Week 1- Reading 1

Hello and welcome! The program you are about to explore is specifically designed to help every type of learner successfully finish the certificate and become an entry-level junior or associate data analyst. No previous data analytics, mathematics, or statistical experience is required. To succeed, you just need to be open to learning how data influences the world.

Become job-ready

Every day, the amount of data out there gets bigger and bigger. So the ability to interpret it effectively is more important than ever before. Data analytics is becoming one of the fastest-growing and most rewarding career choices in the world. In the next decade, the demand for business analytics skills will probably be higher than the demand for any other career (10.9% vs. 5.2%) (Source: Bureau of Labor Statistics). All kinds of companies all over the world need qualified data analysts to solve problems and help them make the best possible business decisions. And right now, fifty-nine percent of companies have plans to add even more positions requiring data analysis skills (Source: SHRM). By the time you are done with this program, you will be well-prepared to make smart, strategic, data-driven recommendations for organizations in all kinds of industries.

During each course of the program, you will complete lots of hands-on assignments and projects based on both day-to-day life and the practical activities of a data analyst. Along the way, you will learn how to ask the right questions and understand objectives. You will also learn how to effectively clean and organize large amounts of data to make it ready for high-quality analysis. On top of that, you will get hands-on experience using all kinds of tools and techniques that will help you recognize patterns and uncover relationships between data points. And to help you communicate the results of your analysis, you will learn how to design visuals and dashboards. There is even an opportunity to create a case study, which you can highlight in your resume to show what you have learned to potential employers.

Course overview

The entire program has eight courses. This is the first course and it covers about five weeks of material.

1. **Foundations: Data, Data, Everywhere** (this course)
2. [Ask Questions to Make Data-Driven Decisions](https://www.coursera.org/learn/ask-questions-make-decisions/home/welcome)
3. [Prepare Data for Exploration](https://www.coursera.org/learn/data-preparation/home/welcome)
4. [Process Data from Dirty to Clean](https://www.coursera.org/learn/process-data/home/welcome)
5. [Analyze Data to Answer Questions](https://www.coursera.org/learn/analyze-data/home/welcome)
6. [S​hare Data Through the Art of Visualization](https://www.coursera.org/learn/visualize-data/home/welcome)
7. [Data Analysis with R Programming](https://coursera.org/learn/data-analysis-r/home/welcome)
8. [Google D​ata Analytics Capstone: Complete a Case Study](https://coursera.org/learn/google-data-analytics-capstone/home/welcome)

## C​ourse content

C​ourse 1– Foundations: Data, Data, Everywhere

1. **I​ntroducing data analytics:** Data helps us make decisions, in everyday life and in business. In this first part of the course, you will learn how data analysts use tools of their trade to inform those decisions. You will also get to know more about this course and the overall program expectations.
2. **T​hinking analytically:** Data analysts balance many different roles in their work. In this part of the course, you will learn about some of these roles and the key skills that are required. You will also explore analytical thinking and how it relates to data-driven decision making.
3. **E​xploring the wonderful world of data:** Data has its own life cycle, and data analysts use an analysis process that cuts across and leverages this life cycle. In this part of the course, you will learn about the data life cycle and data analysis process. They are both relevant to your work in this program and on the job as a future data analyst. You will be introduced to applications that help guide data through the data analysis process.
4. **S​etting up a data toolbox:** Spreadsheets, query languages, and data visualization tools are all a big part of a data analyst’s job. In this part of the course, you will learn the basic concepts to use them for data analysis. You will understand how they work through examples provided.
5. **D​iscovering data career possibilities:** All kinds of businesses value the work that data analysts do. In this part of the course, you will examine different types of businesses and the jobs and tasks that analysts do for them. You will also learn how a Google Data Analytics Certificate will help you meet many of the requirements for a position with these organizations.
6. **C​ompleting the Course Challenge:** At the end of this course, you will be able to put everything you have learned into perspective with the Course Challenge. The Course Challenge will ask you questions about the main concepts you have learned and then give you an opportunity to apply those concepts in two scenarios.

