms from module 1

## **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 challenge

### 1.

Question 1

Fill in the blank: To gain insights, businesses rely on \_\_\_\_\_ to acquire, organize, and interpret the data that informs internal projects and processes.

1 / 1 point

stakeholders

clients

data professionals

information technology professionals

Correct

### 2.

Question 2

Which of the following statements accurately describe machine learning? Select all that apply.

1 / 1 point

Machine learning requires iteration to achieve desired outputs.

Correct

Professionals use machine learning to express how to perform a task by using explicit instructions.

Machine learning involves training a model.

Correct

Professionals use machine learning to express how to perform a task by using data.

Correct

### 3.

Question 3

Before creating predictive models to identify trends and inform best practices, a company must evaluate its data using what type of measurement?

1 / 1 point

SMART methodology

Attributes

Metrics

Best practices

Correct

### 4.

Question 4

What are some key advantages of the Python programming language? Select all that apply.

1 / 1 point

It can be used to build data-driven applications.

Correct

It was created within the data community.

It emphasizes readability.

Correct

It is a very approachable language.

Correct

### 5.

Question 5

What web-based computing platform can be used by data professionals when interacting with Python?

1 / 1 point

Jupyter Notebook

R Markdown

SQL

HTML

Correct

### 6.

Question 6

A data professional prepares to give a presentation to their colleagues. They want to communicate the story told by the data using charts and graphs made with Tableau. This helps them simplify highly technical information for non-technical stakeholders. Which of the following communication practices does this scenario describe? Select all that apply.

1 / 1 point

Creating a statistical model with code

Explaining data using a graphical interface

Correct

Enriching data insights with visual elements

Correct

Sharing complex data

Correct

### 7.

Question 7

Fill in the blank: Edge computing is a way of distributing computational tasks over a bunch of nearby processors that is good for \_\_\_\_ and resilience and does not depend on a single source of computational power.

1 / 1 point

algorithms

speed

augmented reality

artificial intelligence

### 1.

Question 1

Fill in the blank: To gain insights, businesses rely on \_\_\_\_\_ to acquire, organize, and interpret the data that informs internal projects and processes.

1 / 1 point

stakeholders

clients

data professionals

information technology professionals

Correct

### 2.

Question 2

Which of the following statements accurately describe machine learning? Select all that apply.

1 / 1 point

Machine learning requires iteration to achieve desired outputs.

Correct

Professionals use machine learning to express how to perform a task by using explicit instructions.

Machine learning involves training a model.

Correct

Professionals use machine learning to express how to perform a task by using data.

Correct

### 3.

Question 3

Before creating predictive models to identify trends and inform best practices, a company must evaluate its data using what type of measurement?

1 / 1 point

SMART methodology

Attributes

Metrics

Best practices

Correct

### 4.

Question 4

What are some key advantages of the Python programming language? Select all that apply.

1 / 1 point

It can be used to build data-driven applications.

Correct

It was created within the data community.

It emphasizes readability.

Correct

It is a very approachable language.

Correct

### 5.

Question 5

What web-based computing platform can be used by data professionals when interacting with Python?

1 / 1 point

Jupyter Notebook

R Markdown

SQL

HTML

Correct

### 6.

Question 6

A data professional prepares to give a presentation to their colleagues. They want to communicate the story told by the data using charts and graphs made with Tableau. This helps them simplify highly technical information for non-technical stakeholders. Which of the following communication practices does this scenario describe? Select all that apply.

1 / 1 point

Creating a statistical model with code

Explaining data using a graphical interface

Correct

Enriching data insights with visual elements

Correct

Sharing complex data

Correct

### 7.

Question 7

Fill in the blank: Edge computing is a way of distributing computational tasks over a bunch of nearby processors that is good for \_\_\_\_ and resilience and does not depend on a single source of computational power.

1 / 1 point

algorithms

speed

augmented reality

artificial intelligence

Correct

**Module 2:**

What you’ll learn

- The data career field

- Data analytics in different industries

- Roles within data professions

- Responsibilities of data professionals

“A customer engineer at Google is the bridge between the technical and business components or aspects of corporate America.”

“While you might be new to this new role or this new company, you’ve coming with skills that are repeatable, that are transferable.”

Two categories of data careers

-Technical

-Strategic

Technical data professionals

-Expertise in mathematics, statistics and computing

-Build models and make predictions

-Explore datasets

Strategic data professionals

-Interpret information for an organisation’s operations, finance, research, and development.

-Work aligns with business strategy

## Question

What type of data professionals include expert data analysts, machine learning engineers, and statisticians?

Technical data professionals

Industries

**-Finance:**

\*Assess risks

\*Monitor market trends

\*Reduce fraud

\*Create a more stable financial system

**-Healthcare**

\*Process clinical data

\*Support early detection

\*More precise diagnoses

**-Manufacturing**

\*Predict when to perform preventative maintenance

\*Maximise quality assurance

\*Respond to logistical issues

\*Enable clear communication

**-Agriculture**

\*Develop new approaches

\*Improve harvesting technologies

# Profiles of data professionals

You have been learning about data professionals and how valuable their work is to the companies that employ them. You also learned about technical and strategic roles for data analytics professionals. Data analysts in technical roles transform raw data into something useful for decision-making. Strategic data analytics professionals focus on maximizing information to guide the businesses they are working for. In this reading, you will learn more about the data professional career space, and see examples of some general categories of data professional roles.

The classifications of data professional roles presented here are a very general introduction to the careers available within the data career space. When you begin to look for a job, you may encounter entry-level, mid and senior positions within each of these general categories.

## Data scientist and data analyst

Data scientist and data analyst are roles that work directly with data. These professionals gather, clean, analyze, and share insights from data with stakeholders. An increasing number of industries turn to data analysis to create insights that inform various tasks like guide decision-making, identify user preferences, or determine how to use resources more effectively. With more industries looking for data professionals, there is a great chance that you will find one that matches your interests.

### **Key attributes**

* **What they do:** Uncover trends, patterns, and insights from data
* **How they do it:** Employ advanced modeling and statistical analytics techniques
* Entry-level data scientist or analyst positions may not require considerable data analysis experience and be less restrictive about requiring those hired to hold a traditional degree.

### **Sample job titles**

* Data scientist
* Marketing analyst
* Data analyst
* AI analyst
* Business analyst

## Data management and infrastructure

Data professionals that work in data management and infrastructural roles are primarily responsible for the systems that distribute data and maintain its integrity. They work alongside data analytics professionals and help support their work. Their main responsibility is to ensure the functionality of data systems and the compliance with local, state, and federal regulations involving data security and ethics.

### **Key attributes**

* **What they do:** Manage data sources and the overall data infrastructure
* **How they do it:** Work with the tools and databases used to manage data within a business

### **Sample job titles**

* Data engineer
* Technology engineer
* Data manager
* Data steward
* IT architect

## Business intelligence

Data analytics and business intelligence share a lot of commonalities. Both fields have professionals that use data to create insights that inform decision-making. A major difference is that business intelligence is more focused on creating processes and information channels that transform relevant data. Business intelligence professionals create tables, reports, and dashboards that empower stakeholders, giving them access to the data they need to inform the entire decision-making process on a continual basis. These roles often serve as a complement to core data analytics/data science professionals.

### **Key attributes**

* **What they do:** Perform predictive analysis that enables organizations to determine likely future trends
* **How they do it:** Create tables, reports and dashboards that empower their organization

### **Sample job titles**

* BI architect
* BI analyst
* BI solution developer
* BI software engineer
* Data viz & BI analyst

# Additional roles for data professionals

## Product development teams

The professionals in these roles manage a portfolio of customer and stakeholder analytic projects and initiatives. They often manage the analytical strategy for the organization. In these roles, experience is most likely required, and responsibilities are larger and more global.

### **Key attributes**

* **What they do:** Manage analytical strategy within a project team
* **How they do it:** They are less hands-on with data analysis, serving as the person a data scientist or analysts would report to

### **Sample job titles**

* Product manager
* Product developer
* Product lead
* Digital product manager
* Customer product manager

## C-suite

This classification of roles covers high-ranking executives within an organization. The ‘C’ in c-suite stands for chief. In general, there's a trend for the c-suite to build data-driven decision making into their processes. Individuals filling these roles within organizational leadership teams are expected to be familiar with data and analytics.

### **Key attributes**

* **What they do:** Responsible for data and data professionals across an entire organization
* **How they do it:** They are decision makers found at the top end of a company’s hierarchy

### **Examples of job titles**

* Chief marketing officer
* Chief data officer
* Chief analytics officer
* Chief information officer
* Chief data scientist

## Key takeaways

There are a wide variety of roles and responsibilities in the data professional career space. Later, you will learn how these roles work together and collaborate with groups of professionals with diverse expertise on data analysis projects. Having a general understanding of the roles and responsibilities of data professionals can help inform your job search. This information can serve as a reference that can help you understand the needs and expectations of companies as you navigate through job postings. Take inventory of any transferable skills you may already possess from educational or professional experience.

# Where data makes a difference for the future

As you have been learning, the data analytics field is dynamic, spanning a variety of industries. For you and other data professionals searching for new job prospects, there are great opportunities across a spectrum of career fields. As you start to think about your own future role in data analytics, it can help to investigate how data is being used within different industries. In this reading, you’ll consider some examples of industries and how data analytics helps guide them. You’ll also consider the future of data analytics and how the field is still evolving!

There are so many different industries taking advantage of data analytics in so many different ways. Here is just a sample of some of these industries and how they use data:

| **Industry** | **Overview** | **How data is used** |
| --- | --- | --- |
| App-driven business (sharing economy service) | Facilitates users acquiring, providing, or sharing access to goods and services, often through online or app-based communities | * Maintaining functioning mobile applications * Delivering customized content based on user history including discounts * Using machine learning models to send notifications at key times or even locations |
| Automotive | Includes industries associated with the production, wholesaling, retailing, and maintenance of motor vehicles | * Gaining greater control over their supply chains * Improving production line performance, and designing new and more efficient vehicles * Enhancing vehicle safety and new features |
| Cybersecurity | Protects networks, devices, and data from unauthorized access or criminal use and the practice of maintaining confidentiality, integrity, and availability of information | * Locating weak points within networks and systems using predictive analytics * Defending against security attacks * Detecting data breaches through logic, models, and data tools * Improving the ability to identify attacks and respond to them with Artificial Intelligence (AI) |
| Digital marketing | Assists in advertising and promotional efforts of companies using the internet and online technologies | * Translating customer interaction into actionable business data * Predicting user behaviors to personalize content and offers * Identifying patterns and trends that guide innovations * Determining the return on investment (ROI) of marketing efforts |
| Energy | Includes companies that explore, produce, refine, market, store, and transport both renewable and non-renewable energy resources | * Analyzing real-time data from power systems and monitoring devices * Optimizing technologies, monitoring power grids, and predicting failures * Preventing accidents and malfunctions |
| Gaming | Hosts an estimated 2.7 billion gamers worldwide, facilitating the interaction of players across the globe | * Designing world-building and character creation systems * Monitoring character engagement and how the environment reacts to player input * Optimizing game-play by identifying potential new features or upgrades * Regulating in-game purchases and fraud detection systems * Personalizing marketing campaigns |
| Streaming media and entertainment | Provides access to live and recorded content on-demand, delivered via the internet to computers, smart devices, and mobile devices | * Analyzing and monitoring user interactions to better understand customer sentiment * Matching users with advertisers with real-time analytics * Guiding future content decisions * Personalizing marketing campaigns |
| Telecommunications | Primarily involves operating and providing access to facilities for the transmission of voice, data, text, sound, and video | * Assisting the deployment, optimization, and predictive maintenance of telecommunications networks * Optimizing pricing models * Targeting advertisement and incentive campaigns, as well as detecting fraudulent activity * Analyzing customer data to customize subscriber plans |
| Travel and tourism | Encompasses a variety of services from transportation, accommodations, attractions, booking, and much more | * Marketing to individuals based on their previous travel or searched destinations * Directing machine learning systems that can adjust a traveler’s itinerary based on set factors including weather and availability * Generating recommendations based on personal preferences and location-based discounts * Managing reservations and processing transactions |

## Data trends for the future

As you can already tell, data analytics is an emerging field with a wide range of exciting opportunities. And, even more exciting is the fact that big data is getting bigger. The need for people to understand, prioritize, manage, and analyze that information is not slowing down in any industry. Businesses will continue to rely on data-driven decision-making, fueled by both simple trend analyses and more complex techniques like predictive modeling and forecasting.

Additionally, more companies are storing all of their raw data within large repositories accessible across the organization. As companies become more reliant on insights generated by this data, there will be many opportunities for data analytics professionals to use their skills and knowledge to organize that information and make it useful.

## Innovative technologies

Innovations in accessing this data are leading to new approaches in making data interconnected—meaning that there are still new and evolving ways businesses in different industries are going to use data in the future. Technology is also ever-changing and adapting to these new needs. Because of this, there will always be exciting new tools and data solutions to explore.

Artificial intelligence will continue to have a large impact on business, helping to streamline many areas. For example, an increase in sales is understood to be a direct result of forecasting product demand. Artificial intelligence helps companies ensure warehouse supply, keep items in stock, reduce delivery time, and boost operational efficiency through automating processes.

Additionally, artificial intelligence will combine with machine learning, business intelligence, and automation to deliver more personalized services to customers.

Offering additional services will push forward innovation, bringing computer applications and the sources of stored data closer together physically. This concept is referred to as **edge computing**. By closing the gap between data and computation, speed improves. This results in greater support of real-time analytics and the automation necessary to support the increasing number of devices that are becoming linked through the [internet of things](https://en.wikipedia.org/wiki/Internet_of_things).

An increasing number of data analytics tasks will be automated by creating, managing, and analyzing data in edge environments. Artificial intelligence and machine learning systems are only as equitable and inclusive as the people who create and train these systems. You will learn more about how you, as a data analytics professional, will need to take steps to ensure equity in the future.

## Key takeaways

As you progress through your career as a data analytics professional, you will need to stay up-to-date with the latest trends and technologies used across different industries. The one certainty about the future of data analytics is that an increasing amount of data will continue to be generated and that new systems and innovations will continue to be developed, allowing data professionals an opportunity to learn, grow, and develop new skills.

## Nonprofit

A group whose main purpose is to further a social cause or provide a benefit to the public.

Open data

Data that is available to the public and free to use, with guidance on how to navigate the datasets and acknowledge the source.

Hackathon

An event where data professionals and programmers come together and work on a project.

## Question

What type of data is available to the public for free and includes guidance for navigating the datasets and acknowledging the source?

Open data

###### **Transportation**

**Ways data is collected**

* Traffic sensors and cameras
* GPS and mapping apps
* Ride sharing apps

**How data professionals use the data**

* Create predictive analytics models
* Analyze the best way to get from one point to the next
* Determine the impact of development projects

###### **Healthcare**

**Ways data is collected**

* Electronic health records
* Personal health tracking devices
* Clinical trial data

**How data professionals use the data**

* Analyze medical imaging
* Predict genetic factors
* Create machine learning experiments to speed development of treatments

###### **Finance**

**Ways data is collected**

* Real-time analytics of spending activity
* Apps for banking and investing
* Electronic payments

**How data professionals use the data**

* Identify fraudulent behavior
* Facilitate payments
* Provide risk management solutions

###### **Retail**

**Ways data is collected**

* Sales data (online or in store)
* Retailer apps
* Customer loyalty programs

**How data professionals use the data**

* Recommend products
* Optimize inventory and pricing
* Market to individual customers

###### **Restaurants**

**Ways data is collected**

* Refrigerator temperature monitoring
* Reservation apps for customers
* Marketing/promotion responses

**How data professionals use the data**

* Monitor ingredients during storage
* Manage inventory and supply chain
* Anticipate staffing needs
* Gain feedback from customers

###### **Utilities**

**Ways data is collected**

* Sensors in pipeline/equipment deliver real-time data
* Drone data
* Interactive meters

**How data professionals use the data**

* Access usage data by both utility and customer
* Improve detection of risks within physical systems
* Reduce service interruptions

# Test your knowledge: Data-driven careers

Question 1

What type of data professionals are business intelligence professionals and technical project managers?

1 / 1 point

Information technology

**Strategic**

Governance

Technical

Correct

Correct

Business intelligence professionals and technical project managers are strategic data professionals. In their roles, strategic data professionals maximize information to guide how a business works.

Question 2

Fill in the blank: Expert \_\_\_\_\_ explore vast and complex datasets in order to identify worthwhile business initiatives.

1 / 1 point

stakeholders

**data analysts**

data governance professionals

project managers

Correct

Expert data analysts explore vast and complex datasets in order to identify worthwhile business initiatives. In this highly technical role, data analysts help ensure that an organization’s data science efforts are directed as efficiently as possible.

**The top skills needed for a data career**

Data professionals combine a knowledge about how to do practical tasks with an awareness of what makes communication and collaboration successful.

**Interpersonal skills**

Traits that focus on communicating and building relationships.

**Plan**

Communication

**Analyze**

**Construct**

**Execute**

**Active listening**

Allowing team members, bosses, and other collaborative stakeholders to share their own points of view before offering responses.

**Data cleaning**

The process of formatting data and removing unwanted material.

# Ideal qualities for data analytics professionals

You have been learning about skills that can help you enter and excel in the data career space. In this program, you will be building technical abilities, which are necessary before pursuing opportunities in data analytics. Job postings will include a list of the required technical skills prominently displayed. Quite often, organizations will also make note of additional skills and traits that go beyond working with data on a computer. In this reading, you will explore examples of additional skills and traits that employers are seeking when searching for data analytics professionals.

As you begin to search for job opportunities, many employers seek additional skills that are not exclusive to digital fields. In the sections below, you will learn more about these traits through excerpts found within data analytics job postings.

## Being coachable

Coachable individuals are capable of receiving feedback and using that information to make improvements. At the center of being coachable is a positive attitude, and the ability to self-reflect and take steps to grow. People who are coachable usually have a growth mindset, which is a belief that hard work and determination can make them better. As a result, they view feedback from colleagues and supervisors as an opportunity to improve their skill set.

## A passion for data analysis

Employers often seek candidates whose commitment to data analysis extends beyond their professional duties. Volunteering your data skills to help a nonprofit organization is just one example, but it’s not the only way to show your commitment to data analytics. Data analysis is applicable outside of the workplace, but is often not obvious. Community projects, helping a local school organize data, and developing your own side project are a few examples of how you might demonstrate your passion for data analysis outside of the workplace.

Another way to explore your passion for data analysis is to connect with other passionate data analysts and take on data challenges. A great website for exploring data analytics is [Kaggle.com](https://www.kaggle.com/), which hosts an active online community for data scientists and machine learning enthusiasts. Users can collaborate with other users, publish datasets, use GPU-integrated notebooks, and compete with other data scientists to solve data science challenges. Participating in activities hosted by an online data science community like Kaggle can add fuel to your passion for data analysis and provide artifacts for your portfolio.

Employers are seeking passionate candidates. Job postings that identify candidates that have a passion for data analysis often include language like:

* Seeking a passionate data scientist.
* We seek a candidate with deep curiosity directed toward diverse research interests.
* Driven. The prospect of focusing on corporate environmental impact as a data domain excites you, and you are personally motivated.

**Pro tip**: Include in your portfolio your passion for data analysis. Provide examples of passion projects, volunteer work, or analysis outside of employment to relay your commitment to data analysis. If you are conducting data analysis in your free time, it says a lot about your passion.

## Lifelong learning

Earlier, you read about some generational shifts in data workspaces. You may recall that these shifts have included technological advances in computer interfaces, data storage, and the role of data analysis in organizations.  In fact, the one consistent element through all of these eras was change. As you progress in your career, you will continue to learn new techniques, tools, and ways of interacting with other professionals. As you have learned, staying up to date with the latest technologies and techniques is essential for data professionals. Here are samples from actual job posting:

* Candidates must balance their hands-on work with a desire to keep up with trends.
* Seeking candidates with the aptitude and enthusiasm to develop new skills and areas of expertise.
* Wanting those with a drive to learn and master new technologies and techniques.

**Pro tip**: Stay current by reading data-related blogs and attending workshops. Visit business networking websites to connect and learn from subject matter experts.

## Strong interpersonal skills

You may recall that data professionals and business intelligence professionals interact closely; together, they influence an organization's decision-making. Throughout a project, a data analyst will interact with many stakeholders, from different areas of an organization. Often, these interactions will extend to include clients, users, or representatives from other companies. Additionally, data professionals will work closely with a variety of other professionals as a member of a cross-functional team. Examples from job postings include interpersonal skill requirements for candidates and will include these types of statements:

* We seek employees that can build relationships internally to transfer knowledge, consult with fellow data scientists and analysts to guide analysis, and deliver larger projects.
* Wanting to hire those with the ability to negotiate complex and/or sensitive issues; and maintain good working relations.
* Our data professionals value building strong relationships with colleagues and partners.

**Pro tip**: Include interpersonal skills on your list of qualifications on your resume. Revisit the material within this program (including the material on communication you will be introduced to later in this course). Find opportunities to add examples of teamwork, empathy, leadership, mediating, and active listening.

## Communication

Professionals working in data, business, and technology are expected to be fluent communicators. The effectiveness of your work will weigh heavily on your ability to inspire people through the results of your analysis. Job postings can contain requirements like the following:

* Candidates must have excellent communication skills and a friendly, approachable personality.
* You’re a compelling storyteller who can communicate in succinct and inspiring ways to audiences with varied data science experience to influence real world product or feature decisions.
* A storyteller. You know that no data speaks for itself and take pride in the visualizations and narrative that you construct to communicate your careful analysis.

**Pro tip:** Be sure that all correspondences you exchange are professional and free of grammatical and spelling errors. Include any examples of written communication in your portfolio–these may include but are not limited to reflection pieces, executive summaries, or  project proposals. Don’t forget to include writing examples or online blog entries that describe how you have communicated in past situations.

## Problem solver

The ability to resolve problems is an important part of being a data analytics professional, whether it's dealing with incomplete data, resolving issues within an analysis, or finding the best way to communicate your results to your audience. It is important that you also quantify the results of problems that you have solved. Organizations who seek data professionals have included problem-solving in their job descriptions, such as:

* Seeking data professionals who have the capability and the strong desire to solve problems from concept development, customer engagement, and technology transition.
* Candidates must be persistent and have excellent analytical and problem-solving skills.
* A big thinker. You start with "why" when approaching a new problem, and are always wanting to generalize, synthesize, and summarize.

**Pro tip**: Adjust your past working responsibilities into tasks that you were able to achieve. This can be accomplished by stressing the end results of your actions. Detail how your action or task had a direct impact on the organization. If there is a measured or empirical amount of change associated with your actions, be sure to include that. Example: Analyzed data from over 3,000 users to optimize systems, which led to a 32% increase in customer satisfaction.

