**Applications and Careers in Data Science**

In the first lesson, you learn about the power of data science applications and how organizations leverage this power to drive business goals, improve efficiency, make predictions, and even save lives. You also reviewed the process you will follow as a data scientist to help your organization accomplish these ends. In the second lesson, you investigate what companies seek in a competent, experienced data scientist. You will learn how to position yourself to get hired as a data scientist. Amidst the diverse backgrounds from which data scientists emerge, you identify the qualities they share and skills that consistently set them apart from other data-related roles. You will complete a peer-reviewed final project by looking at a job posting for data scientist and identifying commonalities between the job and what you learned in this course. You will also walk through a case study, where you learn about Sarah and her data science journey.

**Learning Objectives**

* Describe the contents of a data science job posting.
* Describe the application of data science in healthcare.
* Explain how companies can start on their data science journey.
* Describe some of the ways in which data is generated by consumers.
* Describe how businesses such as Netflix, Amazon, UPS, Google, and Apple are using data generated by their consumers and employees.
* Compare some of the qualities that differentiate data scientists from qualities of other data professionals.
* Articulate the purpose of the final deliverable of a data science project and the role of storytelling in the final deliverable.
* Describe what the final report of a Data Science project should cover and how it should be structured for best results.
* Demonstrate your understanding of data science by articulating what data scientists do and what a data science report contains.

# **Data Science Application Domains**

## **Lesson Overview: Data Science Application Domains**

In this lesson, "Data Science Application Domains," you'll embark on a journey to explore the vast and impactful realms where data science plays a pivotal role. This engaging module delves into various activities that shed light on the diverse applications of data science in our world today. You'll uncover how data science drives innovation and transformation across different sectors, from revolutionizing industries to saving lives.

Dive into this lesson to discover the real-world applications that define the dynamic landscape of data science.

|  |  |
| --- | --- |
| **Asset name and type** | **Description** |
| “How Should Companies Get Started in Data Science?” video | Gain insights into how organizations can embark on their data science journey effectively. |
| “Old problems with New Data Science Solutions” video | Discover how data science offers innovative solutions to age-old and real-world problems. |
| “Applications of Data Science” video | Explore the wide-ranging applications of data science across various industries and sectors. |
| “How Data Science is saving lives” video | In this video, you will learn about the life-saving potential of data science in healthcare and beyond. |
| “The Final Deliverable” reading | Dive into the details of what constitutes the final deliverable in data science projects. |
| Practice quiz | Test your understanding of the previous reading. |
| “Lesson Summary” video | Recap the essential takeaways from this module with a lesson summary. |
| Practice quiz | Take a practice quiz to evaluate how well you’ve understood the material presented in this lesson. |
| Glossary | Use this glossary of terms to review the terminology presented in this lesson. |

## **How should companies get started with data science**

the importance of measuring data in a business.

In simple terms, businesses need to keep track of their costs and profits to improve and grow. Imagine trying to bake a cake without measuring the ingredients; you might end up with a disaster! Similarly, if a company doesn’t measure its expenses, like how much it spends on materials and labor, it can’t figure out how to save money or make more profit. The first step for any business is to start collecting this important information. Once they have the data, they can analyze it to make better decisions.

Think of data as a treasure map. If you don’t have the map, you won’t know where to dig for treasure. But if you collect and keep the map safe, you can always refer back to it to find the best spots to dig. So, businesses should not only capture data but also keep it organized and accessible for future use.

Here's a brief summary of the topic:

* **Importance of Measurement**: Businesses must measure costs and profits to improve and grow.
* **Data Collection**: The first step is to start capturing data about expenses (like labor and materials) and revenue.
* **Data Analysis**: Once data is collected, it can be analyzed to make informed decisions.
* **Data Preservation**: Companies should archive their data properly, as it remains relevant over time and can provide insights for future improvements.

## [**Old Problems, New Data Science Solutions**](https://www.coursera.org/learn/what-is-datascience/lecture/MQKK1/old-problems-new-data-science-solutions)

data science to solve problems, which is a key theme in the content.

Data Science for Problem Solving: A Simple Explanation  
Data science is like a treasure map that helps organizations find the best solutions to their problems. Just like a detective gathers clues to solve a mystery, data scientists collect and analyze data to understand what’s happening in a situation. For example, Uber uses data to figure out how many drivers are needed in a specific area at a certain time, ensuring that riders get a ride quickly and at a fair price.