## W​hat to expect

Each week of the course includes a series of lessons with many types of learning opportunities. These include:

* **V​ideos** for instructors toteach new concepts and demonstrate the use of tools
* **Readings** to introduce new ideas and build on the concepts from the videos
* [**Discussion forums**](https://www.coursera.org/learn/foundations-data/discussions) to share, explore, and reinforce lesson topics for better understanding
* **D​iscussion prompts** to promote thinking and engagement in the discussion forums
* **Practice quizzes** to prepare you for graded quizzes
* **Graded quizzes** to measure your progress and give you valuable feedback
* Also, be sure to pay attention to the **in-video questions** that will pop up from time to time. They are designed for you to check your learning.

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, they are just a guide. Feel free to move 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. If you prefer, you can extend your deadlines by returning to **Overview** in the navigation panel and clicking **Switch Sessions**. Assessments are based on the approach taken by the course to offer a wide variety of learning materials and activities that reinforce important skills. Graded and ungraded quizzes will help the content sink in. Ungraded practice quizzes are a chance for you to prepare for the graded quizzes. Both types of quizzes can be taken more than one time.

## Optional speed track for those experienced in data analytics

The Google Data Analytics Certificate provides instruction and feedback for learners hoping to earn a position as an entry-level data analyst. While many learners will be brand new to the world of data analytics, others may be familiar with the field and simply wanting to brush up on certain skills.

If you believe this course will be primarily a refresher for you, we recommend taking the practice diagnostic quiz (you can find it in this week's content). It will enable you to determine if you should follow the speed track, which is an opportunity to proceed to Course 2 after having taken each of the Course 1 Weekly Challenges and the overall Course Challenge. Learners who score 100% on the diagnostic quiz can treat Course 1 videos, readings, and activities as optional. Learners following the speed track are still able to earn the certificate.

## Tips

* It is strongly recommended to take these courses—and go through the items in each lesson—in the order they appear because new information and concepts build on previous knowledge.
* Use the additional resources that are linked throughout the program. They are designed to support your learning.
* W​hen you encounter useful links in the course, remember to bookmark them so you can refer to the information for study or review.
* Additional resources are free, but some sites place limits on how many articles can be accessed for free each month. Sometimes you can register on the site for full access, but you can always bookmark a resource and come back to view it later.
* If something is confusing, don’t hesitate to rewatch a video, go through a reading again, and so on.
* Take part in all learning opportunities to gain as much knowledge and experience possible.

Congratulations on choosing to take this first step toward becoming part of the wonderful world of data analytics. Enjoy the journey!

**1. Ask**

**---**First up, the analysts in our example needed to define what the project would look like and what would qualify as a successful result. So, to determine these things, they **asked** effective questions and collaborated with leaders and managers who were interested in the outcome of their people analysis.

**2. Prepare**

**---**It all started with solid **preparation**. The group built a timeline of three months and decided how they wanted to relay their progress to interested parties. Also during this step, the analysts identified what data they needed to achieve the successful result they identified in the previous step - in this case, the analysts chose to gather the data from a survey of new employees. They identified specific questions to ask about employee satisfaction with different business processes, such as hiring, onboarding, and compensation. Rules were established for who would have access to the data collected, what specific information would be gathered, and how best to present the data visually. The analysts brainstormed possible project- and data-related issues and how to avoid them.

**3. Process**

---The group sent the survey out. Great analysts know how to respect both their data and the people who provide it. Since employees provided the data, it was important to make sure all employees gave their consent to participate. The data analysts also made sure employees understood how their data would be **collected, stored, managed, and protected**. In order to maintain confidentiality and protect and store the data effectively, access was restricted to a limited number of analysts. Collecting and using data ethically is one of the responsibilities of a data analyst. Then the data was cleaned up to make sure it was complete, correct, and relevant, and uploaded to an internal data warehouse for an additional layer of security.

**4. Analyze**

**---**Then, the analysts did what they do best: analyze! From the completed surveys, the data analysts would **discover** that a new employee’s experience with the hiring process was a key indicator of overall job satisfaction. The analysts found that employees who experienced an efficient and transparent hiring process were most likely to remain with the company. Employees who experienced a long and complicated hiring process were most likely to leave the company. The group knew it was important to **document** exactly what they found in the analysis, no matter what the results. To do otherwise would decrease trust in the survey process and reduce their ability to collect truthful data from employees in the future.

