Course 1– Foundations: Data, Data, Everywhere

1. **Introducing 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. **Thinking 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. **Exploring 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. **Setting 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. **Discovering 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. **Completing 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.

Week – 1 Introducing Data Analytics

Data

Data is a collection of facts that can be used to draw conclusions, make predictions, and assist in decision-making.

Data Analysis

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

The six steps of the data analysis process are(data analysis life cycle):

- **Ask** questions and define the problem.
- **Prepare** data by collecting and storing the information.

- **Process** data by cleaning and checking the information. Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
- **Analyze** data to find patterns, relationships, and trends.
- **Share** data with your audience.
- **Act** on the data and use the analysis results.

People Analytics

Human resources analytics/Workforce analytics. It is the practice of collecting and analyzing data on the people who make up a company's workforce in order to gain insights to improve how the company operates.

Data science, the discipline of making data useful, is an umbrella term that encompasses three disciplines: ***machine learning, statistics, and analytics***. These are separated by how many decisions you know you want to make before you begin with them.

If you want to make a few important decisions under uncertainty, that is statistics. The excellence of statistics is rigor. They are very, very careful about protecting decision-makers from coming to the wrong conclusion.

If you want to automate, in other words, make many, many, many decisions under uncertainty, that is machine learning and AI. Performance is the excellence of the machine learning and AI engineer.

But what if you don't know how many decisions you want to make before you begin? What if what you're looking for is inspiration? You want to encounter your unknown unknowns. You want to understand your world. That is analytics. The excellence of an analyst is speed. How quickly can you surf through vast amounts of data to explore it and discover the gems, the beautiful potential insights that are worth knowing about and bringing to your decision-makers

---

Data Ecosystem

An ecosystem is a group of elements that interact with one another.

Data ecosystems are made up of various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data. These elements include hardware and software tools, and the people who use them.

For example, you could tap into your retail store's database, which is an ecosystem filled with customer names, addresses, previous purchases, and customer reviews.

Data Science

Data science is defined as creating new ways of modeling and understanding the unknown by using raw data.

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

Data Analytics

Data analytics in the simplest terms is the science of data.

It's a very broad concept that encompasses everything from the job of managing and using data to the tools and methods that data workers use each and every day.

Data, data analysis and the data ecosystem, all fit under the data analytics umbrella.

Data-driven Decision-making

Using facts to guide business strategy.

The first step in data-driven decision-making is figuring out the business need.

Data alone will never be as powerful as data combined with human experience, observation, and sometimes even intuition. To get the most out of data-driven decision-making, it's important to include insights from people who are familiar with

the business problem. These people are called **subject matter experts**, and they have the ability to look at the results of data analysis and identify any inconsistencies, make sense of gray areas, and eventually validate choices being made.

Gut Instinct

Gut instinct is an intuitive understanding of something with little or no explanation. This isn't always something conscious; we often pick up on signals without even realizing. You just have a "feeling" it's right.

Blending data with business knowledge, plus maybe a touch of gut instinct, will be a common part of your process as a junior data analyst. The key is figuring out the exact mix for each particular project. A lot of times, it will depend on the goals of your analysis.

At the heart of data-driven decision-making is data, so analysts are most effective when they ensure that facts are driving strategy.

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.

Glossary

Week 2 – All about analytical thinking

Analytical Skills

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

Five essential aspects to analytical skills:

1. Curiosity
    - It is all about wanting to learn something.
2. Understanding context
    - Context is the condition in which something exists or happens. This can be a structure or an environment.
    - Grouping things into categories.
    - Understanding where information fits into the "big picture".
3. Having technical mindset
    - The ability to break things down into smaller steps or pieces and work with them in an orderly and logical way.
4. Data design
    - Data design is how you organize information.
5. Data strategy
    - Data strategy is the management of the people, processes, and tools used in data analysis.

Data-driven decision-making involves curiosity, understanding context, having a technical mindset, data design, and data strategy.

Analytical Thinking

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

Five key aspects to analytical thinking:

1. Visualization
    - The graphical representation of information.
    - To understand and explain information more effectively.
2. Strategy
    - Strategizing helps data analysts see what they want to achieve with the data and how they can get there. Strategy also helps improve the quality and

usefulness of the data we collect. By strategizing, we know all our data is valuable and can help us accomplish our goals.