Imagine a busy restaurant that wants to serve customers faster. By looking at data about when customers usually come in, the restaurant can schedule more staff during peak hours. This way, they can serve more people without making them wait too long. Similarly, data science helps organizations make informed decisions by understanding patterns and trends in the data they collect.

summary of the key points regarding data science and its applications:

* **Purpose of Data Science**: Organizations use data science to find optimal solutions to existing problems by analyzing large amounts of data.
* **Real-World Examples**:
  + **Uber**: Collects real-time data to manage driver availability and pricing, ensuring efficient service for riders.
  + **Toronto Transportation Commission**: Analyzes traffic data to improve streetcar operations and reduce congestion, leading to less time lost for commuters.
* **Environmental Applications**: Data science is also used to address environmental issues, such as monitoring harmful cyanobacterial blooms in lakes using advanced technology.
* **Process for Solutions**:
  + Identify the problem.
  + Gather relevant data.
  + Choose the right tools and develop a data strategy.
  + Analyze the data to gain insights and develop solutions.

This approach helps organizations refine their practices and make data-driven decisions. Would you like to explore any specific aspect further?

## **How Data Science is saving lives**

**Predictive Analytics in Healthcare**.

Predictive analytics in healthcare is like having a crystal ball that helps doctors see the best treatment options for their patients. By using data science techniques, healthcare professionals can analyze a lot of information about diseases, such as genetic factors, other health conditions, and even environmental influences. This analysis helps them recommend the most suitable tests and treatments tailored to each patient’s unique situation.

Imagine you’re trying to choose the best route for a road trip. You would look at traffic reports, weather conditions, and maybe even ask friends about their experiences. Similarly, doctors use predictive analytics to gather all the relevant data about a patient’s health to make informed decisions. This way, every doctor can access the latest information and provide the best care possible, ultimately improving patient outcomes.

summary of **Predictive Analytics in Healthcare**:

* **Purpose**: Predictive analytics helps healthcare professionals make informed decisions about patient care.
* **Data Analysis**: It involves analyzing various data points, such as genetic markers, health conditions, and environmental factors.
* **Recommendations**: This analysis allows doctors to recommend appropriate tests and treatments tailored to individual patients.
* **Improved Outcomes**: By ensuring all physicians have access to the latest information, predictive analytics enhances patient care and outcomes.

## **The Final Deliverable**

**The Final Deliverable**

The ultimate purpose of analytics is to communicate findings to the concerned who might use these insights to formulate policy or strategy. Analytics summarize findings in tables and plots. The data scientist should then use the insights to build the narrative to communicate the findings. In academia, the final deliverable is in the form of essays and reports. Such deliverables are usually 1,000 to 7,000 words in length.

In consulting and business, the final deliverable takes on several forms. It can be a small document of fewer than 1,500 words illustrated with tables and plots, or it could be a comprehensive document comprising several hundred pages. Large consulting firms, such as McKinsey and Deloitte, I routinely generate analytics-driven reports to communicate their findings and, in the process, establish their expertise in specific knowledge domains.

Let's review the "United States Economic Forecast", a publication by the Deloitte University Press. This document serves as a good example for a deliverable that builds narrative from data and analytics. The 24-page report focuses on the state of the U.S. economy as observed in December 2014. The report opens with a **grabber** highlighting the fact that contrary to popular perception, the economic and job growth has been quite robust in the United States. The report is not merely a statement of facts.

In fact, it is a carefully crafted report that cites Voltaire and follows a distinct theme. The report focuses on the **good news** about the U.S. economy. These include the increased investment in manufacturing equipment in the U.S. and the likelihood of higher consumer consumption resulting from lower oil prices.

The Deloitte report uses time series plots to illustrate trends in markets. The GDP growth chart shows how the economy contracted during the Great Recession and has rebounded since then. The graphic presents four likely scenarios for the future. Another plot shows the changes in consumer spending. The accompanying narrative focuses on income inequality in the U.S. and refers to Thomas Pikkety's book on the same. The Deloitte report mentions many consumers did not experience an increase in their real incomes over the years, while they still maintained their level of spending. Other graphics focused on housing, business, and government sectors, international trade, labor, and financial markets, and prices. The appendix carries four tables documenting data for the four scenarios discussed in the report.

Deloitte's "United States Economic Forecast" serves the very purpose that its authors intended. The report uses data and analytics to generate the likely economic scenarios. It builds a powerful narrative in support of the thesis statement that the U.S. economy is doing much better than most would like to believe. At the same time, the report shows Deloitte to be a competent firm capable of analyzing economic data and prescribing strategies to cope with the economic challenges.

Now consider if we were to exclude the narrative from this report and presented the findings as a deck of PowerPoint slides with eight graphics and four tables. The PowerPoint slides would have failed to communicate the message that the authors carefully crafted in the report citing Piketty and Voltaire. I consider Deloitte's report a good example of storytelling with data and encourage you to read the report to decide for yourself whether the deliverable would have been equally powerful without the narrative.

Now, let us work backward from the Deloitte report. Before the authors started their analysis, they must have discussed the scope of the final deliverable. They would have deliberated the key message of the report and then looked for the data and analytics they needed to make their case. The initial planning and conceptualizing of the final deliverable is therefore extremely important for producing a compelling document. Embarking on analytics, without due consideration to the final deliverable, is likely to result in a poor-quality document where the analytics and narrative would struggle to blend.