## Key takeaways

Employers in a variety of industries are eager to hire data professionals who possess more than just technical expertise. You can highlight your non-technical abilities for potential employers in several ways. Discover ways to successfully and effectively communicate your passions. By demonstrating a commitment to professional growth and sharing examples of your ability to solve problems, you can set yourself apart from other candidates.

# Volunteer data skills to make a positive impact

So far, you have been learning how data analytics professionals are making an impact in a variety of industries. You’ve also learned about nonprofit organizations, and how their efforts are making a difference for so many in need. By volunteering their expertise and time, data analytics professionals contribute to many projects that help nonprofits benefit communities all around the world. In this reading, you will learn more about how you can contribute to global efforts, including  your local community as a data professional.

## Nonprofits

People hoping to further a social cause, or provide a benefit to the public, often create nonprofit groups. The main mission of a nonprofit group is to fill a need or gap that the private sector is not addressing adequately. There are rewarding and inspiring opportunities for data professionals in the nonprofit sector. At the global level, groups like [Children International](https://www.children.org/), [Doctors Without Borders](https://www.doctorswithoutborders.org/), and [Care International](https://www.care-international.org/) lead humanitarian efforts across the globe. In addition to these large-scale efforts, there are many nonprofit organizations whose efforts are regionally specific, like [Christopher’s Promise](http://www.christopherspromise.org/) based in the Columbus, Ohio area.

You can apply your data skills to help nonprofit groups become more effective, allowing them to anticipate and respond to their greatest areas of need.

You may already be aware of nonprofits that fit within your areas of interest or operate in your community. One way that these organizations compensate for their lack of resources is by finding volunteers. Often, regional or local organizations lack the resources to offer full time employment for marketing professionals or data analysts. As a data analytics professional there are many ways that you can make a valuable contribution to nonprofit organizations. In the event that you do not have a specific nonprofit organization in mind, you can research through [Charity Navigator](https://www.charitynavigator.org/), the largest and most trusted online nonprofit evaluation and rating tool in the world. Here you will find ratings on nearly 200,000 nonprofits.

## An example of a community-based nonprofit organization

[Christopher’s Promise](http://www.christopherspromise.org/home.html), a grass-roots nonprofit based in Columbus, Ohio, is a representative example of the types of community-based nonprofits that are making an impact all over the world. This organization was inspired by the efforts to create a bicycle for a fifteen-year old who faced the challenges of Neurofibromatosis, Cerebral Palsy, and Optic Glioma. Since 2011, this nonprofit has facilitated the design and construction of adaptive bicycles for hundreds of children, thanks to outstanding support from the local communities in the central Ohio area.

Christopher’s Promise (2022) mission statement is  “To allow all kids, despite physical limitations, the ability to experience the same hallmark childhood memories as their peers. Helping kids, be kids.”

## Volunteering data skills to nonprofits

As a data analytics professional, your skills can make a great impact on charity and nonprofit organizations within your own community or for a community whose needs resonate with you. Many nonprofits are highly dependent on volunteers for their activities due to limited resources. This situation is especially challenging for community-based initiatives that often lack the funding to hire data analytics professionals. Therefore, offering to volunteer with a local nonprofit is an excellent way to gain experience and support a cause.

### How can data analytics professionals help nonprofits?

Data analytics professionals can help nonprofit organizations in a variety of ways. Below are a few examples.

#### Fundraising

Using data analytics, you can predict with high accuracy which prospects are most likely to donate. In order to maximize return on marketing and outreach expenses, nonprofit organizations can use data analysis to prioritize outreach and reduce overall marketing expenditures.

#### Marketing

There is valuable information in a nonprofit’s data that can benefit marketing efforts. Data analysis can help determine a nonprofit’s target audience and gather insights such as donation methods, contribution history, and other demographics. As a result, the nonprofit can help guide communication with potential donors and determine the effectiveness of campaigns.

#### Monitoring activities

Data analysis can help nonprofits monitor resources, expenses, and daily operational needs. A nonprofit will often need assistance in keeping track of donated materials. For example, a database of donations to a food bank could help the organization determine specific food items being collected and inventory the expiration dates of food items for shelving, cycling, and distribution. Using a system of real-time inventory in conjunction with predictive data analytics could anticipate shortages and identify those donors who have helped provide these items in the past.

Data analytics has a lot to offer a nonprofit organization. Gaining deeper insight into the organization’s efforts can help them to operate in a more effective and efficient way. Analyzing data can improve various nonprofit functions, such as allocation of resources, recruiting and retaining supporters, fundraising, and conducting research. Data analysis also can help reveal patterns and power dynamics that can be used to inform decision-making. In addition to identifying individuals and communities in need of assistance, predictive modeling can also be used to identify potential recipients of a nonprofit's services.

## More opportunities to volunteer

There are organizations who are helping to pair data professionals with projects that are making a difference all around the world.

DataKind is a global nonprofit that harnesses the power of data science and artificial intelligence (“AI”) in the service of humanity. Using data analysis, this organization guides restorative efforts in underserved communities. DataKind brings together volunteer experts from academia and industry to design innovative solutions to tough social challenges. They help social organizations apply their data to predictive analytics, machine learning algorithms, and AI in a way that both increases impact and is sensitive to ethical considerations. Visit the [DataKind website](https://www.datakind.org/join-us/volunteer/) to find out more about their mission and how to become a volunteer.

Another option for putting your data skills to good use are hackathons. A hackathon is an event where programmers and other data professionals come together and collaborate on a particular project. The goal is to create a solution to an existing problem.

There are many hackathon events scheduled across the world each year. These events can take place at a specific location or remotely across the globe. Hackathon events are commonly built around a central theme or question, such as tracking industrial emissions, promoting healthy food options, or how we can improve mental health for cancer patients. If you are interested in finding listings of current hackathons around the world, visit [Devpost.com](https://devpost.com/).

## Key takeaways

Donating your time to a nonprofit is a great way to gain experience, and it can be mutually beneficial for you and the nonprofit. The nonprofit benefits from your data analysis expertise, and you get valuable experience working with data that can be added to your professional portfolio.

Working with a nonprofit organization or participating in a hackathon event introduces you to different types of projects, enhances interpersonal and communications skills, and adds to your professional list of contacts. No matter what your level of data analytics experience, working with nonprofits and hackathons can be a rewarding side project, or even a philanthropic career option.

## Resources for more information

* [Charity Navigator](https://www.charitynavigator.org/) is the world's largest and most trusted nonprofit evaluator.
* [DataKind](https://www.datakind.org/) helps social organizations identify their data and artificial intelligence opportunities, recruits and manages volunteers, and then sees the solutions are used.
* [Devpost](https://devpost.com/) is a place to build products, practice skills, learn technologies, sign up for competitions, and grow your network.
* [IRS.gov tax exempt organizations listing](https://www.irs.gov/charities-non-profits/search-for-tax-exempt-organizations) is a website to find charitable organizations throughout the United States.
* [Statistics without borders](https://www.statisticswithoutborders.org/) contributes to the common good by providing free statistical, data science, and analytical services.

**Personally Identifiable Information (PII)**

Information that permits the identity of an individual to be inferred by either direct or indirect means.

Examples of PII

-Biometric records

-Usernames

-Social security or national identification numbers

**Aggregate information**

Data from a significant number of users that has eliminated personal information.

“Data gathering is a task managed by humans.”

“That process can be informed by different backgrounds, experiences, beliefs, and worldviews.”

## Question

Fill in the blank: A person’s background, experiences, and beliefs lead to \_\_\_\_\_, which may negatively affect data work.

Bias

**Sample**

A segment of a population that is representative of the entire population.

# Critical data security and privacy principles

You have learned how data analytics can be used for good causes, like assisting nonprofit organizations. Also, you learned that data professionals need to protect privacy within data and remain aware of other considerations, like data bias and making assumptions about data.

As a data analytics professional, you have a responsibility to handle data ethically. Data ethics refers to well-founded standards of right and wrong that dictate how data is collected, shared, and used. Throughout your career you will work with a lot of data. This sometimes includes PII, or **personally identifiable information,** which can be used by itself or with other data to track down a person's identity. One element of treating data ethically is ensuring that the privacy and security of that data is maintained throughout its lifetime. In this reading, you will learn more about the importance of data privacy and some strategies for protecting the privacy of data subjects.

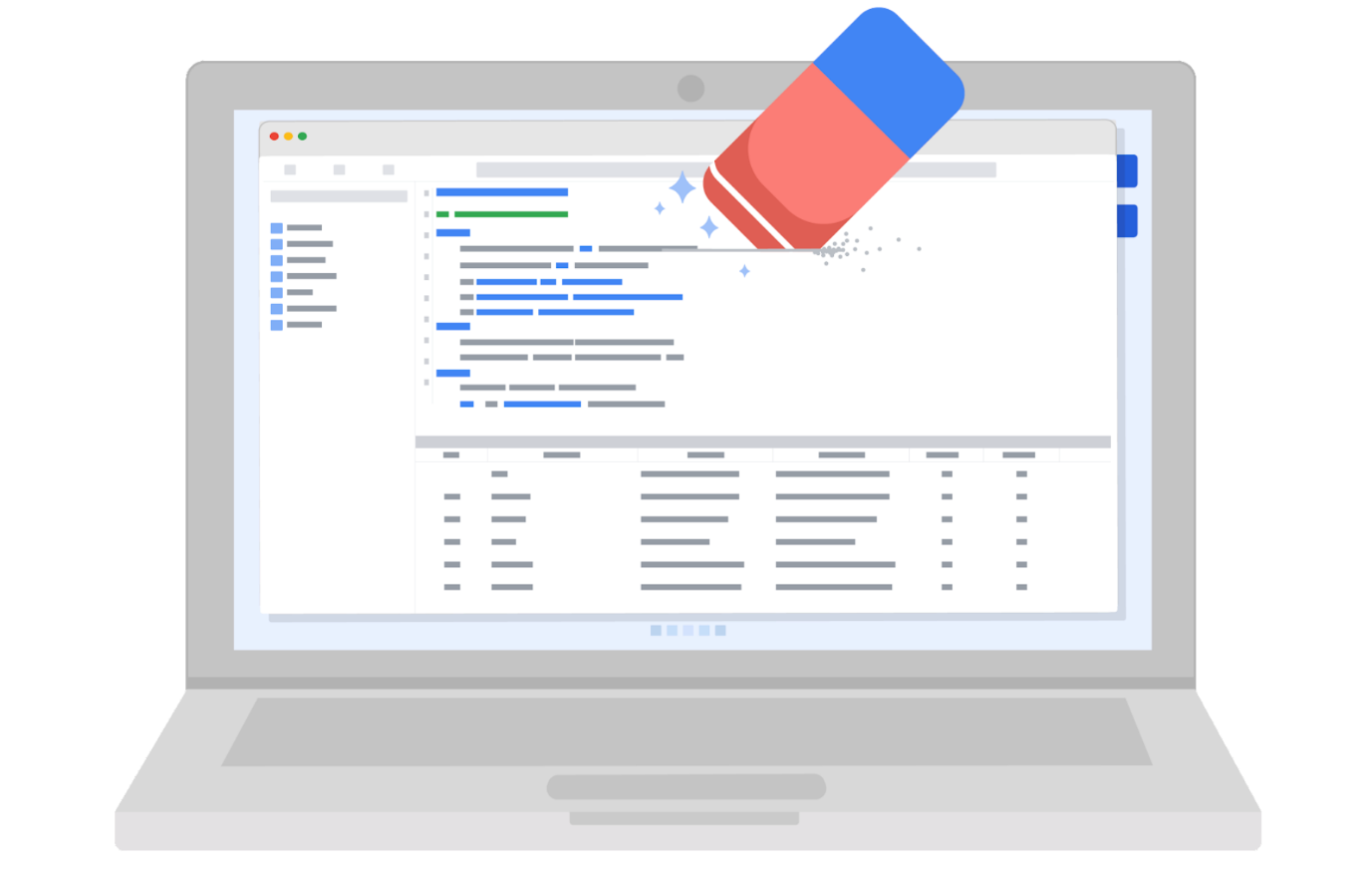
## Privacy matters

Data privacy means preserving a data subject’s information and activity any time a data transaction occurs. This is also called information privacy or data protection. Data privacy is concerned with the access, use, and collection of personal data. For the people whose data is being collected, this means they have the right to:

* Protection from unauthorized access to their private data
* Freedom from inappropriate use of their data
* The right to inspect, update, or correct their data
* Ability to give consent to data collection
* Legal right to access the data

In order to maintain these rights, businesses and organizations have to put privacy measures in place to protect individuals’ data. This is also a matter of trust. The public’s ability to trust companies with personal data is important. It’s what makes people want to use a company’s product, share their information, and more.

## Protecting privacy with data anonymization



Organizations use a lot of different measures to protect the privacy of their data subjects, like incorporating access permissions to ensure that only the people who are supposed to access that information can do so. Another key strategy to maintaining privacy is data anonymization.

**Data anonymization** is the process of protecting people's private or sensitive data by eliminating PII. Typically, data anonymization involves blanking, hashing, or masking personal information, often by using fixed-length codes to represent data columns, or hiding data with altered values.

Data professionals can take additional measures to protect users and their data. **Data aggregation**, for example, is the process of collecting and combining details from a significant number of users in terms of totals or summary. Aggregating data ensures that information contained within datasets is shown in groups; when coupled with other anonymization techniques, data professionals can ensure compliance with data privacy and anonymization standards.

Data anonymization is used in just about every industry. As a data analytics professional, you probably won’t personally be performing anonymization, but it’s useful to understand what kinds of data are often anonymized before you start working with it. This data might include:

* Telephone numbers
* Names
* License plates and license numbers
* Social security numbers
* IP addresses
* Medical records
* Email addresses
* Photographs
* Account numbers

Imagine a world where we all had access to each other’s addresses, account numbers, and other identifiable information. That would invade a lot of people’s privacy and make the world less safe. Data anonymization is one of the ways we can help keep data private and secure!

## Key takeaways

For any professional working with data about actual people, it’s important to consider the safety and privacy of those individuals. That’s why understanding the importance of data privacy and how data that contains PII can be made secure for analysis is so important. We have a responsibility to protect people’s data and the personal information that data might contain.

If you’re interested in learning more about data privacy and ethics, you can check out [the Google Data Analytics Certificate program’s section on bias, credibility, privacy, ethics, and access](https://www.coursera.org/learn/data-preparation/home/week/2).

# The practices and principles of good data stewardship

As you have been learning, all data professionals are responsible for ensuring the quality, integrity, accessibility, and security of data. Data stewardship is the practice of ensuring that data is accessible, usable, and safe. Making data stewardship a normal part of your work habits will benefit everyone who relies on your analysis, both inside and outside of your organization. In this reading, you will learn more about data stewardship and receive some best practices that can assist in guiding your career in data analytics.

## Respect privacy

Earlier in this course, you learned about Information that permits the identity of an individual to be inferred by either direct or indirect means. This kind of information is commonly referred to as personally identifiable information or PII. When users share personal information, they are putting a high level of trust into an organization. It is the responsibility of all who have access within the organization to help protect the privacy of their users. As a data analytics professional, it is important to be thoughtful about any personal data and exhibit great care to protect it. In different parts of the world, laws are in place to guide best practices for data privacy. Laws provide a foundation for best practices as you grow in knowledge and experience on how to support and sustain privacy. One of your responsibilities as a data professional will be to stay up to date with any change in data laws and regulations that govern data. Depending on your organization’s location or industry considerations, there may be additional regulations and policies in place. Here are a couple of regional examples:

* General Data Protection Regulation or [GDPR](https://gdpr.eu/) (European Union law):
  + The GDPR is described on their website as the toughest privacy and security law in the world. It imposes obligations onto organizations anywhere, so long as they target or collect data related to people in the European Union.
* Lei Geral de Proteção de Dados Pessoais or [LGPD](https://www.gov.br/cidadania/pt-br/acesso-a-informacao/lgpd) (Brazil’s general law for the protection of personal data):
  + The LGPD is a data protection law that governs how companies collect, use, disclose, and process personal data belonging to people in Brazil. LGPD applies to companies that process data about individuals in Brazil.
* The California Consumers Privacy Act or [CCPA](https://oag.ca.gov/privacy/ccpa) (Privacy rights for California consumers):
  + The CCPA gives consumers more control over the personal information that businesses collect about them. These regulations provide guidance on how to implement the law.
  + Additionally, states like Virginia, Colorado, New York, Utah, and Connecticut have enacted similar legislation to protect consumer privacy in their states.

## Be cautious of unintentional harm

Data analytics is expanding its influence across an increasing range of industries. Companies are using the results of data analysis to make informed decisions. Many of these decisions have the potential to impact people across a broad range of social and economic factors. It is good practice to continually strive to produce information that is accurate, while respecting cultural and social norms.

Due to the global marketplace, decisions play out differently in different cultures. Taking these issues and considerations into account is very important for the executive team of an organization. Also, companies are known to take a position on particular politicized social and cultural issues, and these can be reflected in their policies. As a data analytics professional, you must be cognizant of your company’s policies. When presented with challenges, it is best to seek guidance from leadership within your organization on how to navigate.

## Avoid creating or reinforcing bias

You have learned about bias within data and how it can have an impact on your analysis. Identifying bias is not always simple. A good practice when working with data is to keep in mind that data gathering is a task managed by humans–and that process is informed by people from different backgrounds, experiences, beliefs, and worldviews. These and other types of biases can affect the data and the results, which in turn can have an impact on business decisions. You will learn more about bias within data as you progress through the program.

## Consider inclusivity

Often in your role as a data analytics professional, you will have access to data collected in a variety of ways. You will need to consider whether the methods of data collection have excluded information from particular populations. Inclusionary approaches can expand how any organization collects and analyzes data. Building diverse research teams, communicating clearly with user communities, and engaging in careful and critical analysis that considers equity and inclusion benefits all stakeholders.

## Uphold high standards of scientific excellence

The processes and technology that you will interact with as a data analytics professional are deeply rooted in the scientific method. As you continue in your data professional journey, embrace inquiry, intellectual discussion, and collaboration. Invite feedback and assess feedback. Remember, artificial intelligence still depends heavily on the instructions provided by data professionals. The more time and consideration that goes into the process of data analytics, the better the results.

Different industries have different standards. In your role as a data analytical professional, you will need to be aware of the standards for the industries you are working in. Each industry will have its own standards based on industry conventions.

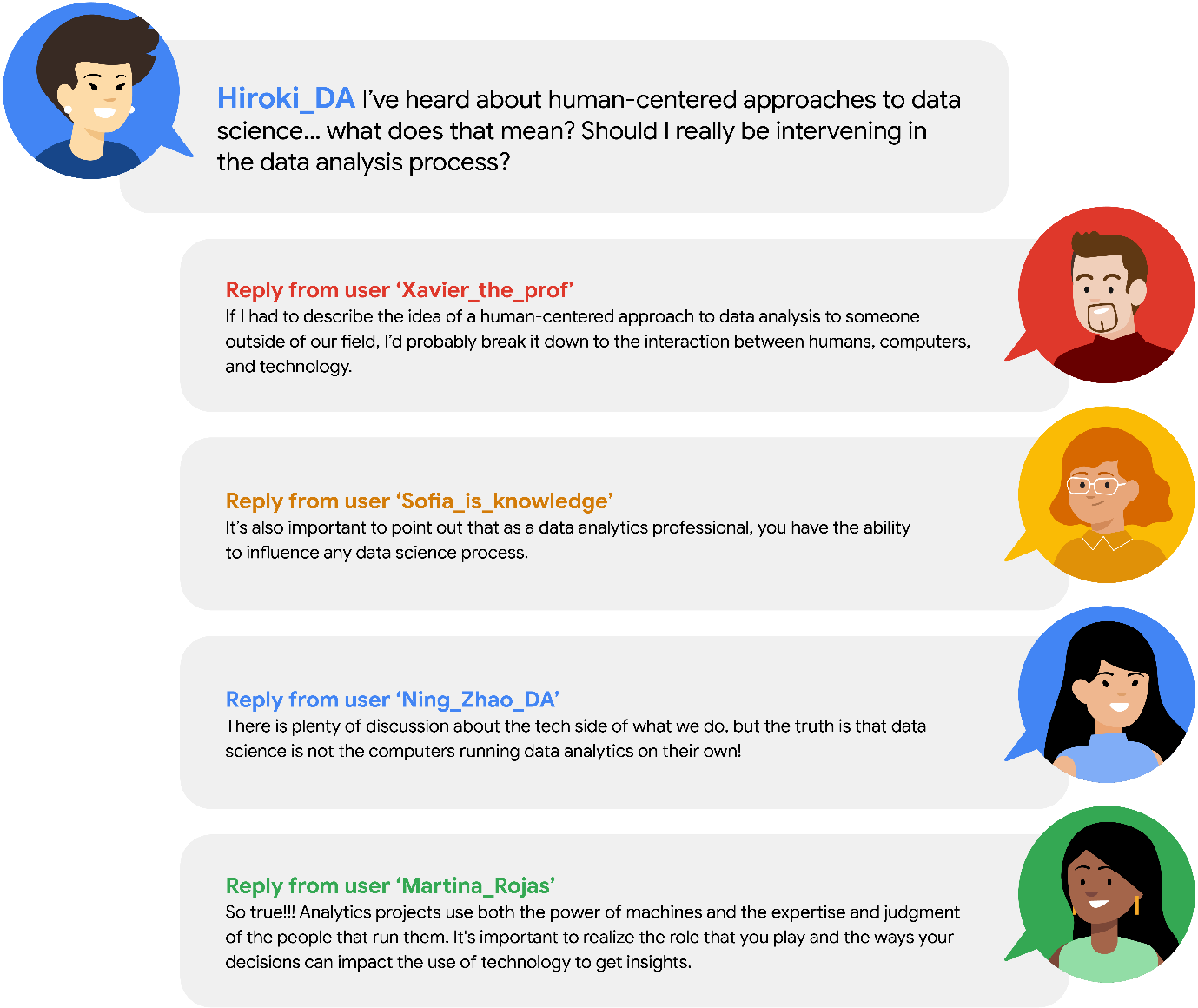
Conventions that work well in the transportation industry may not necessarily be as high of a priority for the healthcare industry. For example, in transportation, data is collected to create predictive analytics models to analyze the best route based on traffic patterns. In the healthcare industry, data is analyzed in medical imaging, predicting genetic factors, and speeding up the development of treatments.

## Data stewardship and ethics conversations

As you progress through your career as a data analytics professional, you will need to consider issues of ethical concern. For example, you may encounter situations where you address questions of bias or need to protect user data and personally identifiable information (PII). When these types of questions arise, many seek guidance and support from online communities of data professionals who have dealt with similar issues. The following graphics present scenarios involving these kinds of issues. You can also find text alternative versions of these conversations in the [Data stewardship and ethics conversations transcript.](https://docs.google.com/document/d/1REAcRzABhtysbB4Nt8h7Gyqu3Sjv4cLfNC-7ETzauSg/template/preview)









## Key takeaways

Data stewardship is the responsibility of every data professional. This responsibility goes beyond interactions with the data. By conducting your work in ways that are socially beneficial and inclusive, you will increase your ability to identify human bias. Guide your efforts through scientific and ethical principles and stay aware of possible bias throughout the data analysis process.