**5.Share**

**---**Just as they made sure the data was carefully protected, the analysts were also careful **sharing the report**. For example, in order for a manager to receive the survey report, a minimum number of their team members had to have participated in the survey. The group presented the results to leaders first to make sure they had the full picture, then asked them to deliver the results to their teams. This gave leaders an opportunity to **communicate the results** with the right context and have productive team conversations about next steps.

**6.Act**

**---**The last stage of the process for the team of analysts was to work with leaders within their company and decide how best to **implement changes and take actions** based on the findings. The analysts recommended standardizing the hiring process for all new hires based on the most efficient and transparent hiring practices. A year later, the same survey was distributed to employees. Analysts anticipated that a comparison between the two sets of results would indicate that the action plan worked. Turns out, the changes improved the retention rate for new employees and the actions taken by leaders were successful!

In a nutshell

The analysts **asked** questions to define both the issue to be solved and what would equal a successful result. Next, they **prepared** by building a timeline and collecting data with employee surveys that were designed to be inclusive. They **processed** the data by cleaning it to make sure it was complete, correct, relevant, and free of errors and outliers. They **analyzed** the clean employee survey data. Then the analysts **shared** their findings and recommendations with team leaders. Afterward, leadership **acted** on the results and focused on improving key areas.

1. **Ask** questions and define the problem.
2. **Prepare** data by collecting and storing the information.
3. **Process** data by cleaning and checking the information.
4. **Analyze** data to find patterns, relationships, and trends.
5. **Share** data with your audience.
6. **Act** on the data and use the analysis results.

Data + business knowledge = mystery solved

Blending facts and data with your business knowledge will be a common part of your process. The key is figuring out the exact mix of data and business knowledge for each particular project. A lot of times it will depend on the goals of your analysis. That is why analysts often ask, “How do I define success for this project?”

Successful analysis needs to be accurate, and fast enough to help decision-makers. So try asking yourself these questions about a project:

* What kind of results are needed?
* Who will be informed?
* Am I answering the question being asked?
* How quickly does a decision need to be made?

For example, if you are working on a rush project, you might need to rely on your own knowledge and experience more than usual. There just isn’t enough time to thoroughly analyze all of the available data. But if you get a project that involves plenty of time and resources, then the best strategy would be to be more data-driven. It’s up to you, the data analyst, to think about the situation and make the best possible choice. You will probably blend facts and knowledge a million different ways over the course of your data analytics career. And the more you practice, the better you will get at finding that perfect blend.

**---Origin of the data analysis process**

When you decided to join this program, you proved that you are a curious person. So let’s tap into your curiosity and talk about the origins of data analysis. We don’t fully know when or why the first person decided to record data about people and things. But we do know it was useful because the idea is still around today!



We also know that data analysis is rooted in statistics, which has a pretty long history itself. Archaeologists mark the start of statistics in ancient Egypt with the building of the pyramids. The Ancient Egyptians were masters of organizing data. They documented their calculations and theories on papyri (paper-like materials), which are now viewed as the earliest examples of spreadsheets and checklists. Today’s data analysts owe a lot to those brilliant scribes, who helped create a more technical and efficient process.

It is time to enter the **data analysis life cycle**—the process of going from data to decision. Data goes through several phases as it gets created, consumed, tested, processed, and reused. With a life cycle model, all key team members can drive success by planning work both up front and at the end of the data analysis process. While the data analysis life cycle is well known among experts, there isn't a single defined structure of those phases. There might not be one single architecture that’s uniformly followed by every data analysis expert, but there are some shared fundamentals in every data analysis process. This reading provides an overview of several, starting with the process that forms the foundation of the Google Data Analytics Certificate.

The process presented as part of the Google Data Analytics Certificate is one that will be valuable to you as you keep moving forward in your career:

1. **Ask**: Business Challenge/Objective/Question
2. **Prepare**: Data generation, collection, storage, and data management
3. **Process**: Data cleaning/data integrity
4. **Analyze**: Data exploration, visualization, and analysis
5. **Share**: Communicating and interpreting results
6. **Act**:  Putting your insights to work to solve the problem

Understanding this process—and all of the iterations that helped make it popular—will be a big part of guiding your own analysis and your work in this program. Let’s go over a few other variations of the data analysis life cycle.