## **Lesson Summary: Data Science and Application Domain**

**data science applications** in simple terms.

Data science applications are like powerful tools that help organizations solve problems and make better decisions using data. Imagine you have a big puzzle, and each piece represents a piece of information. By putting these pieces together, organizations can see the bigger picture, understand what’s happening, and find the best solutions to their challenges. For example, a company like Amazon uses data science to recommend products you might like based on your previous searches, just like a friend suggesting a movie you’d enjoy!

To get started with data science, organizations first need to identify the problem they want to solve. This is like figuring out what puzzle you’re trying to complete. Then, they gather and measure data, which is crucial because if you don’t measure something, it’s hard to improve it. Once they have the data, they can clean it up and analyze it to find patterns and insights. This process can lead to better business strategies, improved efficiency, and even life-saving decisions in fields like healthcare.

summary of **data science applications**:

* **Purpose**: Data science helps organizations solve problems and make informed decisions using data.
* **Process**:
  + **Identify the Problem**: Understand what needs to be solved.
  + **Gather Data**: Collect relevant information, as measurement is key to improvement.
  + **Analyze Data**: Clean and examine the data to find patterns and insights.
* **Examples**:
  + **Amazon**: Uses data to recommend products based on user behavior.
  + **Healthcare**: Utilizes predictive analytics to suggest treatments and improve patient care.
* **Benefits**: Enhances business strategies, improves efficiency, and can save lives.

## **Glossary: Data Science Application Domains**

Welcome! This alphabetized glossary contains many of the terms in this course. These terms are important for you to recognize when working in the industry, participating in user groups, and participating in other certificate programs.

|  |  |  |
| --- | --- | --- |
| **Term** | **Definition** | **Video where the term is introduced** |
| Arithmetic Models | Data science often uses Mathematical models to analyze data and predict outcomes. | Old problems, new data science solutions |
| Case study | In-depth analysis of an instance of a chosen subject to draw insights that inform theory, practice, or decision-making. | Old problems, new data science solutions |
| Data mining | Extracting information from raw data, such as making decisions, predicting trends, or understanding phenomena. | How Data Science is Saving Lives |
| Data Science | The field involves collecting, analyzing, and interpreting data to extract valuable insights and make informed decisions. | Old problems, new data science solutions |
| Data Strategy | A plan that outlines how an organization will collect, manage, and use data to achieve its goals. | Old problems, new data science solutions |
| Predictive analytics | Using data, algorithms, models, and machine learning to make predictions. | How Data Science is Saving Lives |

# **Careers and Recruiting in Data Science**

## **Lesson overview: Careers and Recruiting in Data Science**

In the lesson "Data Science Application Domains," you'll embark on a journey to explore the vast and impactful realms where data science plays a pivotal role. This engaging module delves into various activities that shed light on the diverse applications of data science in our world today.

Dive into this lesson to discover the real-world applications that define the dynamic landscape of data science.

|  |  |
| --- | --- |
| **Asset name and type** | **Description** |
| “How Can Someone Become a Data Scientist?” video | Discover the pathways to becoming a proficient data scientist, exploring the skills and knowledge required. |
| “Recruiting for Data Science” video | Gain insights into organizations' strategies and considerations for recruiting data science talent. |
| “Careers in Data Science” video | Explore the diverse career opportunities and roles available in the dynamic field of data science. |
| “Importance of Mathematics and Statistics for Data Science” video | Understand the fundamental role of mathematics and statistics in data science, emphasizing their significance. |
| “The Report Structure” reading | Delve into the intricacies of structuring reports within data science projects, enhancing your understanding of this essential aspect. |
| Practice quiz | Test your understanding of the previous reading. |
| “Infograph on Roadmap” reading | Explore an informative infographic detailing the roadmap for success in data science careers. |
| “Lesson Summary” video | Recap the essential takeaways from this module with a lesson summary. |
| Practice quiz | Take a practice quiz to evaluate how well you’ve understood the material presented in this lesson. |
| Glossary | Use this glossary of terms to review the terminology presented in this lesson. |
| Grade Quiz | Evaluate your knowledge of data science in a business setting with this graded quiz. |
| “Data Science in Business” Reading | Summarize your learning journey in this module, reviewing the key takeaways and insights gained. |

## **How can Someone Become a Data Scientist?**

skills and background needed to become a data scientist. Here’s a simplified breakdown:

1. **Educational Background**:
   * Many data scientists have advanced degrees (like PhDs) in fields such as **physics**, **statistics**, or **computer science**.
   * A strong foundation in **math** (like algebra, geometry, and calculus) is important.
2. **Programming Skills**:
   * Knowing how to **program** is essential. This means being able to write code to analyze data.
3. **Understanding Data**:
   * Familiarity with **databases** is crucial. This includes knowing how to store and retrieve data effectively.
   * Understanding different types of **statistical distributions** helps in analyzing data correctly.
4. **Hands-On Learning**:
   * Practical experience is key. Engaging with tools like **Jupyter Notebooks** allows you to apply what you learn in real-time.
5. **Self-Learning**:
   * Many data scientists are self-taught. They read articles, watch videos, and experiment with projects to build their skills.
6. **Motivation**:
   * Staying motivated is important. Setting clear goals can help you stay on track in your learning journey.

summary of the topic "How can Someone Become a Data Scientist?":

* **Educational Background**: Many data scientists have advanced degrees in fields like physics, statistics, or computer science, with a strong foundation in math.
* **Programming Skills**: Proficiency in programming is essential for data analysis.
* **Understanding Data**: Knowledge of databases and statistical distributions is crucial for effective data handling.
* **Hands-On Learning**: Practical experience using tools like Jupyter Notebooks is important for applying concepts.
* **Self-Learning**: Many data scientists learn independently through articles, videos, and projects.
* **Motivation**: Staying motivated and setting clear goals is key to success in the field.

**[Careers in Data Science](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[how companies hire data scientists and what qualities they should look for. Here are the key points in simple terms:](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)

* **[Finding the Right Person](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[: Companies often want a "unicorn" – someone who has all the skills and knowledge. However, it's rare to find someone like that.](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)
* **[Important Qualities](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[:](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)
  + **[Curiosity](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[: The person should be curious about many things, not just data science.](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)
  + **[Sense of Humor](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[: A lighthearted attitude helps in problem-solving.](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)
  + **[Storytelling Skills](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[: They should be able to explain their findings in a way that others can understand and appreciate.](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)
* **[Technical Skills](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[: While technical skills are important, they can be taught. What matters more is the person's passion for the field and their ability to communicate effectively.](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)
* **[Understanding Roles](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[: Before hiring, companies need to know what roles they need in their data science team, whether it's for data analysis, engineering, or visualization.](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)
* **[Choosing Tools](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**[: Depending on the type of data work (like market research or big data), different technical skills and tools are required.](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)

[This video emphasizes that while technical skills are important, qualities like curiosity and communication are equally crucial for success in data science.](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)

summary of the video:

* **Hiring Data Scientists**: Companies often seek a "unicorn" with all the skills, but it's rare to find such candidates.
* **Key Qualities**:
  + **Curiosity**: Candidates should be curious about various topics.
  + **Sense of Humor**: A lighthearted approach aids in problem-solving.
  + **Storytelling Skills**: The ability to communicate findings effectively is essential.
* **Technical Skills**: While important, these can be taught. Passion and communication skills are more critical.
* **Understanding Roles**: Companies need to define the roles they require in their data science teams.
* **Choosing Tools**: The type of data work determines the necessary technical skills and tools.

This video highlights the importance of both personal qualities and technical skills in building an effective data science team.

## **[Careers in Data Science](https://www.coursera.org/learn/what-is-datascience/lecture/8ey50/careers-in-data-science)**

we’re focusing on is the growing importance of data science in today's world. In simple terms, data science is the practice of using data to gain insights and make informed decisions. With the rise of the Internet of Things and advanced computing, we now have access to vast amounts of data. Data scientists help organizations understand this data and use it to improve their operations, products, and services.

For example, think of a gardener who wants to grow the best flowers. They would collect information about the soil, weather, and types of flowers that thrive in their area. By analyzing this data, they can make better choices about what to plant and when to water. Similarly, companies use data science to analyze trends and customer preferences, helping them to create better products and services that meet people's needs.

Here's a brief summary of the topic:

* **Data Science**: A field that uses data to uncover insights and make informed decisions.
* **Importance**: With the growth of the Internet of Things and advanced computing, vast amounts of data are available for analysis.
* **Career Opportunities**: Data science is one of the most promising careers, with high demand across various industries.
* **Skills Needed**: To succeed in data science, one should enjoy working with data, coding, and learning math and statistics.
* **Continuous Learning**: Data scientists must keep updating their skills and knowledge due to the evolving nature of the field.

## **Importance of Mathematics and Statistics for Data Science**

**importance of math and statistics** in data science. Here are the key points explained simply:

* **Foundation for Analysis**: Math and statistics provide the tools needed to analyze data effectively. They help in understanding patterns and making predictions.
* **Problem-Solving**: When you encounter a data-related problem, math helps you figure out the best way to approach it. For example, knowing how to calculate averages or probabilities can guide your decisions.
* **Real-World Applications**: Data scientists use math and statistics to interpret data from various fields, such as business, healthcare, and social sciences. This helps organizations make informed choices based on data insights.
* **Learning Curve**: While math may seem challenging, connecting it to real-life problems makes it easier to grasp. Many successful data scientists have improved their math skills over time by applying them to practical situations.

importance of math and statistics for data science:

* **Essential Tools**: Math and statistics are crucial for analyzing data and making predictions.
* **Problem-Solving**: They help data scientists approach and solve data-related problems effectively.
* **Real-World Use**: These skills are applied across various fields, aiding organizations in making informed decisions.
* **Learning and Improvement**: Many successful data scientists enhance their math skills by applying them to real-life scenarios, making the learning process more relatable.