# Test your knowledge: Data career skills

### 1.

Question 1

What are traits related to communication and relationship-building?

Interpersonal skills

### 2.

Question 2

Fill in the blank: A data professional practices \_\_\_\_\_ when they allow others to share their points of view before offering a response.

active listening

### 3.

Question 3

What is the process of formatting data and removing unwanted material?

Data cleaning

“Data tasks and responsibilities are dependent on an organization’s data, team structure, and how they make use of insights and analytics.”

This program refers to the field as a career-space.

Common data professional job titles

-Data Analyst

-Data Scientist

Data professional responsibilities

-Look for patterns and trends within big datasets

-Uncover the stories inside data

-Help guide decision making

-Translate key information into visualizations

-Help guide decision making

Roles that use data and analytical skills

-Data engineer

-Insights or analytics team manager

-Business intelligence engineer or analyst

Data Engineer responsibilities

-Make data accessible

-Ensure data ecosystem produces reliable results

-Deal with infrastructure for data across enterprise

Insights or Analytic Team Manager responsibilities

-Supervise analytical strategy of organization

-Manage multiple groups

Business Intelligence Engineer (or Analyst)

-A highly strategic role focused on organizing information and making it accessible.

# Build the perfect data team

You have been learning about the role of data analytics professionals within organizations and the work they perform. You’ve also examined some general classifications for data professional roles.  Additionally, you explored what a typical data analytics profession might experience during the initial days in a new position. Large organizations often have more data needs, requiring the creation of teams to take on projects. In this reading, you will learn more about how organizations build data analytics teams and look at some best practices when constructing them.

## **Designing data analytics teams**

The idea of companies creating teams of data professionals developed over time and largely out of necessity. Originally, the responsibilities of data collection, management, and analysis fell onto the CIO (chief information officer) of the company. CIOs in the past would typically hand the responsibilities over to the IT (Informational technology) team.

Gradually, organizations began separating data functions as the wealth of information stored grew. Additionally, new technologies emerged and data-related tasks became more specialized.

### Data professional profiles within large organizations

Earlier in this course, you were introduced to some data professional profiles.

* Data Scientist
* Data Analyst
* Data Engineer
* Analytic Team Manager
* Business Intelligence Engineer

These profiles are very generalized categories that can help you focus your eventual job search. In reality, many employers are looking for more than data professionals. They are looking for project stakeholders, dependable team members, and great co-workers.

### Building the team

The data needs of larger companies requires the creation of a team of skilled professionals. The members of these teams each have a specialized area of expertise. Some will come from business backgrounds, some with project or staff management experience, while others will have more technical skills. While there is no checklist available for companies to follow when putting together their team of data professionals, the needs of their organization can help guide them.

Creating a team of data analytics professionals is very similar to how sports franchises put together championship-level rosters. Coaches and general managers are always looking to upgrade and enhance the capabilities of their teams. While having a superstar athlete can help deliver a consistent performance at a certain position, their abrupt departure can compromise the team’s overall performance. A better approach is to develop a well-balanced and collaborative team.

Employers want to hire someone who has more than data analytic skills. They are selecting the best fit for their organization. Candidates who bring additional experience and skills beyond data analytics are most often seen as most favorable. Strong interpersonal and communication skills, experience working in business, or within a team dynamic can enhance data analytics. There is a very good chance that you already possess some of these skills.

## Five principles for data team building

After you join a data team, there are still challenges to building a data-driven organization. Below you will find five principles that can guide any organization to becoming more effective when facing the challenges of data analysis, regardless of the structure of your data teams within the company.

### 1. Adaptability

Data platforms, networks and storage options need to allow flexibility. Each data professional will have their preferences as to the tools and their approaches to analysis. Remote and on-site employees need to have access and the ability to work with all data and use the tools of communication they feel are the most productive.

Organizations need to facilitate a data platform that allows open access to resources for all users. Instead of specifying specific software solutions, organizations can expand their pool of candidates by allowing individuals to integrate the tools and applications they are most comfortable and experienced with. Flexibility invites a wide range of experience and enhances data teams, adding experience and additional perspectives.

Just as organizations need to be adaptable, those seeking opportunities in data fields should be committed to learning new skills and technologies. Data workplaces are undergoing constant change, as you learned earlier. As a data analytics professional, you will continue to learn and grow as newer technologies or regulations emerge.  Accept the challenge presented by new circumstances and let yourself feel energized when presented with opportunities for professional growth.

**Pro tip:** Keep your desire for learning. Expand your knowledge through online data science communities and educational opportunities.

### 2. Activation

Access to data analysis results require someone with the background and experience of a data analyst. Even small-scale data operations require skills that go beyond the scope of other professionals within an organization.

To become more data literate, organizations should cultivate new habits and integrate them into their daily work routines. Dashboards and other accessible interfaces can help promote the use of data analysis, enabling wider usage of data analysis, promoting communication and fostering cross-departmental collaboration.

Vital to the success of an organization's efforts to transform its internal habits and culture are the interpersonal skills of its employees. When a company is seeking to expand data literacy, its data professionals play an important role. There will be opportunities to help promote understanding among peers and colleagues. The time spent answering questions and promoting understanding will help others appreciate you and your contributions to the organization.

**Pro tip:** Maintain positive professional relationships through effective communication. Your ability to share insights is just as important to an organization as your analyzing skills.

### 3. Standardization

An organization needs to set criteria for the standardization of data practices and procedures. Standardization helps to promote best practices, and communication and transferability of information between teams. When users can share optimized code and other assets it saves development time and streamlines projects. An organization that builds a culture of collaboration embeds best practices into work behaviors.

Identifying candidates with capabilities beyond data analysis is essential to developing and standardizing best practices. To develop best practices, organizations need employees that contribute innovative solutions to problems.

**Pro tip:** Become a problem solver. Obstacles within a project can turn into opportunities for innovation, which can transform an organization.

### 4. Accountability

Data analysis is a complex and dynamic process that requires a high level of accountability. To promote responsibility, organizations need a ‘paper trail’ that allows examination of their entire process.

Accountability adds transparency, explainability, and security to data teams and projects. It also helps to eliminate layers within an organization, while aligning business goals and customer values. Transparency in workflow allows organizations to answer specific questions about the data analysis process. Data analytics professionals can use project metadata to examine and communicate specific elements within their prediction models, giving data teams the collaborative ability to make adjustments with more precision.

One way to promote accountability is to extend your ability to communicate throughout your workflow. Within data analytical projects, the transfer of ideas extends beyond individual correspondence. It also includes information in project notes, records kept within projects, and proper tagging within metadata. Individual processes and ideas are the incubators of standards of practice and communicating these ideas effectively can improve an organization’s accountability.

**Pro tip:** Data analytics professionals are more than repositories of information. It benefits everyone involved in a project if you are able to communicate your knowledge and observations. Identify what would be most valuable to others and provide the information.

### 5. Business impact

Often, the inability to estimate the impact on the business can block data analysis projects. Organizations are not considering all available data analytical solutions during the planning stage.

Organizations should look at approaching data projects with the widest field of view in the planning stage. This requires a thoughtful approach including considerations beyond data analysis.  For example, the difficulty of integration, commitment of resources, and changes to the project timeline. Businesses need to consider more data solutions options and identify which benefit the project the most.

Broadening the scope of an organization takes time and consistency of results. To achieve this, a data team must become a trusted resource for insight and a positive influence on an organization's decision-making process. An effective communication strategy, strong interpersonal skills, and a track record of problem-solving will earn the trust of the organization.

**Pro tip:** Focus on communicating clearly with stakeholders. Maintain a commitment to consistency between what you have promised and what you will deliver.

## Key takeaways

Organizations carefully consider the individuals they bring into their organization and seek candidates that embody qualities that go beyond data skills. Additionally, companies can influence their adaptability by hiring data professionals that embody the spirit of lifelong learning, effective communication, interpersonal skills, and the ability to solve problems.

# Activity: Organize your data team

# Activity Exemplar: Organize your data team

Here is a completed exemplar along with an explanation of how the exemplar fulfills the expectations for the activity.

***Note:*** The exemplar represents one possible way to complete the RACI matrix. Different groups, organizations, or teams will have their own unique way of assigning roles and responsibilities to team members; these assignments may fluctuate from project to project or depending on other business variables.

## Completed Exemplar



To review the exemplar for this course item, click the following link and select Use Template.

Link to exemplar:[RACI Matrix Exemplar](https://docs.google.com/document/d/1JktfnpgTj6e4gnETsJGSeDf6mFNIBgvAaIQXEirf6eQ/template/preview)

OR

If you don’t have a Google account, you can download the exemplar directly from the following attachment.

[Activity Exemplar\_ RACI Matrix](https://d3c33hcgiwev3.cloudfront.net/3-34WzQgStuaJ428vCshCQ_e5c610fa07df4b2ab614c445b4fa4bf1_Activity-Exemplar_-RACI-Matrix.docx?Expires=1719878400&Signature=L4lYRRPKpSyuHjB87qcucGarv-zSjL04SOG4pcX-11ZPV4hrp5Qcay55gPxIxC4GmSiROVjgtAy0PmD1GQHxQNdBLS24MyaN5VntknHyt7MDyxMkMfIIgfZlU~IwQ4xeNLrjWVR2Ua64q~uh8a0OUeeZxWT4nWXDqV-pOisulkM_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[DOCX File](https://d3c33hcgiwev3.cloudfront.net/3-34WzQgStuaJ428vCshCQ_e5c610fa07df4b2ab614c445b4fa4bf1_Activity-Exemplar_-RACI-Matrix.docx?Expires=1719878400&Signature=L4lYRRPKpSyuHjB87qcucGarv-zSjL04SOG4pcX-11ZPV4hrp5Qcay55gPxIxC4GmSiROVjgtAy0PmD1GQHxQNdBLS24MyaN5VntknHyt7MDyxMkMfIIgfZlU~IwQ4xeNLrjWVR2Ua64q~uh8a0OUeeZxWT4nWXDqV-pOisulkM_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

## Assessment of Exemplar



Compare the exemplar to your completed activity. Review your work using each of the criteria in the exemplar. What did you do well? Where can you improve? Use your answers to these questions to guide you as you continue to progress through the course.



### **RACI matrix #1:**

* **Access to data:** Based on their general roles within an organization, the business intelligence engineer, analytics team manager, and data engineer have all been identified as responsible for providing access to the data. The data scientist, who is more responsible for working with the analytics team to draw insights from the data, is just consulted.
* **Create models to analyze data**: The data scientist, who is primarily concerned with drawing insights from the data, is responsible for this task. The business intelligence engineer and analytics team manager are consulted. The data engineer, who is concerned with infrastructure, is kept informed.
* **Drive insights and recommendations based on data**: Similarly to the previous task, the data scientist is the primary team member responsible for this task.

### **RACI matrix #2:**

* **Ensure data compliance**: Data compliance is part of developing and managing databases, which is the responsibility of the data engineer. Other team members are consulted or informed, but the data engineer is tasked with the actual responsibility of ensuring compliance.
* **Chief data officer**: Although all of the tasks in this project are the direct responsibility of other team members, the company’s chief data officer is accountable for them—meaning this role has the ultimate responsibility for ensuring this task is completed.

## Key Takeaways

In your work as a data professional, you might encounter the RACI matrix. It is a tool that many organizations use to structure their projects. It also helps to outline, communicate, and understand the responsibilities of data analytics professionals and other cross-functional team members. As a data professional, RACI can also help you consider how to structure your communication with other team members about a project. Understanding how data teams are organized will set you up for better collaboration with your future colleagues!

# Test your knowledge: Work in the field

### 1.

Question 1

What type of data professional is responsible for ensuring an organization’s data ecosystem offers reliable results?

Data engineer

### 2.

Question 2

Fill in the blank: Analytics or \_\_\_\_\_ team managers supervise the analytical strategy of an organization.

Insights

### 3.

Question 3

What is a key part of the role of a business intelligence professional?

Organize information and make it accessible

# Glossary terms from module 2

# **Glossary terms from Course 1, Module 2**

**Aggregate information**:Data from a significant number of users that has eliminated personal information

**Artificial intelligence (AI):** Refers to computer systems able to perform tasks that normally require human intelligence

**Data anonymization**: The process of protecting people's private or sensitive data by eliminating PII

**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

**Hackathon:** An event where programmers and data professionals come together and work on a project

**Nonprofit:** A group organized for purposes other than generating profit; often aims to further a social cause or provide a benefit to the public

**Open data**: Data that is available to the public and free to use, with guidance on how to navigate the datasets and acknowledge the source

**Personally identifiable information (PII)**: Information that permits the identity of an individual to be inferred by either direct or indirect means

**Sample:** A segment of a population, often used to infer parameters of the whole population

# Terms and definitions from the previous module

## D

**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

## E

**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

## J

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

## M

**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

## P

**Python**: A general-purpose programming language

## T

**Tableau**: A business intelligence and analytics platform that helps people visualize, understand, and make decisions with data

# Module 2 challenge (Not all questions are listed here)

### 1.

Question 1 (NOT FINAL)

Which of the following are examples of strategic data professional roles? Select all that apply.

-Business intelligence professionals

-Technical project managers

### 2.

Question 2

Fill in the blank: Data professionals come together during \_\_\_\_\_ to create a solution to an existing problem using technology.

hackathons

### 3.

Question 3 (NOT FINAL)

What are some examples of personally identifiable information? Select all that apply.

National identification numbers

Usernames

Biometric records

Company names

### 4.

Question 4

Fill in the blank: A business \_\_\_\_\_ data by collecting information from a significant number of users. This increases the likelihood that the data will be representative of the population as a whole.

Aggregates

### 5.

Question 5

Fill in the blank: A good \_\_\_\_\_ is a segment of a population that is representative of the entire population.

Sample

### 6.

Question 6

Fill in the blank: Artificial intelligence is the development of \_\_\_\_\_ that are able to perform tasks that normally require human intelligence.

computer systems

### 7.

Question 7

At a business, who is responsible for ensuring socially beneficial and inclusive practices, applying scientific and ethical principles, and staying aware of possible bias?

 All data professionals

### 8.

Question 8

A team of data professionals discusses the potential of their personal backgrounds and beliefs affecting their data findings. They establish processes to ensure that they interpret and communicate sensitive information impartially. What does this scenario describe?

Avoiding subtle biases in data work

**Module 3:**

**What you’ll learn**

-Trajectory of the data career field

-Networking and building relationships

-“It becomes valuable when it’s related to decisions or real world actions.”

“A lot of people don’t realize that without any courses in data analytics, they are already data analysts. We’re all already data analysts.”

**Artificial Intelligence (AI)**

Computer systems able to perform tasks that normally require human intelligence.

**Activities covered by the data professions**

-Statistical inference

-Machine Learning

-Data analytics

## Question

What are the most common activities performed by technical data professionals? Select all that apply.

Statistical inference

Data analytics

Machine learning

# Current and future tools

One of the greatest skills a data professional can have is learning how to apply their knowledge of one tool to another tool. Throughout your career, you might discover that different organizations you work for use different tools—and in the field of data science, new and emerging technologies mean that exciting new tools are being developed all the time. This means there will always be opportunities to expand your data science toolkit! In this reading, you will learn more about tools today, including some of the tools you’re going to learn about in this program. You will also explore some of the exciting ways tools are evolving and what that might mean for your toolkit in the future. Finally, you will explore a demonstration that illustrates how you’ll be using some of these tools in the very near future.

## Tools today

| **Tool** | **Definition** | **Examples** | **Transferable skills** |
| --- | --- | --- | --- |
| Spreadsheets | A digital worksheet where data can be manipulated and used for calculations | * Google Sheets * Microsoft Excel | * Data entry * Mathematical calculations * Manage datasets * Task automation * Data manipulation * Data analysis |
| Databases | A collection of data stored in a computer system | * Google Cloud * CloudSQL * Oracle * Microsoft SQL Server | * Database design * Data storage management * Data integrity |
| Programming languages | A system of words and symbols used to write instructions that computers follow | * SQL * R * Python * Java * C++ | * Communicate with computer systems * Write and input commands * Manage datasets * Data manipulation * Data analysis |
| Data visualization | The graphical representation of data | * Tableau * Matplotlib * Seaborn * Google Charts * InfoGram * ChartBlocks | * Communicate data insights * Design compelling visuals * Identify key metrics |
| Dashboards | A tool that monitors live, incoming data | * Tableau * LookerStudio * Microsoft PowerBI | * Communicate data insights * Monitor real-time data * Develop data visualizations * Design filters and custom calculations |

In this certificate program, you will have the opportunity to learn about many common tools data professionals use every day: spreadsheets, databases, query languages, data visualization, programming languages, and dashboards. Understanding the current tool landscape—and how it’s changing—will help you continue to grow your data science skills throughout your career. And understanding how the skills you learn for one tool can be applied to another means that you can adapt and add more tools to your toolkit!

Already, there are so many tools to choose from as a data professional. This certificate program will focus primarily on Python and data visualizations. As you progress in your career, you might find yourself learning new tools, and using your existing skills in new ways. Being able to recognize where tool skills overlap will help you continuously grow your data toolkit now and in the future.

## You in the near future

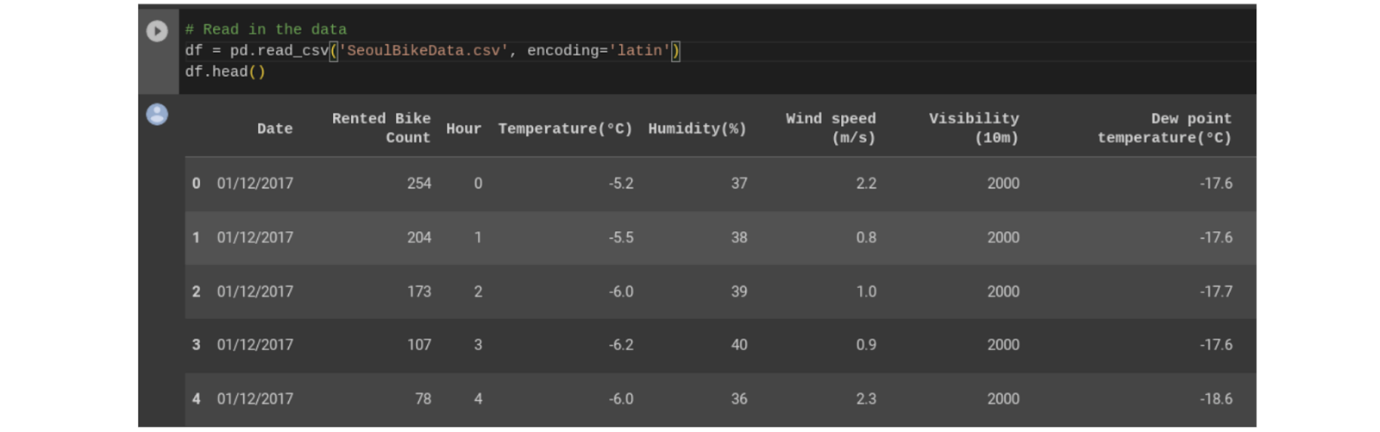
So far in this reading, you have been considering how the skills you’re going to learn in this certificate program will help you use even more tools in the future. As you prepare for your learning journey, you can also think about how you’ll be able to apply these skills soon—not just in the distant future.

This certificate program focuses on some of the most commonly used tools for data analytics and machine learning with Python. More specifically, you will use:

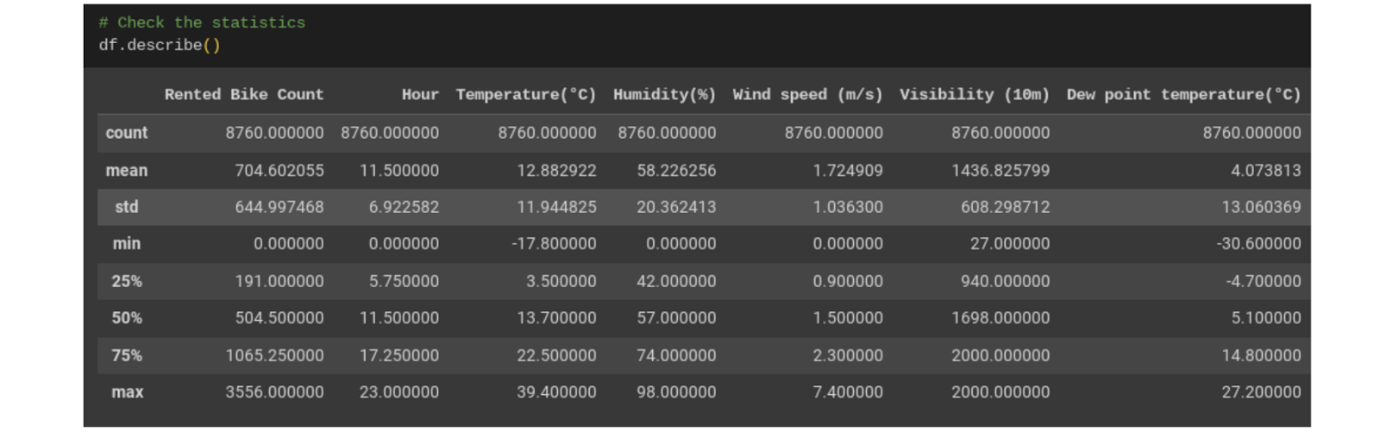
* NumPy and pandas for data processing and manipulation
* matplotlib.pyplot, seaborn, and Tableau for visualizations
* statsmodels for statistical tests and regression modeling
* scikit-learn for building machine learning models

Next, consider the following overview of some of the tools you’ll use to complete tasks in this certificate program.