3. Problem-orientation
    - Data analysts use a problem- oriented approach in order to identify, describe, and solve problems. It's all about keeping the problem top of mind throughout the entire project.
4. Correlation
    - Being able to identify a relationship between two or more pieces of data.
    - Correlation does not equal causation. In other words, just because two pieces of data are both trending in the same direction, that doesn't necessarily mean they are all related.
5. Big-picture and detail-oriented thinking.
    - Being able to see the big picture as well as the details.
    - Detail-oriented thinking is all about figuring out all of the aspects that will help you execute a plan.

The more ways you can think, the easier it is to think outside the box and come up with fresh ideas.

You need to think **critically** to find out the right questions to ask.

You also need to think **creatively** to get new and unexpected answers.

Root cause (What is the root cause of a problem?)

A root cause is the reason why a problem occurs. If we can identify and get rid of a root cause, we can prevent that problem from happening again.

5 whys

In the Five Whys you ask "why" five times to reveal the root cause. The fifth and final answer should give you some useful and sometimes surprising insights.

The Five Whys process is used to reveal a root cause of a problem through the answer to the fifth question.

Gap analysis (Where are the gaps in our process?)

Gap analysis is a method for examining and evaluating how a process works currently in order to get where you want to be in the future. Improving accessibility, increasing efficiency, and reducing carbon emissions are examples of improvements that gap analysis can help accomplish.

The general approach to gap analysis is understanding where you are now compared to where you want to be. Then you can identify the gaps that exist between the current and future state and determine how to bridge them.

(What did we not consider before?)

This is a great way to think about what information or procedure might be missing from a process, so you can identify ways to make better decisions and strategies moving forward.

Quartile

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

Week3-The wonderful world of data

Data life cycle

The entire period of time that data exists in the system.

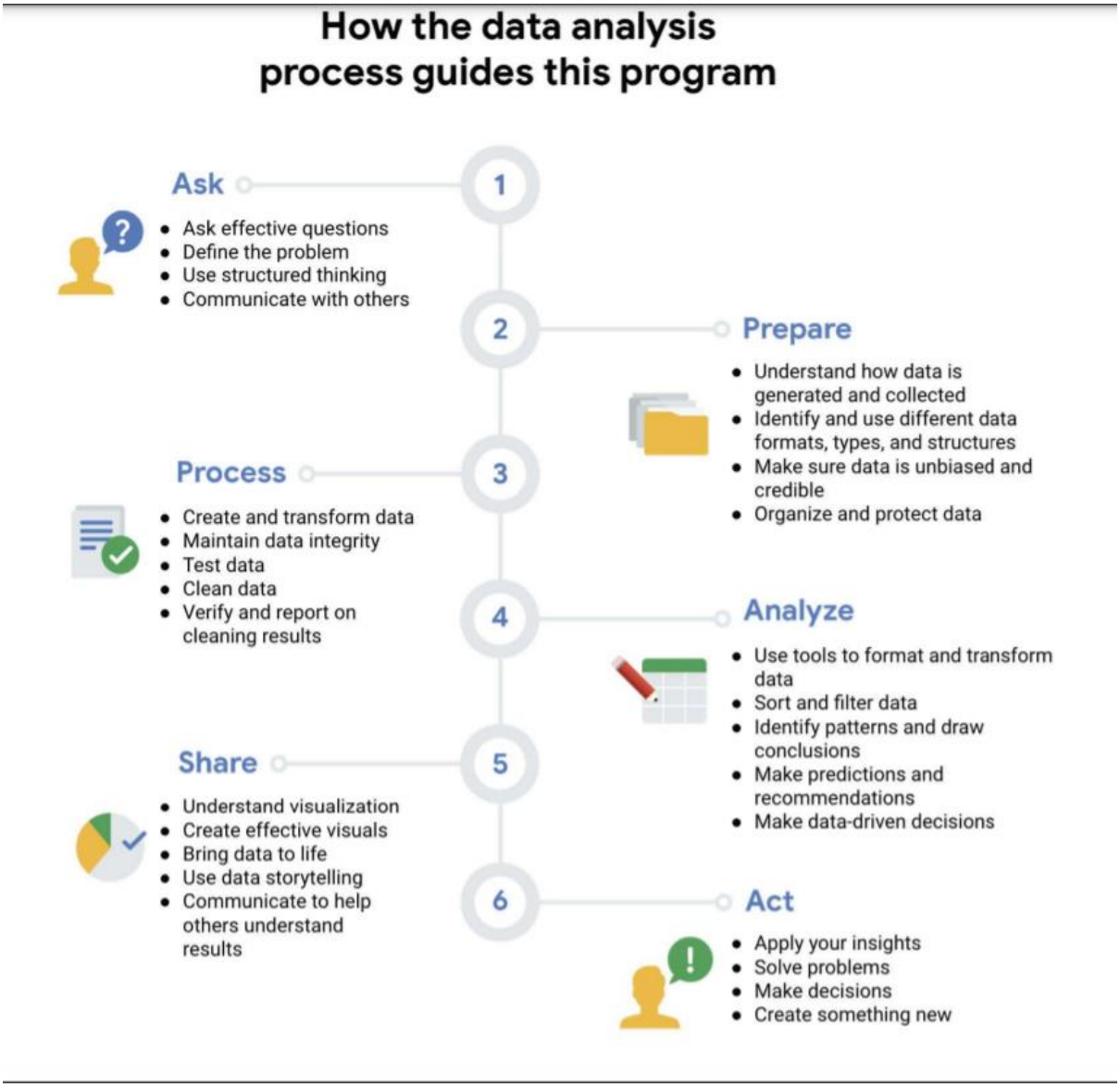
Six stages of data life cycle:

1. **Plan**: During planning, a business decides what kind of data it needs, how it will be managed throughout its life cycle, who will be responsible for it, and the optimal outcomes.
2. **Capture**: This is where data is collected from a variety of different sources and brought into the organization.
3. **Manage**: How we care for our data, how and where it's stored, the tools used to keep it safe and secure, and the actions taken to make sure that it's maintained properly.
4. **Analyze**: In this phase, the data is used to solve problems, make great decisions, and support business goals. A data analyst might use formulas to perform calculations, create a report from the data, or use spreadsheets to aggregate data.
5. **Archive**: Keep relevant data stored for long-term and future reference.
6. **Destroy**: Remove data from storage and delete any shared copies of the data. This is important for protecting a company's private information, as well as private data about its customers.

Individual stages in the data life cycle will vary from company to company or by industry or sector. Although data life cycles vary, one data management principle is universal. Govern how data is handled so that it is accurate, secure, and available to meet your organization's needs.

Stakeholders

They are people who have invested time and resources into a project and are interested in the outcome.



Data analyst tools

Spreadsheet

A spreadsheet is a digital worksheet. It stores, organizes, and sorts data. This is important because the usefulness of your data depends on how well it's structured. When you put your data into a spreadsheet, you can see patterns, group information and easily find the information you need. Some popular spreadsheets are ***Microsoft Excel*** and ***Google Sheets***. Spreadsheets also have some really useful features called **formulas** and **functions**.

- A **formula** is a set of instructions that performs a specific calculation using the data in a spreadsheet.
- A **function** is a preset command that automatically performs a specific process or task using the data in a spreadsheet.

Spreadsheets structure data in a meaningful way by letting you:

- Collect, store, organize, and sort information
- Identify patterns and piece the data together in a way that works for each specific data project
- Create excellent data visualizations, like graphs and charts.

Query language

A query language is a computer programming language that allows you to retrieve and manipulate data from a database.

SQL is a language that lets data analysts communicate with a database.
A **database** is a collection of data stored in a computer system. With SQL, data analysts can access the data they need by making a query. Some popular Structured Query Language (SQL) programs include ***MySQL***, ***Microsoft SQL Server***, and ***BigQuery***.