## **The Report Structure**

**The Report Structure**

Before starting the analysis, think about the structure of the report. Will it be a brief report of five or fewer pages, or will it be a longer document running more than 100 pages in length? The structure of the report depends on the length of the document. A brief report is more to the point and presents a summary of key findings. A detailed report incrementally builds the argument and contains details about other relevant works, research methodology, data sources, and intermediate findings along with the main results.

I have reviewed reports by leading consultants including Deloitte and McKinsey. I found that the length of the reports varied depending largely on the purpose of the report. Brief reports were drafted as commentaries on current trends and developments that attracted public or media attention. Detailed and comprehensive reports offered a critical review of the subject matter with extensive data analysis and commentary. Often, detailed reports collected new data or interviewed industry experts to answer the research questions.

Even if you expect the report to be brief, sporting five or fewer pages, I recommend that the deliverable follow a prescribed format including the cover page, table of contents, executive summary, detailed contents, acknowledgments, references, and appendices (if needed).

I often find the cover page to be missing in documents. It is not the inexperience of undergraduate students that is reflected in submissions that usually miss the cover page. In fact, doctoral candidates also require an explicit reminder to include an informative cover page. I hasten to mention that the business world sleuths are hardly any better. Just search the Internet for reports and you will find plenty of reports from reputed firms that are missing the cover page.

At a minimum, the cover page should include the title of the report, names of authors, their affiliations, and contacts, the name of the institutional publisher (if any), and the date of publication. I have seen numerous reports missing the date of publication, making it impossible to cite them without the year and month of publication. Also, from a business point of view, authors should make it easier for the reader to reach out to them. Having contact details at the front makes the task easier.

"A table of contents (ToC)" is like a map needed for a trip never taken before. You need to have a sense of the journey before embarking on it. A map provides a visual proxy for the actual travel with details about the landmarks that you will pass by in your trip. The ToC with main headings and lists of tables and figures offers a glimpse of what lies ahead in the document. Never shy away from including a ToC, especially if your document, excluding cover page, table of contents, and references, is five or more pages in length.

Even for a short document, I recommend an "abstract" or an "executive summary". Nothing is more powerful than explaining the crux of your arguments in three paragraphs or less. Of course, for larger documents running a few hundred pages, the executive summary could be longer.

An "introductory section" is always helpful in setting up the problem for the reader who might be new to the topic and who might need to be gently introduced to the subject matter before being immersed in intricate details. A good follow-up to the introductory section is a review of available relevant research on the subject matter. The length of the literature review section depends upon how contested the subject matter is. In instances where the vast majority of researchers have concluded in one direction, the literature review could be brief with citations for only the most influential authors on the subject. On the other hand, if the arguments are more nuanced with caveats aplenty, then you must cite the relevant research to offer adequate context before you embark on your analysis. You might use the literature review to highlight gaps in the existing knowledge, which your analysis will try to fill. This is where you formally introduce your research questions and hypothesis.

In the "methodology" section, you introduce the research methods and data sources you used for the analysis. If you have collected new data, explain the data collection exercise in some detail. You will refer to the literature review to bolster your choice for variables, data, and methods and how they will help you answer your research questions.

The results section is where you present your empirical findings. Starting with descriptive statistics (**see Chapter 4, "Serving Tables"**) and illustrative graphics (**see Chapter S, "Graphic Details" for plots and Chapter 10, "Spatial Data Analytics" for maps**), you will move toward formally testing your hypothesis (**see Chapter 6, "Hypothetically Speaking"**).

In case you need to run statistical models, you might turn to regression models (**see Chapter 7, "Why Tall Parents Don't Have Even Taller Children"**) or categorical analysis (**see Chapters 8, "To Be or Not to Be" and 2., "Categorically Speaking About Categorical Data"**). If you are working with time-series data, you can turn to Chapter 11, **Doing Serious Time with Time Series.** You can also report results from other empirical techniques that fall under the general rubric of data mining (**see Chapter 12, "Data Mining for Gold"**). Note that many reports in the business sector present results in a more palatable fashion by holding back the statistical details and relying on illustrative graphics to summarize the results.

The results section is followed by the discussion section, where you craft your main arguments by building on the results you have presented earlier.

The "discussion section" is where you rely on the power of narrative to enable numbers to communicate your thesis to your readers. You refer the reader to the research question and the knowledge gaps you identified earlier. You highlight how your findings provide the ultimate missing piece to the puzzle.