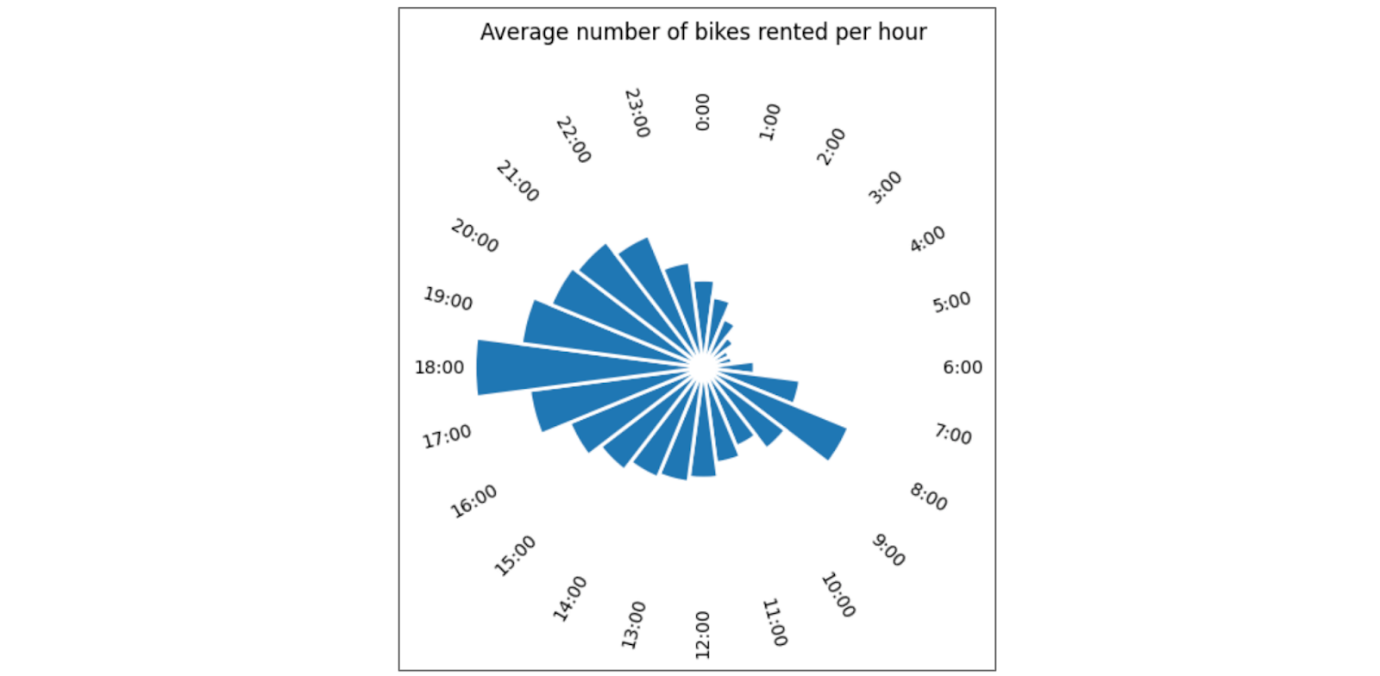
You’ll use pandas to view and manipulate tabular data with Python. In the following example, a comma-separated value (.csv) file is read into a pandas dataframe, of which the first five rows are displayed. A dataframe is basically a table used to organize data. This data is from the [UC Irvine Machine Learning Repository](http://archive.ics.uci.edu/dataset/560/seoul+bike+sharing+demand). It contains the count of public bicycles rented per hour in the Seoul Bike Sharing System, with corresponding weather data and holiday information.

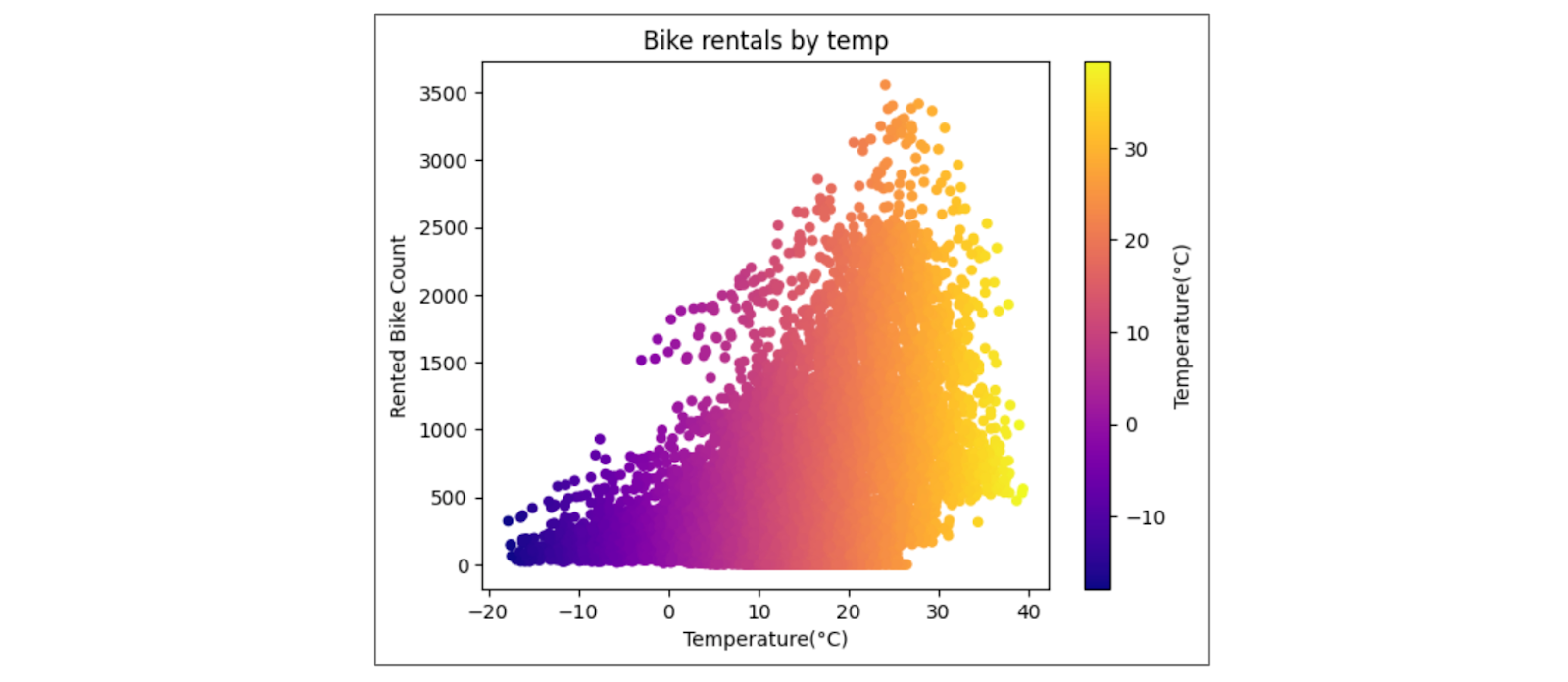


You’ll use NumPy and pandas to perform calculations and get statistics for your data.



You’ll use Tableau, matplotlib.pyplot, and seaborn to create data visualizations.

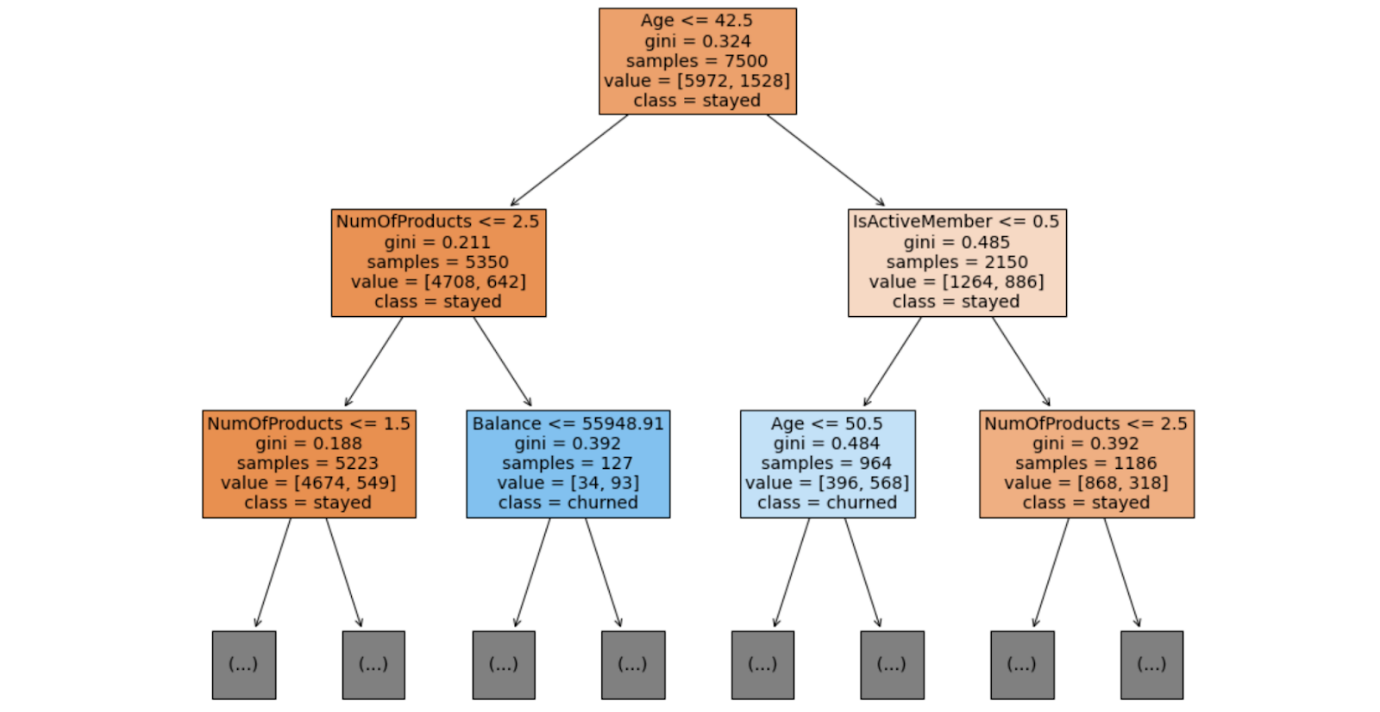




You’ll use statsmodels to practice statistical analysis and linear regression.



And you’ll use scikit-learn to build and analyze machine learning models:



This is just a small sample of the full range of topics you’ll learn about in this certificate. As you gain proficiency with these tools, you’ll be equipped to take on nearly any data task.

## Tools tomorrow and beyond

The world of data science is always growing and evolving—tools you might not have even known about a few years ago can quickly become necessary for professionals working in the field. As you consider the skills you are developing now, it can be useful to consider all the ways you might also use them in the future.

### Artificial intelligence

Artificial intelligence, or AI, refers to computer systems that are able to perform tasks that normally require human intelligence. One of the great benefits of using AI for data science is that it can help provide real-time insights to stakeholders. For example, a business with an e-commerce website might use AI to monitor and deliver insights about how customers use their website in real-time, allowing the team to make quick improvements.

In today's dynamic workplace where leveraging the latest technology is key for productivity and efficiency, having an understanding of AI will boost your career as a data professional. While building a career in this industry, you can start enhancing your AI skills by exploring available AI tools that can assist with your role. One tool that's currently used is Tableau AI. Tableau AI aims to simplify the process of data analysis. This tool has the potential to help data professionals prepare data, reduce repetitive tasks, and suggest appropriate visualizations.

### Machine learning

Machine learning is the use and development of algorithms and statistical models to teach computer systems to analyze and discover patterns in data. Data analysts can train algorithms to analyze large amounts of data that would otherwise take a long time to process. For example, a financial analyst might use machine learning to find patterns in the data that help identify fraud.

One of the most exciting developments in these future technologies is the way they can be used together to automate tasks and provide insights faster than ever.

## Key takeaways

As a data professional, you will continue learning new skills and applying your current skills in new ways throughout your career. Recognizing how skills can be transferable allows you to adapt to different organizations’ needs and evolving technologies. And as you progress through this, you add tools to your data science toolbox that will help you now and in the future!

# How data professionals use AI

Earlier, you briefly considered the role of artificial intelligence in data science. You may recall that **artificial intelligence (AI)** refers to the development of computer systems able to perform tasks that normally require human intelligence. For example, practical applications of AI include voice assistants, autonomous vehicles, automated recommendation systems, and more.

In this reading, you will learn more about the uses of AI for data work, and how AI can help data professionals better understand their data and make more informed decisions. You’ll also learn about the limitations of AI, and the differences between AI and human data professionals.

## The uses of AI for data work

Data professionals can use AI to improve their data analysis, perform essential tasks, and streamline their workflow. For example, data professionals can use AI to:

* Create predictive models to help accurately forecast future events or outcomes.
* Automate time-consuming tasks such as data cleaning, coding, and report writing.
* Analyze extremely large datasets.
* Improve the quality of data by identifying and correcting errors.
* Generate insights from data that would not be obvious to humans.
* Provide guidance on tasks such as choosing the right algorithms and interpreting results.
* Facilitate collaboration among team members.

Data professionals can leverage AI to enhance the quality and efficiency of their data projects, generate valuable insights, and help stakeholders make better business decisions.

### **Conversational AI tools: Gemini and ChatGPT**

Many data professionals now use conversational AI tools to help them analyze their data and boost their productivity. Two of the most frequently used tools are Gemini and ChatGPT.  Gemini was created by Google AI. ChatGPT, also known as Chat Generative Pre-trained Transformer, was developed by OpenAI.

Gemini and ChatGPT are both **large language models (LLMs)** that are trained on massive datasets of text and code. An LLMis a type of AI algorithm that uses deep learning techniques to identify patterns in text and map how different words and phrases relate to each other. This allows LLMs to predict what word should come next. LLMs can generate human-like text in response to a wide range of prompts and questions.

**Note**: This is a general introduction to LLMs. A detailed discussion of the development and computational logic of LLMs is beyond the scope of this course.

Tools like Gemini and ChatGPT can help data professionals in a variety of ways. A data professional might ask Gemini or ChatGPT to:

* Clean a dataset by removing missing values, outliers, and duplicate data.
* Create interactive data visualizations such as dashboards and heatmaps.
* Recommend a specific algorithm for a particular task based on the data professional's input.
* Create a shared document to facilitate a brainstorming session among a team of data professionals.

**Note**: This is a brief list of possible prompts. In another activity, you’ll get a chance to further explore Gemini and discover its capabilities.

A screenshot of a computer

Description automatically generated

## Use cases for AI

Data professionals across industries use AI to help analyze data and generate insights for stakeholders. Here are some examples of how data professionals use AI in specific sectors:

### Finance

* Analyze financial transactions to help prevent fraud and protect customers' money.
* Analyze large datasets of financial data to help identify potential risks and make more informed decisions about investments.
* Analyze historical market data and current market conditions to help generate sound investment recommendations.

### Retail

* Recommend products to customers based on their past purchase history and browsing behavior.
* Track customers' interactions with the retail website to help personalize the shopping experience.
* Analyze sales data and forecast future demand to help optimize the amount of product inventory and reduce costs.

### Manufacturing

* Automate tasks such as welding, painting, and assembly to help improve efficiency.
* Analyze data from sensors and cameras to help identify defects in products before they are shipped to customers.
* Analyze data from production lines to help identify ways to produce more products at a lower cost.

## AI and human data professionals

Data professionals use AI as a tool to help them understand data, make better decisions, and improve efficiency. Like all tools, AI has limitations. Human data professionals possess skills, abilities, and qualities that AI currently lacks. For example:

* Intuition. AI models are trained on data, and they can only make decisions based on the patterns they observe in the data. Humans can use their intuition and personal experience to make decisions that are not explicitly programmed into the AI model. For this reason, it’s important to always verify a model’s output before relying on it.
* Deal with ambiguity. AI models are good at solving problems that are well-defined and have clear parameters. However, humans can identify and understand complex problems that are not well-defined and have ambiguous parameters by considering key details offered in the context of the project.
* Interpersonal communication. AI models can generate reports and presentations, but they cannot communicate with stakeholders in the nuanced way that humans can. Humans can explain the results of their analysis to fit the needs of specific stakeholders, and use their emotional intelligence to address concerns.
* Creativity. AI models are good at following instructions, but they are not imaginative like humans. Humans can be creative in their approach to data analysis, and imagine new and innovative solutions to complex problems.
* Critical thinking. Humans can think critically about their data and identify potential biases and ethical issues. AI models are usually trained on real-world data that contains biases and are therefore likely to reflect those biases in model outputs.
* Leadership. Humans can be leaders, and they can motivate and inspire others. AI may have difficulty understanding the nuances of human emotion, motivation, and communication. This limits AI’s ability to effectively run an organization.
* Factuality. Generative AI models are trained to output text based on patterns in language. Sometimes the model output may be very well-composed and as a result, seem reliable, but may not be factual. As noted above, it’s important to always verify model output.

In the future, product and research teams may develop updates for AI that enlarge its current capabilities. However, human data professionals will continue to play an important role in data science by using their intuition, imagination, and unique experience to solve complex problems.

## Key takeaways

Data professionals can use AI to help automate tasks, make predictions, generate insights, and communicate findings. They can leverage AI to be more productive in their work and more impactful in their organizations. Overall, AI is a powerful tool for data professionals but it is not without limitations. For this reason, human oversight and intervention is critical when working with AI and related tools.

# Activity: Write prompts for Gemini

## Activity Overview



You have learned that many data professionals now use conversational AI tools like Gemini and ChatGPT to help them analyze their data and boost their productivity. Gemini and ChatGPT are both **large language models (LLMs)** that are trained on massive datasets of text and code. LLMs can generate human-like text in response to a wide range of prompts and questions. In this activity, you’ll discover the capabilities of conversational AI by writing your own prompts for Gemini.

To review the role of AI in data work, refer to the reading about [how data professionals use AI](https://www.coursera.org/learn/foundations-of-data-science/supplement/2FH0K/how-data-professionals-use-ai).

Be sure to complete this activity before moving on.

## LLM prompts and best practices



Data professionals can use LLMs to improve their data analysis, perform essential tasks, and collaborate with teammates. Here are some useful prompts for data science workflows:

* Data cleaning. LLMs can automate tasks such as data cleaning and coding. For example, you can ask an LLM to clean a dataset by removing missing values, outliers, and duplicate data.
* Exploratory data analysis (EDA). LLMs can perform exploratory data analysis (EDA) on datasets. For example, you can ask an LLM to create data visualizations, identify patterns and trends, and calculate summary statistics.
* Modeling. LLMs can build and evaluate models. For example, you can ask an LLM to build a machine learning model to predict an outcome, and evaluate the performance of the model.
* Interpreting results. LLMs can interpret the results of models. For example, you can ask an LLM to explain the features that are most important for a model, or generate insights from the results of a model.
* Collaboration. LLMs can help you collaborate with teammates. For example, you can ask an LLM to create a shared document for a brainstorming session with a team of data professionals.

**Pro tip:** Be sure to structure your prompts in a way that makes it easier for the LLM to fulfill your requests and answer your questions.

The following suggestions are best practices for writing prompts for LLMs:

* Be clear and concise in your instructions. It is important to be clear and concise in your instructions so the LLM can understand how to help you. Details are great—just make sure they’re useful and relevant. Avoid giving the LLM unnecessary information.
* Be precise. When posing a question to an LLM, be precise about the input (if any) and the desired output.
* Include a description of LLM’s role. This reinforces the purpose of your prompt. For example, you can tell the LLM to assume the role of a data scientist by writing “Act as a data scientist” or “You are a data scientist.”
* Provide context. Providing context allows the LLM to understand the nuances of the relevant issue and generate more informed responses.
* Try multiple prompts. Trying different prompts can provide different perspectives on a problem and enable the LLM to generate a variety of useful responses.

To help get you started, consider the following specific examples of prompts that data professionals can give an LLM:

* “Act as a data scientist and write a detailed plan for a credit card fraud detection project.”
* “I have a dataset of customer purchases at an online retail store. Act as a data scientist and write Python code for data visualization and exploration.”
* “I have a dataset of customer characteristics and churn for an online video streaming service. Act as a data scientist and create a shared document for a team meeting.”
* “Act as a data generator and use Python code to generate a CSV file that contains mock employee data for a restaurant chain named Fast. The dataset has 100 rows and 5 columns. The columns are name, address, employee\_id, department\_id, email.”
* “Act as a communications expert and share best practices for explaining a data science report to a business executive with no technical background.”

**Note:** LLMsare powerful, but they are still under development – including Gemini, which is still experimental research. As a data professional, it’s important to use your own judgment when interpreting the results. LLMs can generate insights that you may not have thought of on your own; however, it’s ultimately your responsibility to verify the results and make sure they make sense.

## Step-By-Step Instructions



Follow the instructions to complete each step of the activity. Then, answer the questions at the end of the activity before going to the next course item. **Please Note: This activity may be completed on any LLM of your choosing; it is not exclusive to Gemini.**

### **Step 1: Access Gemini**

**Note:** To use Gemini, you’ll need to sign in with a Google Account.

To sign in to Gemini:

1. Go to [gemini.google.com](https://gemini.google.com/).
2. At the top right, select Sign in.
3. Sign in to your personal Google Account.

If you don’t have a Google Account, click on the following link to learn how to create an account: [Create a Google Account](https://support.google.com/accounts/answer/27441?hl=en).

### **Step 2: Give prompts to Gemini based on a workplace scenario**

Review the following fictional workplace scenario. Then follow the instructions for giving prompts to Gemini at different stages of the data project.

Imagine you are a new data professional, recently hired by a healthcare company. The company sells sustainable medical devices to hospitals and clinics in urban communities. Leadership has asked the data team to develop a machine learning model to accurately predict future sales. A powerful model will help company leaders make informed decisions about inventory management, resource allocation, overall sales strategy, and more. As the newest member of the data team, you’re excited to start your first project.

1. Project proposal. To get started, your manager asks you to organize a kickoff meeting with the team to outline the project workflow and timeline. You want to send the invite as soon as possible, and could use some help creating a document for the meeting. **Prompt: Ask Gemini for instructions create a shared document to facilitate a brainstorming session among a team of data professionals.**
2. Data cleaning. Following the team meeting, you help draft a project proposal to outline key deliverables and milestones for the project. Then, the team collects the relevant data. The next step is to clean the dataset. You volunteer to perform this task for the team. The team is using the Python programming language for this project, and you’d like some coding suggestions for data cleaning. **Prompt: Ask Gemini to write Python code to clean data by removing missing values, outliers, and duplicate data.**
3. Data visualization. Now that the team has a clean dataset to work with, the next step is to explore and visualize the data. Your manager asks if you can help create some data visualizations to better understand the relationships between key variables. To get started, you brainstorm with B. **Prompt: Ask Gemini to suggest useful data visualizations for sales data.**
4. Build and test machine learning models. As a new data professional, you are not directly involved in writing code to build and test different machine learning models. However, you want to learn more about the uses of machine learning for data work as this will be an important part of your future career, and will help you better understand the current project. **Prompt: Ask Gemini about the main uses and benefits of machine learning for data work.**
5. Executive summary. The data team successfully builds a model that accurately predicts future sales. Now, the team is ready to share their results and insights with project stakeholders. Your manager asks you to help draft an executive summary for a meeting with company leadership. Before you begin, you want to review best practices so you can create a polished deliverable. **Prompt: Ask Gemini about best practices for creating an executive summary for business executives without a technical background.**

**Note:** Overall, Gemini is a powerful tool for data professionals. However, it’s important to remember that Gemini is not perfect. Be aware of Gemini’s (and other LLMs') limitations. These limitations include the following:

* Gemini is not infallible. Gemini can sometimes make mistakes, such as providing inaccurate information or generating incorrect code.
* Gemini is not an expert in any particular field. Gemini can learn about new topics, but it does not have the same level of understanding as an experienced human data professional.
* Gemini cannot explain its reasoning. Gemini can generate useful output, but it cannot explain why it’s doing what it’s doing. This can make it difficult to understand how Gemini works and to trust its results.
* Gemini can be biased. As an LLM, Gemini is trained on a massive dataset of text and code, and Gemini is likely to reflect the biases that are present in that dataset.