Query languages:

- Allow analysts to isolate specific information from a database(s)
- Make it easier for you to learn and understand the requests made to databases
- Allow analysts to select, create, add, or download data from a database for analysis

Data visualization

Data visualization is the graphical representation of information. Some examples include graphs, maps, and tables. They help data analysts communicate their insights to others, in an effective and compelling way. Some popular visualization tools are ***Tableau*** and ***Looker***.

These tools:

- Turn complex numbers into a story that people can understand
- Help stakeholders come up with conclusions that lead to informed decisions and effective business strategies
- Have multiple features:
    - Tableau's simple drag-and-drop feature lets users create interactive graphs in dashboards and worksheets.
    - Looker communicates directly with a database, allowing you to connect your data right to the visual tool you choose.

A career as a data analyst also involves using programming languages, like R and Python, which are used a lot for statistical analysis, visualization, and other data analysis.

Spreadsheets	Databases
Software applications	Data stores - accessed using a query language (e.g. SQL)
Structure data in a row and column format	Structure data using rules and relationships
Organize information in cells	Organize information in complex collections
Provide access to a limited amount of data	Provide access to huge amounts of data
Manual data entry	Strict and consistent data entry
Generally one user at a time	Multiple users
Controlled by the user	Controlled by a database management system

- What is the relationship between the data life cycle and the data analysis process? How are the two processes similar? How are they different?
    - The data life cycle involves stages for identifying needs and managing data. Data analysis involves process steps to make meaning from data.
    - While the data analysis process will drive your projects and help you reach your business goals, you must understand the life cycle of your data in order to use that process. To analyze your data well, you need to have a thorough understanding of it. Similarly, you can collect all the data you want, but the data is only useful to you if you have a plan for analyzing it.
- What is the relationship between the Ask phase of the data analysis process and the Plan phase of the data life cycle? How are they similar? How are they different?
    - The Plan and Ask phases both involve planning and asking questions, but they tackle different subjects. The Ask phase in the data analysis process focuses on big-picture strategic thinking about business goals. However, the Plan phase focuses on the fundamentals of the project, such as what data you have access to, what data you need, and where you're going to get it.

Week4-Set up your toolbox

Spreadsheet basics

A spreadsheet helps you structure data in rows and columns, prepare data for analysis, and create custom data visualizations. The chart editor enables data analysts to choose the type of chart you're making and customize its appearance. To better analyze your data, clean up

your chart to make it more visually appealing and to clarify what data means by making your chart more descriptive. To do that, it's important to add chart titles and axis titles.

Data Visualization

Data analysts use data visualizations to explain complex data quickly, reinforce data analysis, and create interesting graphs and charts. Data visualizations can clearly demonstrate patterns and trends, help stakeholders understand complex data more quickly, and illustrate relationships between data points.

Misc: A line chart is effective for tracking trends over time. A pie chart shows how a whole is broken down into parts.

Steps to plan a data visualization:

i. **Explore the data for patterns**
ii. **Plan your visuals**: Refine the data and present the results of your analysis. You will want to create a data visualization that explains your findings quickly and effectively to your target audience.
iii. **Create your visuals**: Now that you have decided what kind of information and insights you want to display, it is time to start creating the actual visualizations. Keep in mind that creating the right visualization for a presentation or to share with stakeholders is a process. It involves trying different visualization formats and making adjustments until you get what you are looking for. In this case, a mix of different visuals will best communicate your findings and turn your analysis into the most compelling story for stakeholders.

Data visualization toolkit

- **Spreadsheets**: Spreadsheets are great for creating simple visualizations like bar graphs and pie charts, and even provide some advanced visualizations like maps, and waterfall and funnel diagrams.
- **Visualization software (Tableau)**: Tableau is a popular data visualization tool that lets you pull data from nearly any system and turn it into compelling visuals or actionable insights. The platform offers built-in visual best practices, which makes analyzing and sharing data fast, easy, and (most importantly) useful. Tableau works well with a wide variety of data and includes an interactive dashboard that lets you and your stakeholders click to explore the data interactively. Start exploring Tableau from the How-to Video resources.
- **Programming language (R with RStudio)**: As with Tableau, you can create dashboard-style data visualizations using RStudio. Resources: RStudio, RStudio Cheatsheets, RStudio Visualize Data Primer.