Of course, not all analytics return a smoking gun. At times, more frequently than I would like to acknowledge, the results provide only a partial answer to the question and that, too, with a long list of caveats.

In the "conclusion" section, you generalize your specific findings and take on a rather marketing approach to promote your findings so that the reader does not remain stuck in the caveats that you have voluntarily outlined earlier. You might also identify future possible developments in research and applications that could result from your research.

What remains is housekeeping, including a list of references, the acknowledgment section (**acknowledging the support of those who have enabled your work is always good**), and "appendices", if needed.

**Have You Done Your Job as a Writer?**

As a data scientist, you are expected to do thorough analysis with the appropriate data, deploying the appropriate tools. As a writer, you are responsible for communicating your findings to the readers. Transport Policy, a leading research publication in transportation planning, offers a checklist for authors interested in publishing with the journal. The checklist is a series of questions authors are expected to consider before submitting their manuscripts to the journal. I believe the checklist is useful for budding data scientists and, therefore, I have reproduced it verbatim for their benefit.

* Have you told readers, at the outset, what they might gain by reading your paper?
* Have you made the aim of your work clear?
* Have you explained the significance of your contribution?
* Have you set your work in the appropriate context by giving sufficient background (including a complete set of relevant references) to your work?
* Have you addressed the question of practicality and usefulness?
* Have you identified future developments that might result from your work?
* Have you structured your paper in a clear and logical fashion?

## **Lesson Summary: Careers and Recruiting in Data Science**

concept of teamwork in data science, which is a key point from the content.

**Teamwork in Data Science: A Simple Explanation**  
In the world of data science, companies often look for a team of people with different skills rather than just one person who can do everything. This is because data science involves many areas, like understanding data, programming, and storytelling. Each team member can bring their unique strengths to the table, making the team more effective.

**An Engaging Example**  
Think of a data science team like a sports team. Just as a basketball team has players with different roles—like shooters, defenders, and playmakers—data scientists also have different specialties. One person might be great at analyzing numbers, another might excel at coding, and someone else could be a fantastic storyteller who presents the findings. When they work together, they can achieve amazing results, just like a well-coordinated sports team winning a game!

concise summary of the key points regarding what companies seek in data scientists:

* **Team Composition**: Companies prefer to build teams with diverse skills rather than finding one individual with all the desired qualities.
* **Key Traits**: Important traits for data scientists include **curiosity**, **passion** for their field, and the ability to ask meaningful questions.
* **Skill Set**: Essential skills include:
  + **Mathematics and Statistics**: For analyzing data.
  + **Programming**: Common languages are Python and R.
  + **Data Management**: Understanding data storage and retrieval systems.
  + **Machine Learning**: Familiarity with algorithms to gain insights.
  + **Communication**: Strong storytelling skills to present findings effectively.
* **Excitement and Engagement**: Companies look for individuals who are excited about their industry, as this leads to higher productivity and engagement.

## **Glossary**

**Careers and Recruiting in Data Science Lesson Glossary**

Welcome! This alphabetized glossary contains many of the terms in this course. These terms are important for you to recognize when working in the industry, participating in user groups, and participating in other certificate programs.

|  |  |  |
| --- | --- | --- |
| **Term** | **Definition** | **Video where the term is introduced** |
| Adobe Spark | A suite of software tools that allow users to create and share visual content such as graphics, web pages, and videos. | Recruiting for Data Science |
| Analytical skills | The ability to analyze information systematically, logically, and organized. | Recruiting for Data Science |
| Chief information officer (CIO) | A business executive is responsible for an organization’s information technology systems and tech-related initiatives. | How Can Someone Become a Data Scientist |
| Computational thinking | Breaking problems into smaller parts and using algorithms, logic, and abstraction to develop solutions. Often used but not limited to computer science. | How Can Someone Become a Data Scientist |
| Data clusters | A group of similar, related data points distinct from other clusters. | How Can Someone Become a Data Scientist |
| Executive summary | Usually occurring at the beginning of a research paper, this section summarizes the important parts of the paper, including its key findings. | The Report Structure |
| High-performing computing (HPC) cluster | A computing technology that uses a system of networked computers designed to solve complex and computationally intensive problems in traditional environments. | How Can Someone Become a Data Scientist |
| Mathematical computing | The use of computers to calculate, simulate, and model mathematical problems. | Importance of Mathematics and Statistics for Data Science |
| Matrices | Plural for matric, matrices are a rectangular (tabular) array of numbers often used in mathematics, statistics, and computer science. | Recruiting for Data Science |
| Stata | A software package used for statistical analysis. | Recruiting for Data Science |
| Statistical distributions | A way of describing the likelihood of different outcomes based on a dataset. The “bell curve” is a common statistical distribution. | How Can Someone Become a Data Scientist |
| Structured Query Language (SQL) | A language used for managing data in a relational database. | Importance of Mathematics and Statistics for Data Science |
| TCP/IP network | A network that uses the TCP/IP protocol to communicate between connected devices on that network. The Internet uses TCP/IP. | How Can Someone Become a Data Scientist |