As a result of these limitations, Gemini’s responses may be inaccurate, biased, or insufficient for your purposes. As a data professional, it’s your responsibility to verify the accuracy of Gemini’s output. It’s also your job to modify or supplement Gemini’s output to fit the needs of the specific project you’re working on.

### **Step 3: Experiment with Gemini on your own**

Explore Gemini and discover its capabilities. Feel free to experiment with different types of prompts, use your imagination, and have fun!

### **What to Include in Your Response**



Be sure to address the following elements in your series of prompts to Gemini:

* For Step 2, follow the best practices for the writing prompts outlined above.
* For Step 3, give Gemini at least 5 different unique prompts. Your prompts can include different types of requests, questions, data professional tasks, and workplace scenarios.

### 1.

Question 1

Did you complete this activity?

1 / 1 point

Yes

No

Correct

Thank you for completing this activity! Continually learning how to use new tools such as conversational AI is an important part of being a data professional. Review the quiz question feedback to find out how you did.

### 2.

Question 2

What tasks can Gemini help data professionals perform? Select all that apply.

1 / 1 point

 Lead a data team

Data cleaning

Correct

Gemini can help data professionals perform a number of essential tasks, including data cleaning, exploratory data analysis (EDA), and building and evaluating models.

Building and evaluating models

Correct

Gemini can help data professionals perform a number of essential tasks, including data cleaning, exploratory data analysis (EDA), and building and evaluating models.

Exploratory data analysis (EDA)

Correct

Gemini can help data professionals perform a number of essential tasks, including data cleaning, exploratory data analysis (EDA), and building and evaluating models.

### 3.

Question 3

What are best practices for writing prompts for Gemini? Select all that apply.

0.75 / 1 point

Include a description of Gemini’s role

Correct

Best practices for writing prompts for Gemini include being clear and concise in your instructions, providing context, and including a description of Gemini’s role.

Be clear and concise in your instructions

Correct

Best practices for writing prompts for Gemini include being clear and concise in your instructions, providing context, and including a description of Gemini’s role.

Ask vague questions

This should not be selected

Best practices for writing prompts for Gemini include being clear and concise in your instructions, providing context, and including a description of Gemini’s role. When posing a question to Gemini, be precise about the input (if any) and the desired output.

Provide context

Correct

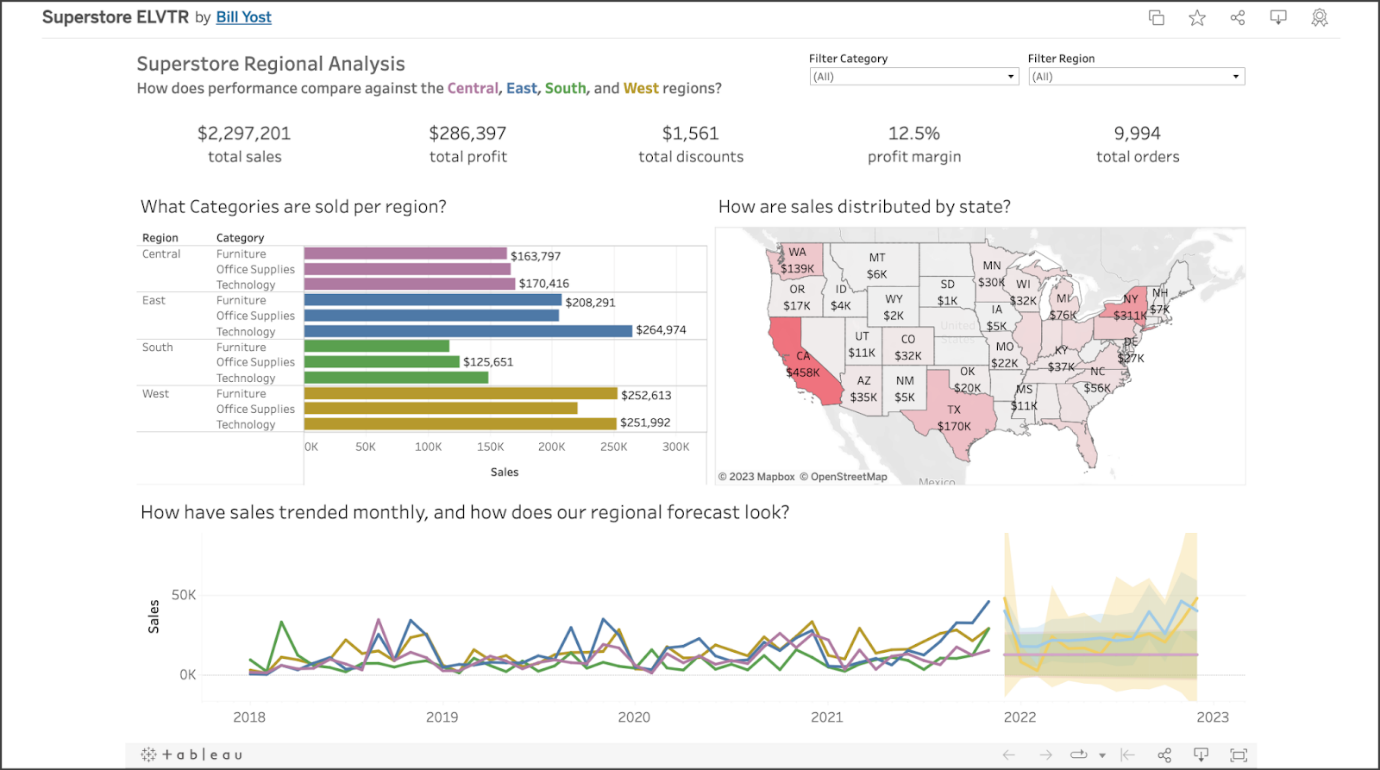
Best practices for writing prompts for Gemini include being clear and concise in your instructions, providing context, and including a description of Gemini’s role.

# The places you’ll go…

By now you’re familiar with many of the tools data professionals use to overcome different challenges in their work. However, as important as these technical skills are for your professional development, creating engaging presentations of your work is just as important. As you complete this certificate program, you’ll create a portfolio of projects that you can showcase to employers, other learners, or to broader audiences. In this reading you’ll find examples of the many ways other data professionals in the field have shared their work, so that you can learn what such presentations entail, get a sense of what appeals to you, and perhaps even inspire you in your own journey.

## **Tableau dashboards**

Bill Yost is a data analyst. One way he leverages his data professional skills is by building interactive data visualization dashboards on Tableau. His [project on office supply store analytics](https://public.tableau.com/app/profile/bill.yost/viz/SuperstoreELVTR/SuperstoreDashboard) is a great example of the dynamic capabilities Tableau offers data professionals.



His  dashboard incorporates several different types of interactive graphs, including a bar chart, a line chart, and a map. You’ll learn how to create these in this certificate program.

Tableau also allows you to create a profile to showcase your work. You can also follow other users and add projects that interest you to a list of favorites.

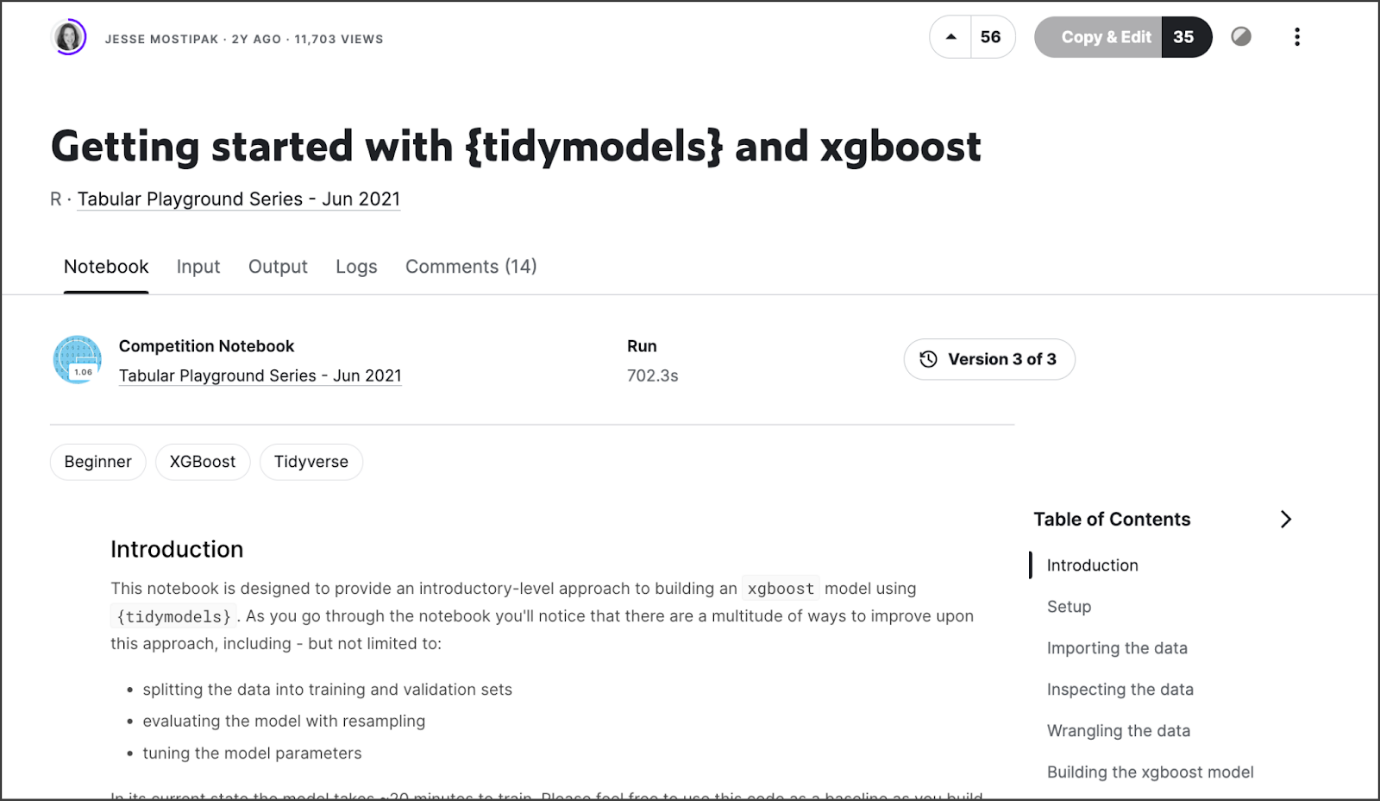


Tableau is a great platform if you want to feature polished visualizations and demonstrate your ability to communicate results in an engaging and elegant way.

## **Notebooks on Kaggle**

If you prefer to compose a more technical exposition of your work, Kaggle is a great place to do it. For example, consider [Jesse Mostipak’s profile](https://www.kaggle.com/jessemostipak). Jesse is a data professional who uses Kaggle to contribute coding exemplars and datasets, participate in modeling competitions, and converse with others in the data community.

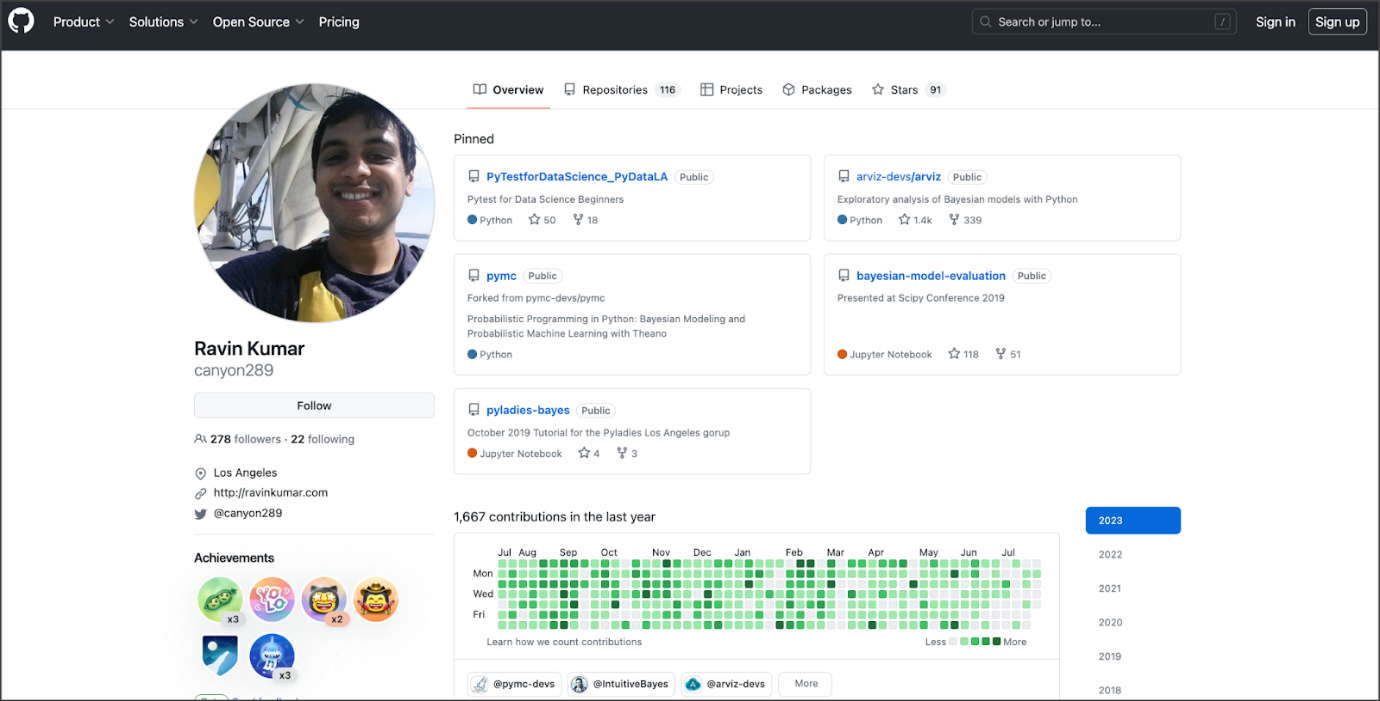
The following is an example of a notebook she submitted as a tutorial on using XGBoost in tidymodels, which is a collection of R packages for modeling and machine learning. You’ll learn how to use XGBoost with Python later in this certificate program.



In addition to supporting your public presence as a data practitioner, publishing your work this way is a great opportunity to generate discussion of different approaches and methodologies with peers in the data community and to learn new skills.

## **GitHub repositories**

Another way to feature your work is by using GitHub. GitHub is probably best known for being a version control system used by coders and developers worldwide to keep track of projects involving large codebases with many simultaneous contributors. GitHub also offers tools to create and customize user profiles and project repositories. Take [Ravin Kumar’s GitHub page](https://github.com/canyon289) for example.

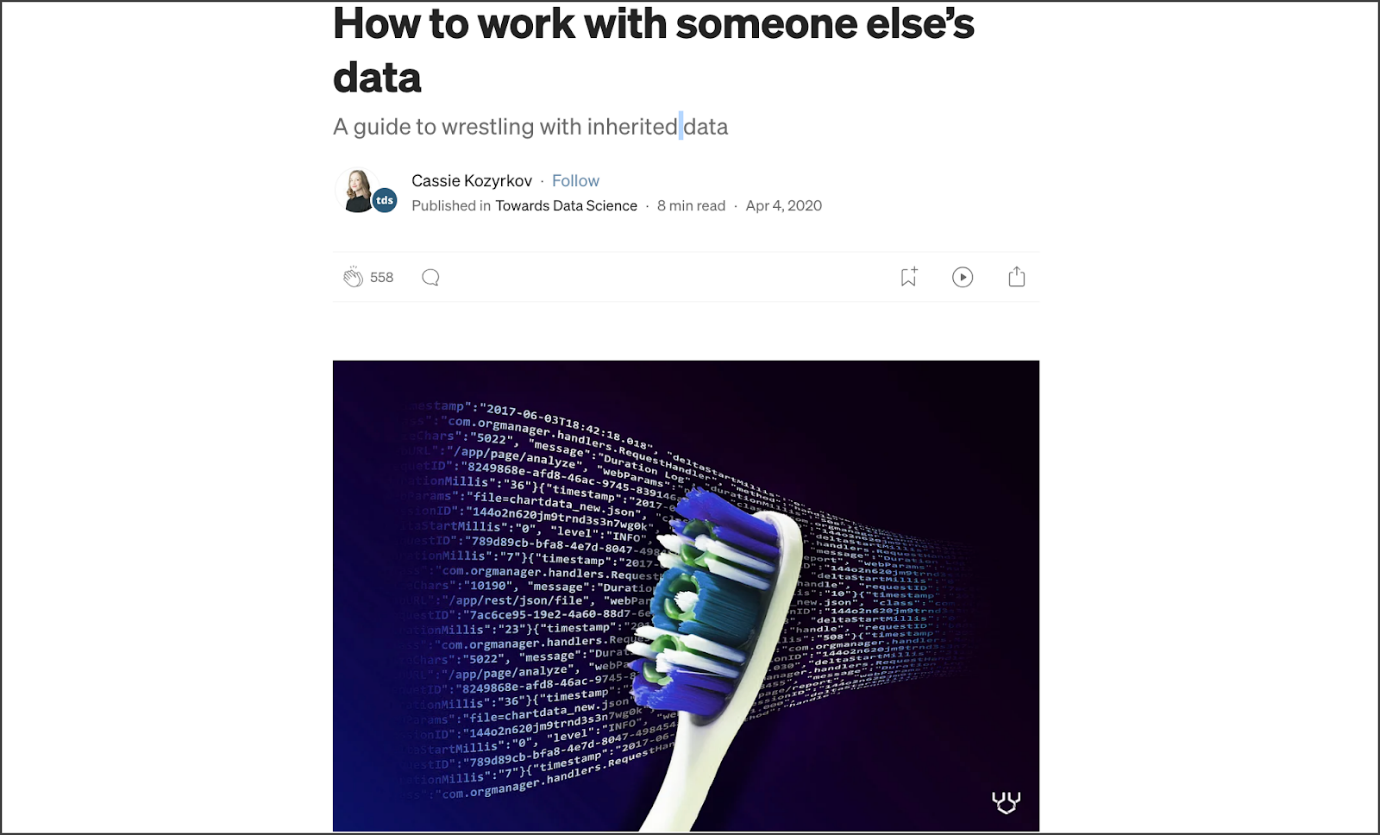


Some advantages of GitHub include the ability to structure your projects into repositories, which are like folders that are accessible to other people via Git. Repositories can include Jupyter notebooks, computer programs, images and figures, and markdown exposition. You’ll work extensively with Jupyter notebooks and markdown text throughout this certificate program. GitHub also lets you watch, star, and follow users or individual projects. The platform is particularly useful for sharing code directly with other people and even working collaboratively with them on projects.

## **Blogs on Medium**

Jupyter notebooks and coding exhibitions aren’t the only means by which you can demonstrate expertise. Blog posts are another great way to propel your professional development. One of the advantages of blogs is that they showcase a wide range of highly sought-after skills beyond just technical proficiency. For example, not only will a well-written blog post convey your understanding of data concepts, tools, and techniques, but it will also emphasize your critical thinking, communication, and writing skills. These types of skills are what distinguish the best data professionals in the field.

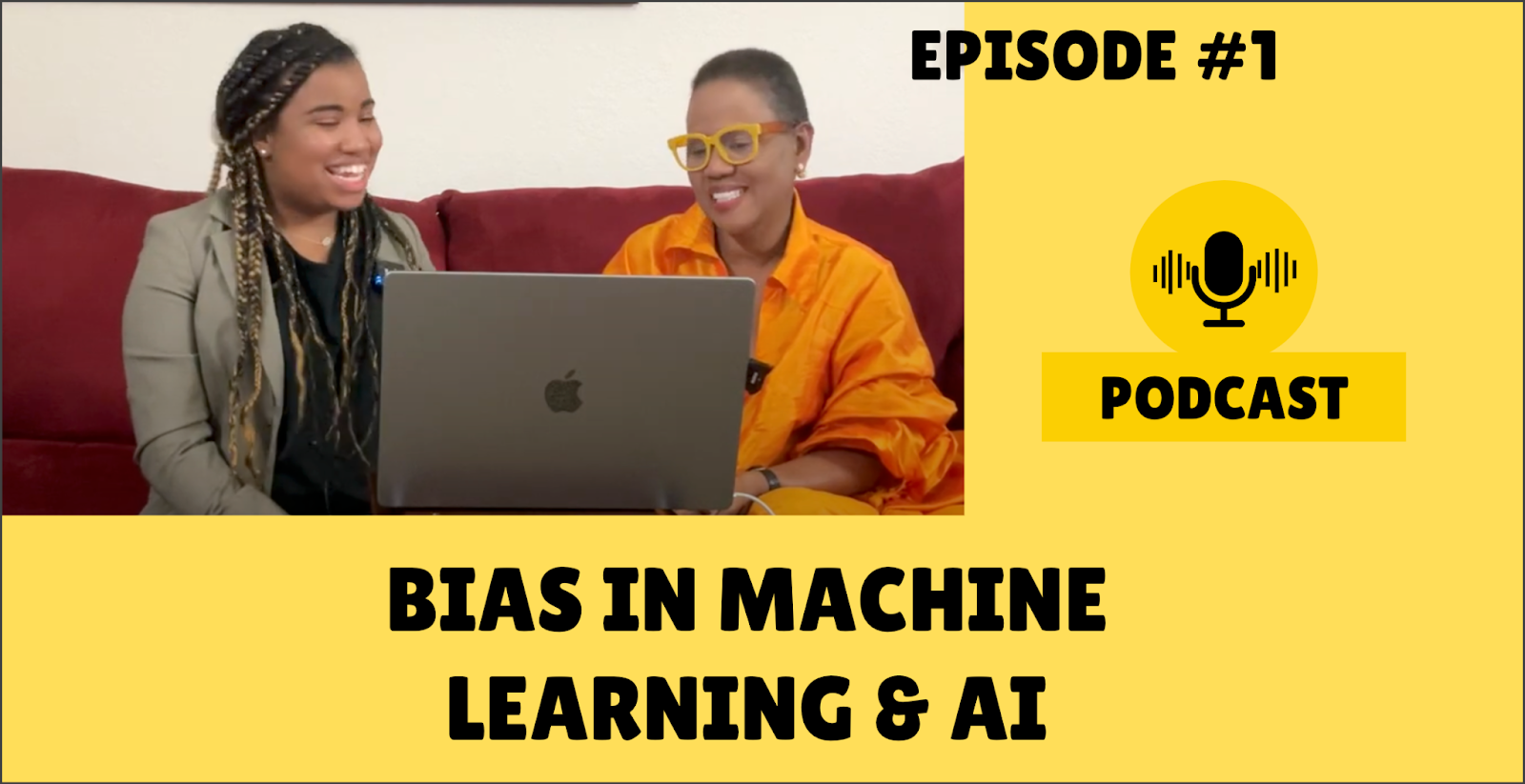
Medium is a great platform to post data science-related blogs. You can also submit to publishers of curated data science content, such as Towards Data Science, Analytics Vidhya, and KDNuggets. As an example, consider the following [article written by Cassie Kozykov](https://towardsdatascience.com/how-to-work-with-someone-elses-data-f33485d79ed4) (the instructor of this course) that was published on Towards Data Science.