Week5-Endless career possibilities

Data analytics helps businesses make better decisions, but getting there is a process. It begins with analyzing a business problem, identifying data about that problem, and then using data analysis to arrive at an answer. Sometimes you get an answer that solves your business problem, but it's often just as likely that you discover other questions to investigate further.

Businesses using data analytics have a common theme. They all have issues to explore, questions to answer, or problems to solve.

- An **issue** is a topic or subject to investigate.
- A **question** is designed to discover information.
- A **problem** is an obstacle or complication that needs to be worked out.

These questions and problems become the foundation for all kinds of business tasks, that you'll help solve as a data analyst. A **business task** is the question or problem data analysis answers for a business. Data analytics helps businesses make better decisions. It all starts with a business task and the question it's trying to answer.

Data analysts have an important responsibility: making sure that their analyses are fair (ensuring that analysis does not create or reinforce bias requires using processes and systems that are fair and inclusive to everyone). **Fairness** means ensuring that your analysis doesn't create or reinforce bias. It's important to think about fairness from the moment you start collecting data for a business task to the time you present your conclusions to your stakeholders. Considering inclusive sample populations, social context, and self-reported data enable fairness in data collection.

A case study:

1. To improve the effectiveness of its teaching staff, the administration of a high school offered the opportunity for all teachers to participate in a workshop. They were not required to attend; instead, the administration encouraged teachers to sign up. Of the 43 teachers on staff, 19 chose to take the workshop.
   At the end of the academic year, the administration collected data on teacher performance for all teachers on staff. The data was collected via student survey. In the survey, students were asked to rank each teacher's effectiveness on a scale of 1 (very poor) to 6 (very good).
   The administration compared data on teachers who attended the workshop to data on teachers who did not. The comparison revealed that teachers who attended the workshop had an average score of 4.95, while teachers who did not attend had an average score of 4.22. The administration concluded that the workshop was a success.

- This is an example of unfair practice. It is tempting to conclude—as the administration did—that the workshop was a success. However, since the workshop was voluntary and not random, it is not appropriate to infer a causal relationship between attending the workshop and the higher rating.
  The workshop might have been effective, but other explanations for the differences in the ratings cannot be ruled out. For example, another explanation could be that the staff volunteering for the workshop were the better, more motivated teachers. This group of teachers would be rated higher whether or not the workshop was effective. It's also notable that there is no direct connection between student survey responses and workshop attendance. The data analyst could correct this by asking for the teachers to be selected randomly to participate in the workshop. They could also collect data that measures something more directly related to workshop attendance, such as the success of a technique the teachers learned in that workshop.

Data analyst roles

The data analyst role is one of many job titles that contain the word "analyst."

To name a few others that sound similar but may not be the same role:

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

Data analysts, data scientists, and data specialists sound very similar but focus on different tasks.

Decoding the job description

	Data Analysts	Data Scientists	Data Specialists
Problem solving	Use existing tools and methods to solve problems with existing types of data	Invent new tools and models, ask open-ended questions, and collect new types of data	Use in-depth knowledge of databases as a tool to solve problems and manage data
Analysis	Analyze collected data to help stakeholders make better decisions	Analyze and interpret complex data to make business predictions	Organize large volumes of data for use in data analytics or business operations
Other relevant skills	• Database queries • Data visualization • Dashboards • Reports • Spreadsheets	• Advanced statistics • Machine learning • Deep learning • Data optimization • Programming	• Data manipulation • Information security • Data models • Scalability of data • Disaster recovery

Job Specializations

The role of data specialist (concentrates on in-depth knowledge of databases) is one of many specializations within data analytics. Other specialist roles for data analysts can focus on in-depth knowledge of specific industries.

Other industry-specific specialist positions that you might come across in your data analyst job search include:

- Marketing analyst — analyzes market conditions to assess the potential sales of products and services
- HR/payroll analyst — analyzes payroll data for inefficiencies and errors
- Financial analyst — analyzes financial status by collecting, monitoring, and reviewing data
- Risk analyst — analyzes financial documents, economic conditions, and client data to help companies determine the level of risk involved in making a particular business decision
- Healthcare analyst — analyzes medical data to improve the business aspect of hospitals and medical facilities