## **Summary: Careers and Recruiting in Data Science**

Congratulations! You have completed this module. At this point, you know that:

* Data Science helps physicians provide the best treatment for their patients  , helps meteorologists predict the extent of local weather events, and can even help predict natural disasters like earthquakes and tornadoes.
* Companies can start on their data science journey by capturing data. Once they have data, they can begin analyzing it.
* Everyone who uses the Internet generates mass amounts of data daily.
* Amazon and Netflix use recommendation engines, and UPS uses data from customers, drivers, and vehicles to use  the drivers’ time and fuel efficiently.
* The purpose of the final deliverable of a Data Science project is to communicate new information and insights from the data analysis to key decision-makers.
* The report should present a thorough analysis of the data and communicate the project findings.
* Companies should look for someone excited about working with the data in their particular industry. They should seek out someone curious who can ask interesting, meaningful questions about the types of data they intend to collect. They should hire people who love working with data, are fluent in statistics, and are competent in applying machine learning algorithms.
* A clearly organized and logical report should communicate the following to the reader:
  + What they gain by reading the report
  + Clearly defined goals
  + The significance of your contribution
  + Appropriate context by giving sufficient background
  + Why this work is practical and useful
  + Conjecture plausible future developments that might result from your work

**Data Science Application Domains**

## **Case Study**

**Case Study: Lila's Journey to Becoming a Data Scientist: Her Working Approach on the First Task**

This case study explores the data scientist's career path and key attributes, highlighting the skills, education, and experiences required to excel in this dynamic field. We'll follow the story of Lila, a fictional individual who aspires to become a successful data scientist.

There will be a quiz after this reading based on the contents of this case study.

**Education and Skill Acquisition**

With an economics undergraduate degree and a substantial data analysis background, Lila finds data science and its potential to drive meaningful change captivating. Inspired by her experiences, she makes a determined decision to transition her career and step into the role of a data scientist.

Lila realizes that to embark on her data science journey, she needs to enhance her skills and knowledge. She enrolled in the IBM Data Science Professional Certificate online program that covers key topics like statistics, machine learning, data analysis, and programming languages like Python and SQL. She diligently completes coursework and practices her coding skills on real datasets.

**Building a Strong Foundation**

As she progresses in her studies, Lila gains a deep understanding of data science fundamentals such as data manipulation and visualization with Python libraries like NumPy, Pandas, and Matplotlib. This strong foundation equips her with essential skills for data analysis.

**Visualization for Storytelling**

Lila knows she must communicate her findings effectively, so she learns which types of data visualizations will be most informative. She learns to create charts and graphs that visually represent data like sales trends, customer segmentation, and product popularity, allowing stakeholders to grasp the data's significance. These visualizations help in storytelling and decision-making.

**Hands-On Experience**

Lila understands that practical experience is invaluable in data science. She started participating in Kaggle competitions and working on personal data projects. These experiences expose her to real-world data problems and help her develop problem-solving skills. Furthermore, she created her GitHub account and uploaded her projects to build her profile.

**Data Wrangling and Preprocessing**

Lila learns that data scientists spend a significant portion of their time on data cleaning and preprocessing. She worked on various datasets, learned data preprocessing as she used sed NumPy and pandas Python libraries, and became skilled in handling missing data, outlier detection, and feature engineering to improve model performance.

**Communication and Storytelling**

Recognizing that data scientists must communicate their findings effectively, Lila honed her data storytelling skills. She learned various tools like matplotlib and plotly while she pursued her IBM Data Science Professional Certificate. She learned how to create compelling visualizations and present her insights in a clear and understandable manner.

**Networking and Collaboration**

Lila actively participates in data science communities and attends meetups and conferences. She collaborates on open-source projects, connects with fellow data scientists, and gains exposure to various industries when she attended the IBM TechXchange Conference.

**Domain Expertise**

Understanding that domain knowledge is crucial, Lila chooses a niche that aligns with her interests. She looks deeply into several domains, including e-commerce, healthcare, finance, and several other fields to which she could apply her data science skills effectively. Since her master's in economics, she chose e-commerce as her core domain to land herself a data science career.

**Landing the First Job**

After months of preparation, Lila started applying for data scientist positions. She tailors her resume to highlight her relevant skills and projects. Her online portfolio showcases her capabilities and demonstrates her commitment to the field.

**Lila's Approach to Working on Her First Task as a Data Scientist**

As a newly hired junior data scientist at a retail company, Lila uses data insights to improve customer service. Her first assignment involves diving into customer data to identify patterns and anomalies that could impact customer service. She uses data analysis to enhance the overall customer experience.