This post isn’t about coding or complex computer science or mathematics at all! It’s a practical guide for avoiding pitfalls when working with third-party data. Such topics are highly relevant, and writing about them is another great way to access exciting career opportunities.

## **Other approaches**

This list isn’t exhaustive. It’s intended to give you an idea of the possibilities for presenting your work and becoming an active participant in the field. Some people share Jupyter notebooks, others write helpful articles, and others create podcasts or YouTube videos. For example, consider [this YouTube video on bias in machine learning and AI, produced by Angelica Spratley](https://www.youtube.com/watch?v=J2gTrr0SKuI), a data scientist and educator.



By the way, bias is something you’ll learn about in the course on machine learning. Find what excites you, and your enthusiasm will be evident. The possibilities are vast if you follow your imagination and motivation to contribute to the exciting world of data.

## **Key Takeaways**

These are just a few ways you can share your work. One of the great things about doing so is that it gives you a public-facing profile that employers can review. If someone searches for your name and finds an impressive portfolio, it gives you more opportunity to stand out. Additionally, presenting your work forces you to clearly explain (and thus understand) your methodologies, making you a stronger data practitioner. Whichever way you choose to share your work, be creative, have fun, and don’t be afraid to put yourself out there. This is your opportunity to impress!

# Test your knowledge: Trajectory of the field

Question 1

Fill in the blank: Artificial intelligence is the development of \_\_\_\_\_ able to perform tasks that normally require human intelligence.

computer systems

Question 2

In what way is building diverse teams an effective method for countering human bias in data work? Select all that apply.

- It promotes wider representation.

- It yields more accurate project results.

Question 3

Data professionals can use AI to help them perform which of the following tasks? Select all that apply.

- Automate time-consuming tasks such as data cleaning.

- Analyze large datasets.

-Create predictive models to forecast future events.

Question 4

What tool can generate human-like text in response to a wide range of prompts and questions from data professionals?

Large language model (LLM)

“Over time, I became more and more confident of what I could accomplish and the things that I could reach for.”

LinkedIn

-Follow industry trends

-Learn from thought-leaders

-Stay engaged with the global data analytics community

-Job boards and recruiters

Opportunities to build relationships

-Follow organizations and business leaders on social platforms.

-Search for data field webinars

-Attend data science and data analytics events

Mentor

-Someone who shares knowledge, skills, and experience to help another grow both professionally and personally.

# Make the most out of mentorships

As you have been learning, professional relationships can help you find job opportunities. Exploring job boards and online resources is only one part of your job search process; it is just as important to connect with other professionals in your field, build your network, and interact with the data science community. A great way to achieve these goals is by building a relationship with a mentor. In this reading, you will learn more about mentors, the benefits of mentorship, and how to connect with potential mentors.

## Considering mentorship

Mentors are professionals who share knowledge, skills, and experiences to help you grow and develop. They can offer guidance at different points in your career. Mentors can be advisors, sounding boards, honest critics, resources, and more. You can even have multiple mentors to gain more diverse perspectives!

There are a few things to consider along the way:

* **Decide what you are searching for in a mentor.** Think about your strengths and weaknesses, what challenges you have encountered, and how you would like to grow as a data professional. Share these ideas with potential mentors who might have had similar experiences and have guidance to share.
* **Consider common interests.** Often you can find great mentorships with people who share interests and backgrounds with you. This could include someone who had a similar career path or even someone from your hometown.
* **Respect their time**. Often, mentors are busy! Make sure the person you are asking to mentor you has time to support your growth. It’s also important for you to put in the effort necessary to maintain the relationship and stay connected with them.

Note that mentors don't have to be directly related to data science. Mentors can be friends of friends, more experienced coworkers, former colleagues, or even teammates. For example, if you find a family friend who has a lot of experience in their own non-data field, but shares a similar background as you and understands what you're trying to achieve, that person may become an invaluable mentor to you. Or, you might happen to meet someone at a casual work outing or a professional conference or meetup with whom you develop an instant connection over shared interests or hobbies.

No one mentor may be able (or willing) to advise in all areas, so think about the skills, insights, or values you appreciate in that individual. Then, build a network of individuals that you may approach with different questions about different topics (job searches, public speaking, technical topics, different industries, etc.).

## Build the relationship

Once you have considered what you’re looking for in a mentor and found someone with time and experience to share, you’ll need to build that relationship. Sometimes, the connection happens naturally, but usually, you need to formally ask them to mentor you.

One great way to reach out is with a friendly email or a message on a professional networking website. Describe your career goals, explain how you think those goals align with their own experiences, and talk about something you admire about them professionally. Then you can suggest a coffee chat, virtual meetup, or email exchange as a first step.

Be sure to check in with yourself. It’s important that you feel like it is a natural fit and that you’re getting the mentorship you need. Mentor-mentee relationships are equal partnerships, so the more honest you are with them, the more they can help you. Most importantly, remember to thank them for their time and effort!

As you get in touch with potential mentors, you might feel nervous about being a bother or taking up too much of their time. But, mentorship is meaningful for mentors too. They often genuinely want to help you succeed and are invested in your growth. Your success brings them joy! Many mentors enjoy recounting their experiences and sharing their successes with you as well. Mentors often learn a lot from their mentees. Both sides of the mentoring relationship are meaningful!

## Resources

There are a lot of great resources you can use to help you connect with potential mentors. Here are just a few:

* **Meetups** are meetings that are usually local to your geography. Enter a search for “data science meetups near me” to check out what results you get. There is usually a posted schedule for upcoming meetings. Find out more information about [meetups happening around the world](https://www.meetup.com/find/?keywords=data%20science&source=EVENTS).
* **Platforms** including LinkedIn® and Twitter are a great way to reach out to other professionals. Use a search on either platform to find data science or data analysis hashtags to follow. Post your own questions or articles to generate responses and build connections that way.
* **Webinars** may showcase a panel of speakers and are usually recorded for convenient access and playback. You can see who is on a webinar panel and follow them too. Plus, a lot of webinars are free. One interesting pick is the [Tableau on Tableau webinar series](https://www.tableau.com/learn/series/how-we-do-data). Find out how Tableau has used Tableau in its internal departments. There are also a number of other data science related webinars available at [Brighttalk.com](https://www.brighttalk.com/topic/data-science/)
* **Conferences** present innovative ideas and topics. The cost varies, and some are expensive. But, many offer discounts to students, and some conferences like [Women in Analytics](https://womeninanalytics.com/about/) aim to increase the number of under-represented groups in the field.
* **Associations** or **societies** gather members to promote a field such as data analytics. Many memberships are free. The [Association of Data Scientists](https://www.adasci.org/) is just one example. The [Cape Fear Community College Library](https://libguides.cfcc.edu/c.php?g=604688&p=4191138) also has a list of professional associations for analytics, business intelligence, and business analysis.
* **User communities** and **summits** offer events for users of professional tools; this is a chance to learn from the best. Have you explored the [Tableau](https://community.tableau.com/s/) or [Python](https://www.python.org/community/) communities?
* **Non-profit organizations** that promote the ethical use of data science often offer events for the professional advancement of their members. The [Data Science Association](https://www.datascienceassn.org/) is one example.

## Key takeaways

Finding and connecting with a mentor is a great way to build your network, access career opportunities, and learn from someone who has already experienced some of the challenges you’re facing in your career. Whether your mentor is a senior coworker, someone you connect with on LinkedIn®, or someone from home on a similar career path, mentorship can bring you great benefits as a data analytics professional.

Career-building activities and resources

-Complete hands-on activities based on data-driven scenarios.

-Find and apply for jobs

-Prepare for the interview process

-Put together an online portfolio

“I look for candidates with the experience that they leverage advanced analytics solutions or machine learning solutions to drive business impact.”

“The successful candidates are those that are able to relate their past projects or their school work to any type of problems.”

“Give your best, but also know when to stop.”

# Showcase your skills: How to prepare for the interview

As you’ve been learning, there are many steps that go into preparing for your job search in the data career space. Before you even start applying for jobs, it’s important that you take two key steps. You need to build a professional online presence and develop your portfolio to showcase your skills to potential employers. This is also a great time to connect with mentors, who can provide professional insight and help you prepare for the application process. The interview will often be the final step in this process; in addition to all of the other job prep you’ll do, preparing for the interview will help you approach this last step confidently. To help get you ready for the job market, this reading will explore four different interview types and offer interview questions that this certificate program will help you answer.

## Establishing your professional narrative

Establishing a professional narrative, or brand, can help you build connections between your daily work and the positive influence it has on something bigger. This something bigger could be an exciting project that you helped finish or a positive change in your organization or community! This applies to the work you have done previously and the work you hope to do in the roles you apply for. Framing your experience and goals around your professional narrative allows you to develop a stronger sense of the value you bring to an organization and your own career goals. Keeping your professional narrative top of mind is important as you prepare to enter or navigate the data professional career space. Your personal brand is the value you offer potential employers. It can often provide answers to questions posed during the interview process.

## Interview question types

There are many different types of interview questions and each organization has its own priorities for what they want to know about each applicant for each role. Interview questions typically belong to one of four categories:

* **Behavioral questions**: These questions ask you to describe how you have handled specific situations in the past, and your personal characteristics. They are designed to assess your skills, experience, and problem-solving abilities, as well as your fit for the company culture.
* **Technical questions**: These questions ask you to demonstrate the knowledge and skills presented in your resume or portfolio. In job interviews for technical positions, an employer may ask for a demonstration of specific tasks, prior projects, or even a take-home assessment. Often these technical demonstrations are presented as a separate assignment that you will complete outside of the interview itself. These questions are designed to assess your technical skills and expertise.
* **Situational questions**: These questions ask you how you would handle hypothetical situations. Similar to behavioral questions, employers typically use situational questions to develop a preliminary understanding of how your skills fit the role. Situational questions are designed to assess your judgment, critical thinking skills, and ability to apply knowledge to new situations.
* **Subject questions**: These questions ask you about your knowledge of a specific subject or area, usually pertaining to the field or industry that you’re applying for. These questions are designed to determine how well you understand the relationship between the role you’re applying for and the broader context of the company. Employers may also use these questions to assess your understanding of how the company works in contrast to direct competitors in the marketplace.

These are just a few examples of the types of questions you might be asked in interviews. These categories aren’t universal, and different organizations have different interview styles– they may even ask questions that combine categories. For example, an interviewer might give you a hypothetical situation and ask you for an example of a previous situation you’ve encountered that relates. This is a combination of a behavioral and situational question. Generally, the goal of interview questions is to assess your skills, experience, and general fit for the role; so keep that in mind as you prepare.

## Applying course skills

As you progress through this certificate program, you will learn industry skills that interviewers will be interested in asking about. Throughout your learning journey, it will be useful to identify and keep in mind key skills you will need to be able to discuss. The following is a list of questions that you might be asked in an interview for data professional positions. You will find questions in this list that are representative of the four interview question type categories explained in the previous section. Finishing your certificate will mean you’re prepared to answer all of these questions!

### 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?

## Begin with the end in mind

At this point in the certificate program, you are still early in your learning journey. Because of that, you have the opportunity to consider everything you’re going to learn in the context of your final goal: taking the next step in your data professional career. Part of taking that next step involves interviewing with potential employers. As you learn more and more skills and become familiar with new tools, keeping these interview questions in mind can help you frame how what you’re learning now applies to future job roles. These questions can also help you frame your focus in each course–by considering how you might use the new skills and knowledge you’re learning, you can better understand why the work you’re doing now is so important!

Starting now, you can keep your final goals in mind and continue to build them into your professional narrative. That way, by the end of this program, you will already have a strong framework for communicating with potential employers.

## Key Takeaways

The interview is an opportunity to share how you can add value to an organization. Recognizing your growing skillset and how you might communicate those skills to potential employers is a great way to showcase not just your technical know-how, but your ability to communicate effectively too. This reading is a great resource to keep in mind as you build your skills and your professional narrative in preparation for your job search.

# Glossary terms from module 3

# Terms and definitions from Course 1, Module 3

**Active listening:** Refers to allowing team members, leadership, and other collaborative stakeholders to share their own points of view before offering responses

**Analytics Team Manager**: A data professional who supervises analytical strategy for an organization, often managing multiple groups

**Business Intelligence Analyst**: (Refer to **Business Intelligence Engineer**)

**Business Intelligence Engineer:** A data professional who uses their knowledge of business trends and databases to organize information and make it accessible; also referred to as a Business Intelligence Analyst

**Chief Data Officer**: An executive-level data professional who is responsible for the consistency, accuracy, relevancy, interpretability, and reliability of the data a team provides

**Data cleaning**: The process of formatting data and removing unwanted material

**Data Engineer:** A data professional who makes data accessible, ensures data ecosystems offer reliable results, and manages infrastructure for data across enterprises

**Data Scientist**: A data professional who works closely with analytics to provide meaningful insights that help improve current business operations

**Interpersonal skills:** Traits that focus on communicating and building relationships

**Mentor:** Someone who shares knowledge, skills, and experience to help another grow both professionally and personally

**RACI chart**:A visual that helps to define roles and responsibilities for individuals or teams to ensure work gets done efficiently; lists who is responsible, accountable, consulted, and informed for project tasks

# Terms and definitions from previous modules

## A

**Aggregate information**: Data from a significant number of users that has eliminated personal information

**Artificial intelligence (AI):** Refers to computer systems able to perform tasks that normally require human intelligence

## D

**Data anonymization**: The process of protecting people's private or sensitive data by eliminating PII

**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

## E

**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

## H

**Hackathon:** An event where programmers and data professionals come together and work on a project

## J

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

## M

**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

## N

**Nonprofit:** A  group organized for purposes other than generating profit; often aims to further a social cause or provide a benefit to the public

## O

**Open data**: Data that is available to the public and free to use, with guidance on how to navigate the datasets and acknowledge the source

## P

**Personally identifiable information (PII):** Information that permits the identity of an individual to be inferred by either direct or indirect means

**Python**: A general-purpose programming language

## S

**Sample:** A segment of a population that is representative of the entire population

## T

**Tableau**: A business intelligence and analytics platform that helps people visualize, understand, and make decisions with data

# Module 3 challenge

Question 1

Fill in the blank: A data professional practices active listening when they allow others to \_\_\_\_\_ before offering a response.

share their points of view

Question 2

What is the main objective of data cleaning?

Remove anything that could cause an error during analysis

Question 3

Fill in the blank: Data engineers are responsible for making data \_\_\_\_\_.

accessible

Question 4

Which data professionals are responsible for supervising an organization’s analytical strategy?

Insights or analytics team managers

Question 5

Fill in the blank: Business intelligence engineers are responsible for organizing data and making it \_\_\_\_\_.

Accessible

Question 6

Which of the following statements accurately describe a RACI matrix? Select all that apply.

Someone who is labeled “informed” is kept aware of progress and the concerns of people working on a project.

Someone who is labeled “consulted” is typically a subject matter expert who offers input on a task.

The acronym RACI stands for responsible, accountable, consulted, and informed.

Question 7

A data professional sets criteria to ensure consistent data practices and procedures across the organization. They want to promote best practices, effective communication, and transferability of information among teams. Which principle for data team building does this scenario describe?

(NOT FINAL)

Accountability

Or Standardization

**Module 4:**

**What you’ll learn**

-Workflow within data-driven careers

-The PACE model

-Effective communication best practices

“I don’t think it’s possible to have a successful data science career relying only on technical know-how.”

“Be asked to be invited to meetings where you don’t need to contribute, but you can use it as an opportunity to observe people who are good at communicating.”

**Data workflow structure**

-Ask

-Prepare

-Process

-Analyze

-Share

## Question

What does the acronym PACE stand for?

Plan, analyze, construct, and execute

**Plan**

-What are the goals of the project?

-What strategies will be needed?

-What will be the business or operational impacts of this plan?

**Analyze**

-Acquire data from primary and secondary sources

-Clean, reorganize, and transform data for analysis

-Engage in EDA

-Work with stakeholders

**Construct**

-Build and revise machine learning models

-Uncover relationships in the data

-Apply statistical inferences about data relationships

**Execute**

-Present findings to internal and external stakeholders

-Answer questions

-Consider differing viewpoints

-Present recommendations based on the data

# The PACE Stages

So far, you have been introduced to the PACE framework and learned how it provides a clear foundation and structure for data analysis projects. You also learned that PACE is an acronym; each one of the letters represents an actionable stage in a project: plan, analyze, construct, and execute. In this reading, you will learn more about the PACE workflow and discover how each stage of the process can help guide data analysis.

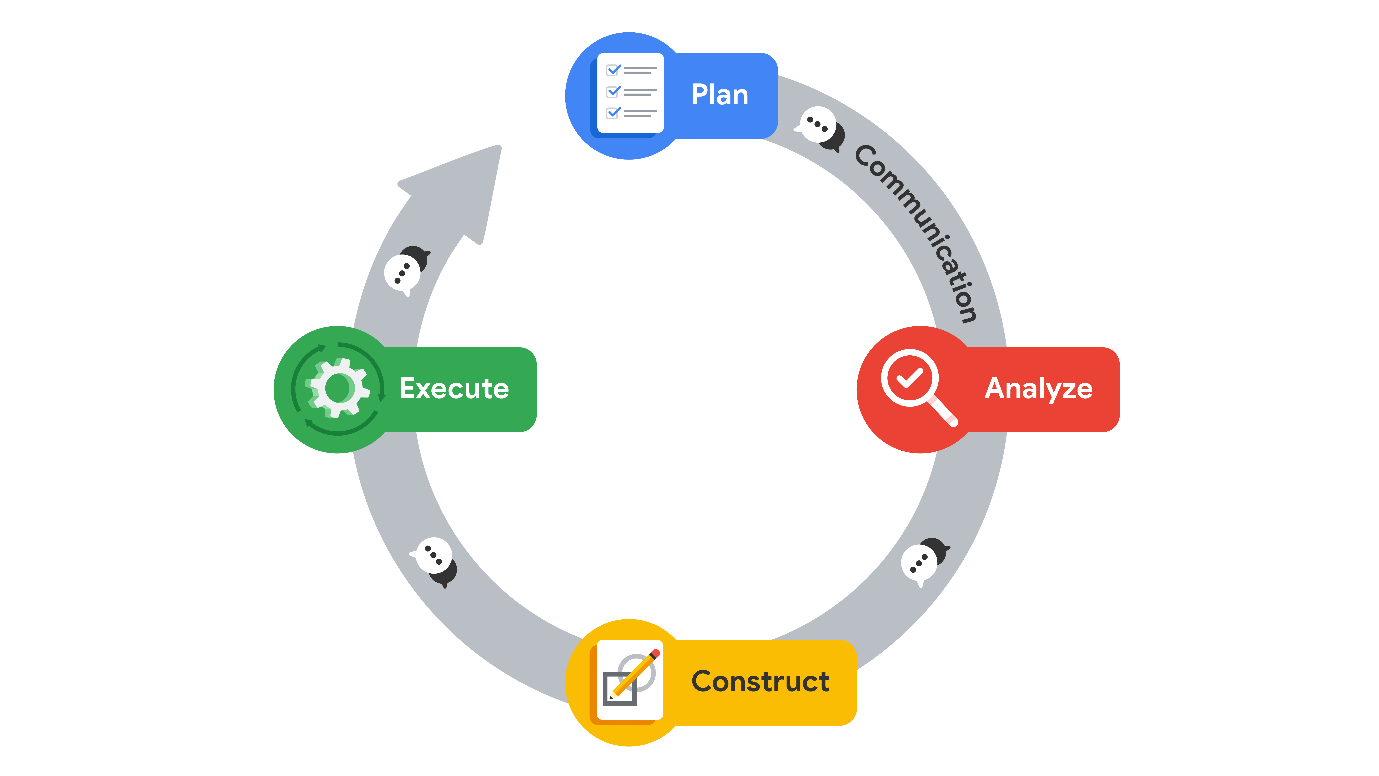
## Why do we use a workflow structure?

As a general rule, data professionals rely on workflow structures to guide them through the duration of data projects. Within a large-scale project, there can be a number of tasks that require a certain order of operations. Identifying complexities and finding consistent ways to work together can make projects more efficient and enable more productive communication. Identifying these and other types of potential blockers early can help you plan and prepare resources in advance before they can negatively affect a project.

Our team of data professionals who assisted in creating this program developed PACE as a flexible model; you are encouraged to revisit each stage without interrupting the entire workflow. Through PACE, you will identify areas of action and contexts for when they will need to be considered. All in all, PACE offers professionals a customizable scaffold that can support their efforts while working through every stage of a data project.

## A closer look at the PACE model

Let’s take a closer look at each stage of the PACE model.



### **Plan**

At the beginning of a project, it is important to establish a solid foundation for success. Here you will define the scope of your project. This is when you will begin by identifying the informational needs of the organization. During the planning stage, you will have the widest viewpoint of a project. By assessing all of the factors and processes involved, you are mapping a path to completion, using your creativity to conceptualize a course of action. Here you will also take special note of tasks that may require an innovative approach within your workflow.

**Summary**: The planning stage is where you conceptualize the scope of the project and develop the steps that will guide you through the process of completing a project.

Here are a few examples of the types of planning stage tasks:

* Research business data
* Define the project scope
* Develop a workflow
* Assess project and/or stakeholder needs

### **Analyze**

In the analyzing stage, you will interact with the data for the first time. Here you will acquire all of the data you will need for the project. Some datasets could come from primary sources within your organization. Others may need to be collected from secondary sources outside your company. You may even find that you need governmental or open source data. The analyzing stage is also where you will engage in exploratory data analysis or EDA. This involves cleaning, reorganizing and analyzing all of the necessary data for the project.

**Summary**: The analyzing stage is where you will collect, prepare, and analyze all of the data for your project.

Here are a few examples of the types of analyzing stage tasks:

* Format database
* Scrub data
* Convert data into usable formats

### **Construct**

Just as the name suggests, the construction stage is all about building. In this stage of PACE, you will be building, interpreting, and revising models. Some projects will require machine learning algorithms to uncover correlations within your data. You will use these correlations to uncover information from the data that would otherwise go unused. These relationships can help your organization make informed decisions about the future.