EMC's data analysis life cycle

EMC Corporation's data analytics life cycle is cyclical with six steps:

1. Discovery
2. Pre-processing data
3. Model planning
4. Model building
5. Communicate results
6. Operationalize

EMC Corporation is now Dell EMC. This model, created by David Dietrich, reflects the cyclical nature of real-world projects. The phases aren’t static milestones; each step connects and leads to the next, and eventually repeats. Key questions help analysts test whether they have accomplished enough to move forward and ensure that teams have spent enough time on each of the phases and don’t start modeling before the data is ready. It is a little different from the data analysis life cycle this program is based on, but it has some core ideas in common: the first phase is interested in discovering and asking questions; data has to be prepared before it can be analyzed and used; and then findings should be shared and acted on.

For more information, refer to [The Genesis of EMC's Data Analytics Lifecycle](https://infocus.delltechnologies.com/david_dietrich/the-genesis-of-emcs-data-analytics-lifecycle/).

SAS' iterative life cycle

An iterative life cycle was created by a company called **SAS**, a leading data analytics solutions provider. It can be used to produce repeatable, reliable, and predictive results:

1. Ask
2. Prepare
3. Explore
4. Model
5. Implement
6. Act
7. Evaluate

The SAS model emphasizes the cyclical nature of their model by visualizing it as an infinity symbol. Their life cycle has seven steps, many of which we have seen in the other models, like Ask, Prepare, Model, and Act. But this life cycle is also a little different; it includes a step after the act phase designed to help analysts evaluate their solutions and potentially return to the ask phase again.

For more information, refer to [Managing the Analytics LIfe Cycle for Decisions at Scale](https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/manage-analytical-life-cycle-continuous-innovation-106179.pdf).

Project-based data analytics life cycle

A project-based data analytics life cycle has five simple steps:

1. Identifying the problem
2. Designing data requirements
3. Pre-processing data
4. Data analysis
5. Data visualizing

This data analytics project life cycle was developed by Vignesh Prajapati. It doesn’t include the sixth phase, or what we have been referring to as the Act phase. However, it still covers a lot of the same steps as the life cycles we have already described. It begins with identifying the problem, preparing and processing data before analysis, and ends with data visualization.

For more information, refer to [Understanding the data analytics project life cycle](http://pingax.com/understanding-data-analytics-project-life-cycle/).

Big data analytics life cycle

Authors Thomas Erl, Wajid Khattak, and Paul Buhler proposed a big data analytics life cycle in their book, **Big Data Fundamentals: Concepts, Drivers & Techniques**. Their life cycle suggests phases divided into nine steps:

1. Business case evaluation
2. Data identification
3. Data acquisition and filtering
4. Data extraction
5. Data validation and cleaning
6. Data aggregation and representation
7. Data analysis
8. Data visualization
9. Utilization of analysis results

This life cycle appears to have three or four more steps than the previous life cycle models. But in reality, they have just broken down what we have been referring to as Prepare and Process into smaller steps. It emphasizes the individual tasks required for gathering, preparing, and cleaning data before the analysis phase.

For more information, refer to [Big Data Adoption and Planning Considerations](https://www.informit.com/articles/article.aspx?p=2473128&seqNum=11&ranMID=24808).

Data life cycle based on research

One final data life cycle informed by Harvard University research has eight phases:

1. Generation
2. Collection
3. Processing
4. Storage
5. Management
6. Analysis
7. Visualization
8. Interpretation

This version includes storage, management, and interpretation phases, and excludes the Act phase that has appeared in other models.

For more information, refer to [8 Steps in the Data Life Cycle](https://online.hbs.edu/blog/post/data-life-cycle).

Key takeaway

From our journey to the pyramids and data in Ancient Egypt to now, the way we analyze data has evolved (and continues to do so). The data analysis process is like real life architecture, there are different ways to do things but the same core ideas still appear in each model of the process. Whether you use the structure of this Google Data Analytics Certificate or one of the many other iterations you have learned about, we are here to help guide you as you continue on your data journey.