**Dataset Selection and Sourcing**

In the initial phase of her data science journey, Lila faced the challenge of selecting a suitable dataset and procuring it from different sources. Apart from the historical data available for the organizations for the past four years, she scoured various repositories, websites, and databases to find the right datasets for her project. Upon collecting data from diverse sources, Lila encountered another crucial decision point. She had to decide how to harmonize and integrate these disparate datasets into a cohesive whole. She reached out to product professionals, data engineers, and domain specialists, seeking their input and expertise in merging datasets.

**Data Understanding and Cleaning**

Lila begins by importing the dataset into her data analysis environment using Python and SQL. She loads the data and examines the first few rows to understand its structure and contents. Upon acquiring the dataset, Lila encounters her first challenge: data cleaning. Lila checks for missing values, duplicates, and outliers in the dataset. She addresses missing data by imputing or removing rows or columns with missing values. Outliers are identified and treated appropriately based on their impact on the analysis.

**Exploratory Data Analysis (EDA)**

As she delves into exploratory data analysis, Lila faces numerous choices. She must determine which summary statistics, visualizations, and distribution analyses will best reveal insights into customer behavior and sales trends. Each choice she makes during EDA influences the story the data will tell. Lila conducts EDA to gain insights into the dataset. She generates summary statistics and visualizations (histograms, scatter plots) and explores the distribution of variables. EDA helps her understand customer behavior, popular products, and sales trends.

**Feature Engineering**

Lila recognizes the potential for feature engineering to enhance her analysis. She assesses whether creating new features, such as calculating total purchase amounts, will improve the dataset's utility for her project.

**Statistical Analysis, Machine Learning**

Lila evaluates whether statistical tests or machine learning algorithms are necessary. She employs regression analysis to understand relationships between variables and explore machine learning models for demand forecasting or customer segmentation tasks. Lila also performs statistical tests to uncover patterns in the data. She uses regression analysis to understand the impact of unit price on sales.

**Presentation and Reporting**

At the culmination of her analysis, Lila faces the challenge of presenting her findings. Lila compiles her analysis and findings using a Jupyter Notebook into a comprehensive report and presentation. She highlights actionable insights and recommendations for the e-commerce platform's stakeholders.

**Continuous Learning**

After completing her first project, Lila continues to refine her skills, explores more complex datasets, and tackles increasingly challenging data science tasks.

**Machine Learning Skills**

Although Lila took an introductory course on Machine Learning as part of the IBM Data Science Professional Certificate, the field intrigues her, and she wants to develop her skills further by taking the IBM Machine Learning Professional Certificate. She identified Machine Learning Repository datasets in the course and experimented with various algorithms. Lila dives into machine learning to excel as a data scientist, wherein she studies various algorithms, such as linear regression, decision trees, and deep learning models. She continues to gain expertise in selecting and fine-tuning algorithms based on specific data problems.

## **Explore Data Science Job Listings**

**Review and evaluate a data science job post.**

**Assignment overview**

For this project, you should find a data science job posting on a job board of your choice, such as LinkedIn, Indeed, Zip Recruiter, Glassdoor, Monster, Naukri, or USAjobs.gov, that interests you.

Analyze the posting by responding to the following questions and statements. You do not need to submit your responses. This is an exercise to familiarize yourself with actual data science related jobs.

Identify the following aspects of data science job post:

1. What is the company name that is advertising the job?
2. What is the job title?
3. Where is the role located?
4. What is the expected salary or salary range?
5. What is the total number of results from the search for the job post?
6. What is one technical responsibility from the job post related to something you learned about in this course?
7. What are two required technical skills from the job post?
8. What are at least two ideas or concepts you learned about in this course relevant to these job posts?

# **Course Summary**

Congratulations! You have completed this course. At this point, you know that:

* Data science is the practice of extracting valuable insights from vast datasets to guide strategic decision-making.
* Data science careers offer diverse paths, often involving mathematics, programming, and a curiosity for data exploration.
* Successful data scientists exhibit qualities like curiosity, critical judgment, and an aptitude for constructive argumentation.
* The data science field is characterized by high demand, resulting in attractive remuneration for skilled professionals.
* A Data Scientist's daily routine can vary significantly depending on the project's nature.
* A wide array of algorithms is available for extracting insights from data.
* Big Data plays a pivotal role in driving digital transformation across industries.
* Cloud computing is a fundamental technology in modern data science.
* Data mining techniques are essential for uncovering patterns and knowledge from data.
* Tools like Hadoop, HDFS, Hive, and Spark are employed for processing Big Data.
* Deep learning, machine learning, and regression are critical data science topics.
* Data science applications span diverse domains, solving complex problems.
* Companies can harness data science to address age-old challenges with innovative solutions.
* Data science contributes significantly to saving lives and improving various aspects of society.
* Careers in data science offer exciting opportunities, with mathematics and statistics being essential foundations.
* Reports in data science adhere to specific structures, and career roadmaps provide guidance.
* Case studies and projects offered practical application of the knowledge acquired during the course.