**Summary**: In the construction stage you will build models that will allow you access to hidden relationships locked within data.

Here are a few examples of the types of construction stage tasks:

* Select modeling approach
* Build models
* Build machine learning algorithms

### **Execute**

In the execution stage, you will put your analysis and construction into action. Here you will deliver your findings to the internal (inside of your organization) and external (outside of your organization) stakeholders. Quite often, this will involve stakeholders from the business-side of the companies you are working with. Presenting your findings is only a part of the execution stage. Stakeholders will provide feedback, ask questions, and make recommendations that you will collect and incorporate.

**Summary**: In the execution stage you will present the finding of your analysis, receive feedback, and make revisions as necessary.

Here are a few examples of execution stage tasks:

* Share results
* Present findings to other stakeholders
* Address feedback

## Communication and PACE

Regardless of where you might be within the PACE workflow, communication is essential to moving the framework through to the realization of the project. One way to think of this is by visualizing the four stages of PACE as a completed circuit and with communication being represented by the flow of electricity.

At each stage, there will always be a need for communication to improve the workflow. This could be asking questions about your data, gathering additional sources, updating stakeholders on progress, or presenting findings and receiving feedback.

## Adaptability of PACE

At the start of a project, the PACE model offers a good structure to guide you. At the beginning, you have the planning stage, where you gather the information and tools you will need and set up a framework to guide you. When you are analyzing data and constructing models, the analyzing and construction stages assist you. After these steps, the execution stage is where you share results and gather feedback.

Although the PACE model is first presented as stages in a certain order, you will discover that the open flow of communication allows you to easily move to the stages you need. New information and feedback can be incorporated at any part of the process. You might need to return to the analyzing stage to clarify some aspect of the data and then go to the executing stage to present this aspect to your stakeholders, without the need to construct new models. The PACE framework can be adapted to fit any project. Its adaptability will prepare you for a dynamic profession that requires a high degree of professional flexibility and communication.

## Key takeaways

Data professionals need structured workflows to help them manage the large number of tasks within data projects. The PACE professional workflow was designed specifically for this program to assist you in developing your professional structures and practices. PACE functions like a completed circuit, with communication flowing between each stage. The design of PACE promotes flexibility, allowing for free movement between stages as needed.

# Test your knowledge: The data project workflow

Question 1

“In which PACE stage would a data professional ask, “What are the goals of the project?”

Plan

Question 2

Fill in the blank: In the \_\_\_\_\_ stage of the PACE model, a methodical examination of the data is conducted.

analyze

Question 3

In the execute stage of the PACE model, what is shared with stakeholders?

The story told by the data

Questions to keep in mind

-What does my audience already know?

-What do they need to know?

Sender

The person responsible for crafting the message or communication

Questions to keep in mind

-What’s your relationship to the receivers?

-What’s your role in this exchange?

-What personal biases might affect the message?

## Question

Which element of communicative exchange refers to crafting a message or communication?

Sender

# Best communication practices for data professionals

# Effective communication drives PACE

Throughout the stages of PACE, stakeholders can offer feedback, pose questions, or ask for clarification. Being able to communicate is key; at every point in a project’s life cycle, you will need to be able to share obstacles and results, and provide necessary information to guide decision-making. Communication drives each stage of PACE, from collecting data to constructing models to sharing results. As a data professional, you will need a combination of strong communication skills and the technical skills you’re learning to succeed in the data career space. In this reading, you will learn some tips for effective communication you can apply throughout the stages of PACE. You will also discover some best practices for sharing your findings through presentations that you can use in the future.

# Seven tips for effective communication

Over the course of a typical work day, you will interact with stakeholders in a variety of ways. Some of your interactions will be through emails and messaging, others through meetings and one-on-one conversations, and even formal presentations. Each interaction requires an individualized approach to ensure that your audience understands what you are trying to communicate. The following seven tips can help guide your communication, no matter what form it takes or what stage of PACE your project is in:

### 1. Speak the language of your audience

Identify the needs of your audience. It’s important to know the objectives of the person you are communicating with. Focusing on their needs allows you to gain insight and assess how technical the conversation should be. Consider the following questions:

* Why has this person contacted you?
* What does your stakeholder want from this interaction?
* What’s important to them, their team, or their organization?

In your role as a data professional, you will interact with a variety of stakeholders. Determine what they know, what they need to know, and what might go beyond their level of involvement in a project.

* Break down technical concepts into simpler terms.
* Use shorter sentences so main ideas are easier to understand and remember.
* Use direct language and minimize embellishments or unnecessary detail.
* Pay attention to diverse backgrounds and respect the lived experiences of others.

**Pro tip**: Avoid jargon, acronyms, and technical “buzzwords” that could lead to confusion.

### 2. Invite questions and welcome feedback

Everyone can use feedback–whether it is positive reinforcement or areas for improvement. When you are focused on the details of a task, it’s possible to overlook something. Another person’s feedback offers you a great way to gather insights for your personal growth and professional excellence. By accepting the challenge that feedback and questions present, you will strengthen your own skills and help the overall project.

* Merge your passion for finding solutions with the goals of the project.
* Continue to strive for greater understanding of the results.
* Elicit feedback and questions to improve communication about your projects.
* Consider opportunities to reflect on your communication skills.

**Pro tip**: Analyze feedback. Is it valid? Does the person have a complete understanding of the goals of the project or data analytical process? If not, set up an additional meeting to help clarify.

### 3. Be the connection to the data

## You are your team’s direct connection to the insights your data offers. Your goal is to help other stakeholders understand the process and how it addresses their needs. When everyone understands the process, communication can be highly effective.

* Focus on the objectives to help others better understand your data process.
* Tell the story of the data with a compelling and cohesive narrative.
* Respond to questions in a timely manner.
* Demonstrate your value to the team.
* Find opportunities to address stakeholder questions.

**Pro tip**: Continue to proactively identify ways the data and tools you have access to can address the objectives of your team and drive new insights.

### 4. Let your visualizations help tell the story

Visualizations are one of the best ways to communicate ideas, especially when dealing with big data. Visual references help bring to life the details inside your data. Graphs, charts, and infographics can promote general understanding. Later you will explore Tableau, a visualization tool that you can use to create compelling visuals from data.

* Be sure that your visuals tell the story within the data.
* Design visualizations for inclusivity.
* Use labels and text to clarify, not clutter.
* Use fonts that are easy to read.
* Use high contrast, shading, and other customizations to communicate your messages clearly.
* Offer handouts, slides, and other material in accessible formats.

**Pro tip**: Keep visualizations simple. When deciding what to include in a presentation, less is more.

### 5. Build positive professional relationships

When you consider the responsibilities and objectives of others, your communication will reflect that consideration. This builds credibility and influence in your workplace and allows you to continue growing throughout your career.

* Focus on what matters to your audience.
* Invite feedback and discussion.
* Be a trusted subject matter expert who communicates clearly and inclusively.
* Cultivate positive interactions to strengthen working relationships and improve morale.

**Pro tip**: When a stakeholder contacts you, be accessible and engaged in your communication.

### 6. Identify assumptions about the data

The backgrounds, experiences, beliefs, and worldviews of people can influence the information contained within data. In your role as a data analytics professional, you will want to consider ways that these various factors can introduce bias.

If they are not recognized, assumptions can have a powerful effect on outcomes. Without information, there is often a tendency to fill in the gaps in understanding with assumptions. The most effective methods to reduce the impact of assumptions are practicing active listening and effectively asking questions. For example, these questions can help identify any assumptions:

* Is there something I am taking for granted?
* Am I assuming something here that I shouldn’t?
* Can I determine if the assumption is correct?

**Pro tip:** Data professionals need to identify their own assumptions as well as any assumptions their audience might have. Make sure you consider any bias you might have, too!

### 7. Identify limitations in the data

As a data professional, you will also encounter limitations within data that can impede your analysis. These limitations will need to be addressed before you can proceed. To assist you in identifying data limitations, ask the following questions:

* Is the data complete? Are there missing values or sections?
* Are the datasets formatted correctly?
* Is this a sufficient sample size to conduct analysis of an entire population or group?
* What are the biases present in the data set?
* Does this data contain personally identifiable information? What steps will I take to protect this information?

**Pro tip:** In addition to identifying and communicating about data limitations before analysis, you will also need to make sure stakeholders are aware of any limitations still affecting the results during your presentation.

## Share Findings

One of the most important communications you will have with stakeholders will be sharing your findings, often through presentations. This will mean translating the results, concepts, and terms of your analysis for wider audiences. When sharing the results of your analysis with stakeholders, there are some best practices that you should apply:

* Craft results to the needs of your stakeholders. Communicate why this data will help them achieve their goals.
* Determine the visualizations and/or dashboards that are the most effective. What data will you need to show and how do you want stakeholders to interact with it?
* Think about the design carefully. A simple yet visually appealing approach to visualizations is always the best.
* Use a hierarchy of data in your visualizations/dashboards. Information that is most important should be easily accessible, but you should provide a path for more details.

### What should I keep in mind when I share results?

* What information is the most important to my audience?
* What is the most efficient way to share with the tools available and the time I’m allotted?
* What can I do to make the key points effectively?

### Presentations

Presenting information clearly and effectively is key to a data scientist's workflow. Communication skills that pertain to presentations include but are not limited to: the structure of your presentation, slide design, the tone of your voice and the body language you convey, and more. It’s also important to consider accessibility as you are creating any assets. Check with your organization about accessibility guidelines. You may also refer to online resources like [U.S. accessibility guidelines](https://accessibility.digital.gov/) or [The World Wide Web Consortium’s (W3C) web accessibility initiative](https://www.w3.org/WAI/).

#### Tips for presentations

* Structure your presentation. Be sure there is a logical structure: a beginning, middle, and end.
* Presentation slides are not scripts. Don’t read or put complete paragraphs on presentation slides.
* Make sure your data can be understood visually and consider potential accessibility challenges for your audience.
* Focus most on the points your data illustrates.
* Share one—and only one—major point from each chart.
* Label chart components clearly.
* Visually highlight “Aha!” zones.
* Write a slide title that reinforces the data’s point.

A solid presentation can help others understand the findings of your data analysis and ensure that you are effectively communicating.

## Key takeaways

Effective communication is important for data professionals. Regardless of where you are in the framework of a project, communication can inform and empower your stakeholders. Identify the needs of your audience and invite feedback. Remember that your role is to connect the data, technology, and the stakeholders. When sharing the results of your data analysis, make your presentations clear, focused, accessible, and take into account any assumptions or limitations presented by the data. Demonstrate your value by being available and ready to share insights in a simple way that promotes general understanding.

# Activity: Communicate with stakeholders in different roles

Review Learning Objectives

## Activity Overview



You have learned about the vital role of communication in meeting the needs of different stakeholders in the data professional career space. In this activity, you will write emails to two stakeholders in different roles at a government agency. This activity is based on a fictional workplace scenario.

To review tips for communicating effectively as a data professional, refer to the reading about [essential data professional communication practices](https://www.coursera.org/learn/foundations-of-data-science/supplement/AiGjd/best-communication-practices-for-data-professionals).

Be sure to complete this activity before moving on. The next course item will provide you with a completed exemplar to compare to your own work. You will not be able to access the exemplar until you have completed this activity.

## Scenario



Review the following scenario. Then complete the step-by-step instructions.

You are a data professional working for the U.S. National Park Service (NPS). The NPS manages 63 National Parks, including the Grand Canyon, Great Smoky Mountains, Yellowstone, and Yosemite. The NPS’ mission is to preserve the natural and cultural resources of the parks for the enjoyment, education, and inspiration of present and future generations. Your team analyzes data to improve visitor experiences and protect park resources.

One of the NPS’ main concerns is the recent increase in visitation. In 2022, the NPS received about 312 million recreation visits, an increase of 15 million visits from 2021. 26 percent of total visits occurred in the top 8 most visited parks. While the NPS strives to serve as many people as possible, unexpected increases in visitation can stress the natural environments and wildlife within the parks. Further, overcrowding decreases the overall quality of visitor experiences.

NPS leadership has asked the data team to build a machine learning model to predict future visitation at the most visited parks. To develop the model, the data team will analyze historical data on recreation visits, and also consider other factors affecting visitation rates: weather, temperature, user fees, traffic conditions, and more. The proposed timeline for the project is 12 weeks. The goal is to build a model with at least 90% accuracy.

A powerful model can help park managers better understand trends in future visitation, and take proactive measures to protect the parks’ natural and cultural resources. Managers can also use predictive data to make informed decisions about implementing changes in park operations or infrastructure. For example, increasing staffing, renovating facilities, or upgrading the visitors’ reservation system.

Your team leader has asked you to email two new employees and give them an overview of the project:

* A data professional who will be part of your team. This person will need relevant information about the data team’s workflow and project goals.
* A writer for NPS public relations. This person will need relevant information to create non-technical articles that promote the NPS’ efforts to improve visitor experience.

## Step-By-Step Instructions



Follow the instructions to complete each step of the activity. Then, answer the three questions at the end of the activity before going to the next course item to compare your work to a completed exemplar.

### **Step 1: Access the template**

To use the template for this course item, click the following link and select Use Template.

Link to template:[Emails to different stakeholders](https://docs.google.com/document/d/102j4WYkZRBBZrKHa9R582hMgVOxOlJp1A3N0M_V_z0E/template/preview)

OR

If you don’t have a Google account, you can download the template directly from the following attachment.

[Emails to different stakeholders](https://d3c33hcgiwev3.cloudfront.net/MfLtEUmER7WFoMwdqrloJg_30e5e984ac12436db1ba372083085bf1_Emails-to-different-stakeholders.docx?Expires=1719878400&Signature=F5M6VxynRyEmF4EMzZOsG2PWIhQbJU9IdVfC3LxZxAFWGXqFDz7pnLbg2vF5a27bTJkAQmyiY8NO3Ppt3nj4RVSe5rD8ymBrXAqbh1gbyzI0Sp5dFlktIpK6GMGgR8rNUvA6jnL~Hti1UXthGtyBnAi3LrP3X5Q8LVXPyQXmNPE_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[DOCX File](https://d3c33hcgiwev3.cloudfront.net/MfLtEUmER7WFoMwdqrloJg_30e5e984ac12436db1ba372083085bf1_Emails-to-different-stakeholders.docx?Expires=1719878400&Signature=F5M6VxynRyEmF4EMzZOsG2PWIhQbJU9IdVfC3LxZxAFWGXqFDz7pnLbg2vF5a27bTJkAQmyiY8NO3Ppt3nj4RVSe5rD8ymBrXAqbh1gbyzI0Sp5dFlktIpK6GMGgR8rNUvA6jnL~Hti1UXthGtyBnAi3LrP3X5Q8LVXPyQXmNPE_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

### **Step 2: Complete the email to the new data professional**

Review the scenario and identify the needs of your audience (in this case, the new data professional joining your team):

* Why are you emailing this stakeholder?
* What does this stakeholder need from this interaction?
* What is important to them?

Consider what information would be most relevant to the data professional based on the scenario provided. Then complete **Email #1** with that information, and be sure to invite them to ask follow-up questions.

### **Step 3: Complete the email to the new public relations writer**

### Pro Tip: Save the template

Finally, be sure to save a blank copy of the template you used to complete this activity. You can use it for further practice or in your professional projects. These templates will help you work through your thought processes and demonstrate your experience to potential employers.

## What to Include in Your Response



Be sure to address the following elements in your completed emails to the new data professional and the new writer:

* Indicate why you are communicating with the recipient
* Address what they need from the interaction
* Identify what is important to them
* Invite them to ask follow-up questions

Question 1

Did you complete this activity?

Yes

Question 2

What information is most relevant to include in the email to the new data professional? Select all that apply.

An invitation to ask follow-up questions

The accuracy goal for the visitation prediction model

An overview of the data team’s workflow

Question 3

What information is most relevant to include in the email to the new writer? Select all that apply.

An invitation to ask follow-up questions

The purpose and benefits of the visitation prediction project

The problem the project is addressing

# Activity Exemplar: Communicate with stakeholders in different roles

Here is a completed exemplar along with an explanation of how the exemplar fulfills the expectations for the activity.

## Completed Exemplar



To review the exemplar for this course item, click the following link and select Use Template.

Link to exemplar:[Emails to different stakeholders](https://docs.google.com/document/d/1vkUh8QZbLc0YGbV1K7-wrTfw6770j9UrZwgLWUQXwMs/template/preview)

OR

If you don’t have a Google account, you can download the exemplar directly from the following attachment.

[Emails to different stakeholders](https://d3c33hcgiwev3.cloudfront.net/3dZk-859TRaUpsJFwqTJhQ_69db6d6b9fc1404da2b2624359cc61f1_Emails-to-different-stakeholders.docx?Expires=1719878400&Signature=KXhg~6hqDiywGnc-kwUyn1QgWEJ0Id950H4Gg4GQkhChZgQ3xWtiVi3X1S5sR15QG4L8lTZMJBsPM~X786eG0WQ66zBdP8G5l1jzNMftJBH-4cL6US2oUpL6uQZnFHP~G0Zcgs7jvZ9-vIapyiIwIjAoCuRSYwV~OijhiG9E-Io_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[DOCX File](https://d3c33hcgiwev3.cloudfront.net/3dZk-859TRaUpsJFwqTJhQ_69db6d6b9fc1404da2b2624359cc61f1_Emails-to-different-stakeholders.docx?Expires=1719878400&Signature=KXhg~6hqDiywGnc-kwUyn1QgWEJ0Id950H4Gg4GQkhChZgQ3xWtiVi3X1S5sR15QG4L8lTZMJBsPM~X786eG0WQ66zBdP8G5l1jzNMftJBH-4cL6US2oUpL6uQZnFHP~G0Zcgs7jvZ9-vIapyiIwIjAoCuRSYwV~OijhiG9E-Io_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

## Assessment of Exemplar



Compare the exemplar to your completed activity. Review your work using each of the criteria in the exemplar. What did you do well? Where can you improve? Use your answers to these questions to guide you as you continue to progress through the course.

***Note:*** The exemplar represents one possible way to complete the emails. Your emails will likely differ in certain ways. What’s important is that your emails clearly identify information that will help each recipient fulfill their job responsibilities.



Email #1 to the new data professional provides:

* An overview of the data team’s workflow. This information gives the new data professional insight into how the data team shares their results and invites feedback from internal stakeholders.
* The accuracy goal for the visitation prediction model. Knowing the project objective is important for a data professional working to develop an effective ML model.

Email #2 to the new public relations writer provides:

* Relevant information for creating non-technical articles that promote the NPS’ mission and work. This information includes the purpose and benefits of the visitation prediction project, and the problem the project is addressing.
* Direct, non-technical language that does not include unnecessary detail.

Both emails:

* Invite the recipient to ask follow-up questions. Questions are an opportunity to improve communication and learn more about the project.

# Elements of successful communication

As you have been learning, communication is the driving force behind PACE because data professionals need to be able to communicate effectively with stakeholders while working through different stages of a project. You have already learned some important tips for effective communication. But, there are a few more elements for successful communication you should consider. This reading provides best practices for successful communication that you can follow in your workplace communications.

## Understanding why

Having a clear vision of why you are communicating is the first thing you need to consider. Your “why” depends on the context set by the business or organization you work for as well as the goals orienting the project. When crafting any form of communication, use your why to guide main ideas so that your audience can identify how to act or respond with purpose.

When you prepare to communicate, take a moment to outline important goals and expectations you have, like:

* Goals of the project you are communicating about
* What you hope to gain from this communication
* What you’re asking your audience to do
* What you need your audience to understand

Understanding the why behind your communication will help you organize your thoughts and develop clearer, more direct communication.

## Set the stage

When you are developing effective communication, you have to consider more than just the why–you also need to think about where the communication is taking place. Setting will have a direct impact on how your message is delivered and how you shape it. As you prepare to communicate, consider the most appropriate way to communicate for the setting you plan to be in.

On the job, it’s possible that you will communicate in a variety of settings. What information you share, how you share it, and how you follow up will depend on the context of that communication. For example, you might:

* Ask a coworker for advice about a recent obstacle over lunch
* Send an email updating all the stakeholders about an important project
* Share progress with your team in a weekly meeting
* Present the results of your analysis to a boardroom of executives

Each of these settings will require you to consider how you’re communicating, what each of those audiences need, and what you need from them in return. As you develop your communication skills, don’t forget that the setting can be just as important as the actual communication.

### How to work one-on-one and in small groups

One of the most common settings you will work in as a data professional is a one-on-one or small group meeting. As you prepare to communicate in these settings, remember that it is important to:

* Respect your colleagues' time by scheduling a meeting in advance
* Convey interest by practicing active listening
* Check for alignment by asking questions

## All about time

Time is a currency in the professional world. It’s very important to be efficient–this includes making sure your communication is understandable so that stakeholders can quickly comprehend your message. To ensure your message is clear and concise, remember to always:

* Use direct language
* Minimize wordiness
* Avoid unnecessary details
* Always strive for clarity
* Use proper grammar and punctuation
* Keep vocabulary simple and avoid technical language
* Break complex ideas into shorter sentences to make concepts easier to understand and remember

Not only will these suggestions help make your communication efficient and easier to understand, they will also save you time having to re-explain important concepts. And more than that, your colleagues will be grateful that you respected their time.

## Active Listening

As you begin your career as a data professional, you will spend a lot of time in meetings and in conversation. Many stakeholders are from different departments both inside and outside of your organization. The information shared during these interactions is valuable. Often, it’s where you gather insight into how the business operates, its goals, key milestones, and parameters within projects.

When you listen actively, you:

* Invite understanding of others
* Develop empathy for others and their responsibilities
* Build a connection with colleagues
* Promote trust

When you are practicing active listening, you make the effort to understand the speaker’s point of view. This helps you understand what other people are trying to communicate and sets you up to ask better, more insightful questions.

## Asking Questions

Data professionals aren’t automatically developing solutions. For data analysis to be effective, data professionals need to ask the right questions. In fact, the entire data analytical process depends on it.

Asking questions is a powerful communication tool. Asking the right questions can lead to institutional learning and a fruitful exchange of ideas. Many times, questions invite innovation and initiate efforts that can help improve projects and overall workflow.

Asking questions builds rapport and trust among team members. The right questions can often help mitigate business risks by uncovering unforeseen pitfalls and hazards. Here are a few tips to help guide your questioning:

* Ask questions that haven’t been answered already
* Ask questions that reveal the bigger picture
* Ask questions that gather information or further the knowledge of the team
* Ask questions that can help clarify misunderstandings

Effective questions are more likely to get you the answers you need to do your best work– which is good for the whole team.

## Key takeaways

In this reading, you learned that communication is present in all aspects of data professional work. Focus on the purpose, setting, and timing of your communication in order to promote more successful discussions with your team. Additionally, becoming an active listener that asks relevant questions enables more efficient communication and invites the perspective of other stakeholders. All these factors add up and make a big difference in how effectively you communicate as a data professional.

# Test your knowledge: Elements of communication

### 1.

Question 1

Which element of communicative exchange involves thinking about the reason why the communication is taking place?

Purpose

### 2.

Question 2

To work successfully, PACE must be employed in a linear progression.

False

### 3. Question 3

Fill in the blank: Regardless of the PACE workflow stage, \_\_\_\_\_ drives the framework through to project realization.

Communication

## Question

Fill in the blank: During the lessons involving career resources, you will be operating in the \_\_\_\_\_ stage as you collect information about your future as a data professional.

planning

# The value of the PACE strategy document

You have learned about the PACE workflow and how it helps bring structure to the data analytical process. This reading introduces the PACE strategy document. It is a resource designed to assist you in this program and throughout your career as a data analytics professional.

## Why do I need the PACE strategy document?

The videos, readings, activities, and projects in this program are a foundation for advancing your data skills and knowledge. But, success in the data career space involves more than data analytical skills. Data professionals are often involved in organizational decision making. This requires them to communicate about data and the results of data analysis with a variety of stakeholders.

In the advanced data analysis program, you’ll be introduced to different aspects of the field, which data analytics professionals encounter. Concepts and skills are arranged in a logical order and organized into courses that will prepare you for more advanced analytical tasks.

After each course, you'll be asked to develop an end-of-course project that demonstrates your mastery of expected data professional skills, such as your analytical and communication skills. To assist, you will receive an individual PACE strategy document for each of the course projects. Inside the PACE strategy documents, you’ll find helpful tips and opportunities to reflect on what you have learned and consider how to apply it to your job as a data analytics professional. Additionally, your responses within these strategy documents will help you create executive summaries that will inform decision makers and stakeholders of a project’s progress.

PACE strategy documents provide evidence of your expanded knowledge and can serve as a powerful motivator when acquiring new skills. Commitment to an educational program requires dedication and persistence. Your knowledge base and technical proficiency will be expanded through this journey. The gradual nature of personal growth makes it difficult to monitor day-to-day progress. Each PACE strategy document is designed to function as a record of your progress while developing new data analytical skills. The strategy documents will also help you improve communication skills by providing thoughtful questions designed to help identify and detail each step of your data analysis.

## A look at the PACE strategy document

You will find a PACE strategy document included with the end-of-course project for each course in this program. Even though these strategy documents were designed for the specific needs of a particular course, there are consistent elements in each of them.

### Instructions

In the “Instructions” section of the PACE strategy document, you will find general guidelines and special considerations for completing the document.

### Course project recap

In this section, the goals for each end-of-course project are outlined. By completing each task and the items needed for your projects, you will achieve these goals. No matter which workplace scenarios you select, the project goals will align with the knowledge you have gained in that course.

### Relevant interview questions

This section of the PACE strategy document builds context around your data tasks. The questions unlock a deeper understanding of data analytics, previewing the way you'll be able to speak after completing each course and its corresponding projects. During your job search, these questions can help you prepare for the types of questions you'll encounter during technical interviews.

### Review relevant course materials

In this section of the PACE strategy document you are given links to course materials you can review and reference as you complete each project. Links to relevant course materials provide you with quick access to relevant course information you need to complete your project.

### Reference guide

In the “Reference Guide,” you can find an outline of tasks required to complete the end-of-course project and the stage of PACE that each task addresses. Additionally, you can refer to the material when you are on the job completing similar tasks and projects later in your career.

### Data project questions & considerations

The questions in this section are specific to different stages of the PACE workflow. These “Data Project Questions & Considerations“ directly correspond to the questions you will encounter in the Jupyter notebooks for courses 2 - 6.  The answers you give to these questions will help you map out your thought process through each stage of the project.

## PACE strategy documents and Jupyter notebooks

Each end-of-course project  will also include a specially created Jupyter notebook containing helpful tips to assist you when programming project elements. Similar to the strategy documents, each notebook includes thought-provoking questions that will help guide you through the tasks of each end-of-course project. You find some questions appear in both locations, signifying relevance in both areas of the project. The information in these documents can be referenced together during the executing stages of each project to produce executive summaries to inform your team members and stakeholders.

## The benefits of the PACE strategy documents

Gathering content for a portfolio can be challenging without the proper resources. This is where preserving detailed records of your decision making may pay off. The more you consider the questions at each stage of a project, and preserve thoughtful responses, the more valuable these PACE strategy documents can be for  your future. Each PACE strategy document offers you a wealth of content that you can use to create a cohesive and branded portfolio. By creating a collection of your thoughts and internal processes, you develop a valuable resource you can reference throughout your career.

## Key takeaways

The PACE strategy document will help you complete the end-of-course projects by providing you with questions that deepen your understanding of the data analysis process. PACE strategy documents offer valuable insight into your personal workflow and provide a source of information that enhances your resume, portfolio, and job interview preparation.

# Communicate objectives with a project proposal

In this reading you will continue examining communication within the data workspace. As you have learned, communication is a key part of all aspects of data professional work. On a data team, project tasks and responsibilities are shared by different data professionals. Effective communication and collaboration among all team members and stakeholders is key to the success of any data project.

A **project proposal** can provide the structure and communication needed for tracking tasks. In addition, project proposals are beneficial for teams when facing challenges that require a high degree of flexibility. As projects progress, the expectations, resources, or even team members can change. This will require adjustments within a project that can impact the overall workflow and delivery date.

## Project proposals

A project proposal's main function is to outline objectives and requirements. Project proposals present ideas in more detailed and actionable segments often called **milestones**. Proposals are commonly created with input from team members and other stakeholders. It may also be the case that project proposals are shared with clients or executives to gain approval and inform them of a project’s path to completion. Project proposals are used across a multitude of industries and organizations. Although the design and layout of project proposals can vary, there are key elements that are common among them.

### Elements of a project proposal

Each project proposal contains important information that a team will need to consider before work begins. Below is a brief explanation of some common sections you will find in project proposals. Note that the format of project proposals will vary, so not every section described here will be included in every project proposal.

**Project title:** The title of the project is prominent, usually placed near the top of a document. Effective titles are brief and purposeful. Depending on the context and circumstances surrounding a project, the title can change over time.

**Project objective:** The objective statement is a one to three sentence explanation of what the project is trying to achieve.

**Milestones:** Milestones are groupings of tasks within a project, breaking the work needed into smaller, manageable goals. Milestones assist in the delegation and scheduling of work that needs to be completed within projects.

* The milestones in the provided example are representative of future end-of-course projects.

**Tasks:** Tasks detail the work that needs to be completed within a milestone.

* The tasks in the provided example parallel some of the work you will complete in upcoming end-of-course projects.

**Outcomes:** Outcomes are the completed actions or results that allow a project to continue.

**Deliverables:** Deliverables are items that can be shared amongst team members or with stakeholders. These are the end products of work undertaken for a project.

**Stakeholders:** The individuals or groups who are directly involved and have a vested interest in the success of a project. Input from stakeholders can serve as a basis for making decisions throughout a project.

**Estimated time:** At the beginning of a project, the time needed to complete milestones is estimated. As a project develops, these estimates will often need to be updated to account for adjustments to timelines or changes in team members.

### Sample project proposal

The sample project proposal, linked below, deals with a fictional visitation prediction project undertaken by the U.S. National Park Service (NPS). Use this document as a reference as you review each of the following sections.

This project proposal’s audience is the NPS data team. The purpose is to gather a comprehensive list of project tasks and divide them into smaller actionable groupings or milestones. Project proposals assist project managers in setting up task tracking, scheduling, and allocating resources. Furthermore, they serve as a reference for the team and as a valuable tool when new members are added to the project.

To access the sample project proposal, click the following link and select “Use Template.”

Link to sample project proposal: [Visitation prediction project proposal](https://docs.google.com/presentation/d/1ZfZLuP4eXtcuaVuKhkXSe1I66fRoH70Sjaze68IA3V0/template/preview)

OR

If you don’t have a Google account, you can download the file directly from the following attachment.

[Visitation prediction project proposal](https://d3c33hcgiwev3.cloudfront.net/DqL9Lw8RQSSe4lKRnk-0-w_1293cb61222c4e278986ab10148232f1_Visitation-prediction-project-proposal-.pptx?Expires=1719878400&Signature=SLPjPgKTYtEf9ONJdNP9cD-Rju1bmHgz-E7ORpB5NyWgTm~oXJQMz0d769ohBTwcRZ8beCcR2puKEbYo3t9yxwyuBCDAGEllDD40YTqRS00RFFH79Gl8x3Yb3PWZE0MFdBD5rZ162m3bpUsleUkYcGYDiEbGnIdZHHUOHzHXYMA_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[PPTX File](https://d3c33hcgiwev3.cloudfront.net/DqL9Lw8RQSSe4lKRnk-0-w_1293cb61222c4e278986ab10148232f1_Visitation-prediction-project-proposal-.pptx?Expires=1719878400&Signature=SLPjPgKTYtEf9ONJdNP9cD-Rju1bmHgz-E7ORpB5NyWgTm~oXJQMz0d769ohBTwcRZ8beCcR2puKEbYo3t9yxwyuBCDAGEllDD40YTqRS00RFFH79Gl8x3Yb3PWZE0MFdBD5rZ162m3bpUsleUkYcGYDiEbGnIdZHHUOHzHXYMA_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

## PACE and project proposals

In this course, you learned about the scalability of PACE (Plan, Analyze, Construct, and Execute). Through the PACE framework, projects can be organized globally by outlining their main tasks and deliverables. At the same time, each individual task within a project can be broken down into smaller actions.

You will discover that PACE strategy documents are a great reference when working on project proposals. During your end-of-course projects, you’ll be presented with questions that will assist you in identifying the planning, analyzing, constructing, and executing stages. The more time you spend considering and answering each question, the more information you’ll have available to you when creating project proposals.

## Key takeaways

A project proposal is a plan of action that outlines what needs to be accomplished and how to achieve your intended goals and outcomes. Proposals define a project’s purpose and scope, and list key milestones, deliverables, timelines, and schedules. It’s important to update proposals throughout the course of a project, as the project’s scope, objectives, and stakeholders may change over time.

# Connect PACE with executive summaries

In this reading, you will continue to examine communication within the data professional workplace. Regardless of workflow, data professionals need ways to share and communicate plans, updates, and summaries about projects. A document called an **executive summary** is used to update decision makers who may not be directly involved in the tasks of a project. In your role as a data professional, you will often be involved in creating executive summaries.

## Executive summaries

Executive summaries are documents that summarize the most important points about a project, giving decision makers a brief overview of the most relevant information. They can also be used to help new team members quickly become acquainted with a project. The format is designed to respect the responsibilities of decision makers and/or executives who may not have time to read and understand an entire report.

Executive summaries are used across numerous industries and organizations. There are many ways to present information within an executive summary, including software options built specifically for that purpose. In this program, you will primarily consider a one-page format within a presentation slide. Although the design and layout of executive summaries can vary, there are key elements that are common among them.

### Elements of an executive summary

Executive summaries are used across a wide variety of businesses and typically include the following elements:

**Project title:** A project's theme is incorporated into the executive summary title to create an immediate connection with the target audience.

**The problem:** A statement that focuses on the need or concern being targeted or addressed by the project. Note that the problem can also be referred to as the hypothesis that you’re trying to prove through data analysis.

**The solution:** This statement summarizes a project’s main goal. In this section, actions are described that address the concerns outlined in the problem statement.

**Details/Key Insights:** The purpose of this section is to provide any additional background information that may assist the target audience in understanding the project's objectives. Determining what details to include depends on the intended audience.

**Next steps/Recommendations:** Information that supports the actions the team plans to take. This can also include recommendations for decision makers based on the insights gained over the course of the project. In this section, a data professional may also include general project reflections. When you are adding to this section, include at least one point for recommendations and one for the suggested next steps.

### Sample executive summary

The following linked sample executive summary deals with a fictional visitation prediction project undertaken by the U.S. National Park Service (NPS). The intended audience of this summary is a group of decision makers from NPS leadership. The purpose of this summary is to share the insights gained through data analysis of recreational park visits. Each section delivers a short statement without embellishment. This allows decision makers to quickly grasp the most relevant points about a project. Reference this document as you review each of the following sections.

To access the sample executive summary, click the following link and select “Use Template.”

Link to sample executive summary: [Visitation prediction project executive summary](https://docs.google.com/presentation/d/1MQ2cXMvsWQfzTD6iAJqnj5Tvynrv4RCZQUnZie1ZuMo/template/preview?resourcekey=0-_Ck9PU4Bw_JHxrLqobHWXw)

OR

If you don’t have a Google account, you can download the file directly from the following attachment.

## PACE and executive summaries

In this course, you explored the PACE (Plan, Analyze, Construct, Execute) workflow and how it can help guide projects. Through PACE, the tasks and deliverables of a project are clearly identified and recorded in a PACE strategy document.

You will discover that PACE strategy documents are a great reference when working on executive summaries. When planning, analyzing, constructing, and executing your end-of-course and capstone projects, the PACE strategy documents provide questions to guide you. The more time you spend considering and answering each question, the more information you’ll have available to you when creating executive summaries.

## Key takeaways

Executive summaries offer an effective way to share information with decision makers, clients, and executives. These documents summarize the most important information within a project or plan of action, and share key insights and results. Typically, an executive summary reports on an identified problem and outlines the solutions used to address the problem.

# Activity: Create a project proposal

## Activity Overview



As you have learned, a project proposal is a plan of action that describes what needs to be accomplished in order to achieve a project’s intended goals and outcome. The main function of this resource is to outline a project’s objectives and requirements. Project proposals are typically generated at the beginning of a project and are used by team members throughout the project’s duration.

In this activity, you will create a project proposal based on a fictional workplace scenario.

To review tips for creating an effective project proposal, refer to the reading about [communicating objectives with a project proposal](https://www.coursera.org/learn/foundations-of-data-science/supplement/79Ysh/communicate-objectives-with-a-project-proposal).

Be sure to complete this activity before moving on. The next course item will provide you with a completed exemplar to compare to your own work. You will not be able to access the exemplar until you have completed this activity.

## Scenario



Review the following scenario. Then complete the step-by-step instructions.

You are a data professional working for an international delivery company.  Company leadership has asked the data team to develop a machine learning model for predictive maintenance on its fleet of delivery vehicles.Predictive maintenance uses machine learning to help predict equipment failures before they occur. The ML model analyzes both historical data such as performance and maintenance records, and real-time data collected from sensors placed on vehicles.

Leadership wants to use the power of predictive maintenance to monitor the performance of delivery vehicles, and to identify and fix issues before they cause costly downtime or delays. For example, if the engine of a delivery vehicle breaks down, the company has to deal with delivery delays, dissatisfied customers, vehicle repair or replacement, and potential safety issues. Predictive maintenance offers benefits such as optimized delivery, increased safety, improved customer service, and reduced costs.

Your manager asks you to create a project proposal to identify and organize key milestones and deliverables for the project. The proposed timeline for the project is 12 weeks. The goal is to build a model with at least 90% accuracy.

## Step-By-Step Instructions



Follow the instructions to complete each step of the activity. Then, answer the five questions at the end of the activity before going to the next course item to compare your work to a completed exemplar.

### **Step 1: Access the template**



To use the supporting materials for this course item, click the following link and select “Use Template.”

Link to supporting materials:  [Predictive maintenance project proposal](https://docs.google.com/presentation/d/1zHQmJzwSIE-atFCjLC46R_eZ62_LyFASLPEXbUmUrdM/template/preview)

OR

If you don’t have a Google account, you can download the supporting materials directly from the following attachment.

[Predictive maintenance project proposal](https://d3c33hcgiwev3.cloudfront.net/6QuF3hW1SLO335h2QKCZ_A_52087ea3c604413098c38495f39497f1_Predictive-maintenance-project-proposal-.pptx?Expires=1719878400&Signature=Uq19BLMV5x2hDl90y9bdP02BETPloOCTReHATKXHYg5P1zkmPgN38Qtl8m9ACos~9d4EBrdLfPAewDg~nApPvtUmBp3gGpA4LfLZqNTMnwA1WfFnwM2itKOL-Y7G9LXUWxbBqdwGWzgp7-f6A6~8mmC8Urja~sMGuEx4SNiURss_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[PPTX File](https://d3c33hcgiwev3.cloudfront.net/6QuF3hW1SLO335h2QKCZ_A_52087ea3c604413098c38495f39497f1_Predictive-maintenance-project-proposal-.pptx?Expires=1719878400&Signature=Uq19BLMV5x2hDl90y9bdP02BETPloOCTReHATKXHYg5P1zkmPgN38Qtl8m9ACos~9d4EBrdLfPAewDg~nApPvtUmBp3gGpA4LfLZqNTMnwA1WfFnwM2itKOL-Y7G9LXUWxbBqdwGWzgp7-f6A6~8mmC8Urja~sMGuEx4SNiURss_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

### **Step 2: Complete the project proposal**

Review the scenario and think about the objective, scope, timeline, and key tasks of the project. The project proposal should include the following tasks:

* Data collection
* Data cleaning
* Data exploration
* Building and testing of ML models
* Sharing results/insights with stakeholders

Consider what information would be most relevant to the data team. Then complete the project proposal with that information.

Note that the project template has been organized according to the PACE framework (Plan, Analyze, Construct, Execute).

### **Pro Tip: Save the template**

Finally, be sure to save a blank copy of the template you used to complete this activity. You can use it for further practice or in your professional projects. These templates will help you work through your thought processes and demonstrate your experience to potential employers.



Be sure to address the following elements in your completed project proposal::

* A project title
* An objective statement that describes the goal of the project
* A list of key tasks and deliverables to be completed for each project milestone
* A time estimate for achieving each project milestone

Question 1

## Did you complete this activity?

Yes

Question 2

What is the main goal of the project?

Predict vehicle failures before they occur

Question 3

Approximately how long should it take to complete the planning and analysis stages of the project?

4-6 weeks

Question 4

During which stages of the project should the team share project progress and findings with stakeholders? Select all that apply.

- Analyzing

-Constructing

-Planning

-Executing

Question 5

Which deliverables are associated with the tasks of finalizing modeling and machine learning decisions, constructing a model, and testing machine learning techniques for accuracy? Select all that apply.

- Machine learning model

- Stakeholders updated

# Test your knowledge: Communicate like a data professional

* 1. Which PACE stage involves acquiring new skills for data work?

Plan

* 1. Which PACE stage involves preparing data in order to reveal the stories data tells?

Analyze

# Glossary terms from module 4

# Terms and definitions from Course 1, Module 4

**PACE workflow**:A framework that provides an initial structure to guide the process of data analytics; PACE stands for plan, analyze, construct, and execute

**Plan stage**:Stage of the PACE workflow where the scope of a project is defined and the informational needs of the organization are identified

**Analyze stage**:Stage of the PACE workflow where the necessary data is acquired from primary and secondary sources and then cleaned, reorganized, and analyzed

**Construct stage**:Stage of the PACE workflow where data models and machine learning algorithms are built, interpreted, and revised to uncover relationships within the data and help unlock insights from those relationships

**Execute stage**: Stage of the PACE workflow where a data professional will present findings with internal and external stakeholders, answer questions, consider different viewpoints, and make recommendations

# Terms and definitions from previous modules

## A

**Active listening:** Refers to allowing team members, bosses, and other collaborative stakeholders to share their own points of view before offering responses

**Aggregate information**:Data from a significant number of users that has eliminated personal information

**Analytics Team Manager**: A data professional who supervises analytical strategy for an organization, often managing multiple groups

**Artificial intelligence (AI):** Refers to computer systems able to perform tasks that normally require human intelligence

## B

**Business Intelligence Analyst**: (Refer to **Business Intelligence Engineer**)

**Business Intelligence Engineer:** A data professional who uses their knowledge of business trends and databases to organize information and make it accessible; also referred to as a Business Intelligence Analyst

## C

**Chief Data Officer**: An executive-level data professional who is responsible for the consistency, accuracy, relevancy, interpretability, and reliability of the data a team provides

## D

**Data anonymization**: The process of protecting people's private or sensitive data by eliminating PII

**Data cleaning**: The process of formatting data and removing unwanted material

**Data Engineer:** A data professional who makes data accessible, ensures data ecosystems offer reliable results, and manages infrastructure for data across enterprises

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

**Data science**: The discipline of making data useful

**Data Scientist**: A data professional who works closely with analytics to provide meaningful insights that help improve current business operations

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

## E

**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

## H

**Hackathon:** An event where programmers and data professionals come together and work on a project

## I

**Interpersonal skills:** Traits that focus on communicating and building relationships

## J

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

## M

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

**Mentor:** Someone who shares knowledge, skills, and experience to help another grow both professionally and personally

**Metrics**: Methods and criteria used to evaluate data

## N

**Nonprofit**: A group organized for purposes other than generating profit; often aims to further a social cause or provide a benefit to the public

## O

**Open data**: Data that is available to the public and free to use, with guidance on how to navigate the datasets and acknowledge the source

## P

**Personally identifiable information (PII):** Information that permits the identity of an individual to be inferred by either direct or indirect means

**Python**: A general-purpose programming language

## R

**RACI chart**: A visual that helps to define roles and responsibilities for individuals or teams to ensure work gets done efficiently; lists who is responsible, accountable, consulted, and informed for project tasks

## S

**Sample:** A segment of a population that is representative of the entire population

## T

**Tableau**: A business intelligence and analytics platform that helps people visualize, understand, and make decisions with data

# Module 4 challenge (Not all questions are listed here)

Question 1

In the plan stage of the PACE model, what questions might a data professional ask themself? Select all that apply.

-How will the business or operation be affected by the plan?

-What are the goals of the project?

-What strategies will be needed?

Question 2

Fill in the blank: In the construct stage of the PACE model, data professionals build, interpret, and revise models and \_\_\_\_\_ in order to uncover relationships within the data and unlock related insights.

machine learning algorithms

Question 3 (NOT FINAL)

A data professional uses their communication skills to present findings from a data project to others. They also answer questions and consider different viewpoints. What stage of the PACE model does this scenario describe?

- Construct

Question 4

All communicative exchanges contain key elements. Which element involves considering the audience of a particular communication?

Receiver

Question 5

Which of the following are best practices when communicating about data in a professional setting? Select all that apply.

Use proper punctuation.

Use direct language.

Be efficient.

Question 6

Fill in the blank: Effective \_\_\_\_\_ is used to improve the workflow in all PACE framework stages.

Communication

Question 7

What are some strategies to use when practicing active listening? Select all that apply.

Place a high value on the insights gained from others

 Be curious

Take notes in meetings for later reference

Question 8

Which section of a project proposal is a brief explanation of what the project is trying to achieve?

Project objective