Week I

Welcome to the Google Analytics Certificate

Data ✧.*
Data is what companies in e-commerce, entertainment, healthcare, manufacturing, marketing, finance, tech, and hundreds of other industries all have in common.

Organizations of all kinds need **data analysts** to help them improve their processes, identify opportunities and trends, launch new products, provide great customer service, and make thoughtful decisions**.**

Data Analysis ✧.*
Data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making. (ex. reading reviews of a product before deciding whether to buy it or not.

Data evolves over time which means this analysis or analytics, as we call it, can give us new information throughout data's entire life cycle. **Data** is everywhere.

Data Analyst ✧.*
Data analyst is someone who collects, transforms, and organizes data in order to help make informed decisions.

The **Google Data Analytics Certificate** is split into courses based on different processes for data analysis. Those are *ask, prepare, process, analyze, share,* and *act.*

Introduction to Data Analytics

“Data! Data! Data!... I can’t make bricks without clay!” - Sherlock Holmes by Sir Arthur Conan Doyle

Data Analytics ✧.*
The science of data.

Data ✧.*

Data is a collection of facts or information.

Analysis ✧.*
Through analysis, you'll learn how to use the data to draw conclusions, and make predictions, and decisions.

Computers + Brain + Skills Traits = Job Success

Data Analytics in Everyday Life

Controlling data can ✧.*
Improve processes, identify opportunities, launch new products, serve customers, and make thoughtful decisions.

Analysis ✧.*
Turning data into insights.

Data Analytics ✧.*
Can help organizations completely rethink something they do or point them into totally new decisions.

People Analytics ✧.*
Also known as 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.

Case Study: New Data Perspectives

Six phases of the data analysis process: ✧.*
Ask - ask effective questions
Prepare - came up with solid preparation. Decide on how to relay on progress.
Process - data is collected, stored, managed, and protected.
Analyze - discovering the findings and its result.
Share - sharing and communicating the result
Act - implement changes and take actions.

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.

Dimensions of Data Analytics

Decision Intelligence ✧.*
It is a combination of applied data science and the social and managerial sciences. It is all about harnessing the power and beauty of data.

"Data analyst is an explorer, a detective, and an artist all rolled into one."

Data Science ✧.*
It is an umbrella term that encompasses three disciplines: machine learning, statistics, and analytics.

- ***Machine learning*** *- automate, or make many many many decisions under uncertainty*
- ***Statistics*** *- make few important decisions under uncertainty*
- ***Analytics*** *- where an individual dont know how many decisions she/he wants to make before they begin. Looking for inspiration, encountering unknown unknowns.*

What is the Data Ecosystem?

Ecosystem ✧.*
a group of elements that interact with one another. It can be large or small.

Data Ecosystem ✧.*

The various elements that interact with one another in order to produce, manage, store, organie, analyze, and share data. These elements include hardware, software tools, and end-user.

Cloud ✧.*
a place to keep data online rather than a computer hard drive. A virtual location.

Data science ✧.*
Creating new ways of modeling and understanding the unknown by using raw data.

The difference between data science and data analyst ✧.*
Data scientists create new questions using data, while analysts find answers to existing questions by creating insights from data sources.

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

How data informs better decisions

Data-driven decision-making ✧.*
Using facts to guide business strategy. First step is figuring out the business needs. Problem that needs to be solved.

Data analyst finds data, analyzes it, and uses it to uncover trends, patterns, and relationships.

Subject matter experts ✧.*
People who are familiar with the business problem. 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.

Data and Gut Instinct

Analysts use **data-driven decision-making** and follow a **step-by-step process**. You have learned that there are six steps to this process:

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.

Gut instinct ✧.*
An intuitive understanding of something with little or no explanation. Feeling of something right.

"Data + business knowledge = mystery solved"

Questions to find the perfect balance in data:

- **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?**

Origins of the data analysis process

Key takeaway notes:

Statistics - where data analysis is rooted
Archaeologists - mark the start of statistics in ancient Egypt with the building of the pyraminds.
Ancient Egyptians - masters of organizing data by documenting calculations to papyri.
Papyri - paper-like materials that is now viewed as an earliest example of spreadsheets and checklists.

Data analysis life cycle ✧.*
Process of going from data to decision.

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

EMC's data analysis life cycle ✧.*
By David Dietrich. EMC Corporation is now Dell EMC.

1. Discovery
2. Pre-processing data
3. Model planning

4. Model building
5. Communicate results
6. Operationalize

SAS's iterative life cycle ✧.*

SAS is a leading data analytics solution provider. It can be used to produce repeatable, reliable, and predictive results. Visualizes as an infinity symbol wherein after you evaluate, you potentially return to the ask phase again.

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

Big data analytics life cycle ✧.*

Thomas Erl, Wajid Khattak,, and Paul Buuhler proposed this in their book, "Big Data Fundamentals: Concepts, Drivers & Techniques"

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

Week II

Key Data Analyst Skills

Analytical Skills ✧.*

Qualities and characteristics associated with solving problems using facts

Aspects of Analytical Skills ✧.*

1. **Curiosity** - all about wanting to learn something.

- **Curious people** seek out new challenges and experiences that lead to knowledge.

2. **Understanding context -** how you group things into categories.
    - Pay close attention. Listening and trying to understand the full picture is critical. Know what's context included and out of context.
    - **Context** is the condition in which something exists or happens.
3. **Having a technical mindset** - involves breaking processes down into smaller steps and working with them in an orderly, logical way
    - **Technical mindset** is the ability to break things down into smaller steps or pieces and work with them in an orderly and logical way.
    - using technical mindset means taking something that seems like a single task, and breaking it into smaller steps with an orderly process
    - Example: Paying bills [*sorting* by the date due, *add* them up and *compare* the amount in the balance of your bank account, and *pay* them]
4. **Data design** - how you organize information.
    - As a data analyst, design typically has to do with an actual database.
5. **Data strategy** - management of the people, processes, and tools used in data analysis.
    - You **manage people by making** sure they know how to use the right data to find solutions to the problem you're working on.
    - **Processes** are about making sure the path to that solution is clear and accessible.
    - **Tools** are making sure the right technology is being used for the job.

All About Thinking Analytically

Analytical Thinking ✧.*
Identifying and defining a problem and then solving it by using data in an organized, step-by-step manner.

Five key aspect to analytical thinking ✧.*

1. **Visualization** - the graphical representation of information. important because visuals can help data analysts understand and explain information more effectively.
    - Examples: graphs, maps, or other design elements.
2. **Strategy** - key to staying focused and on track. 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.
3. **Problem-orientation -** identify, describe, and solve problems. It's all about keeping the problem top of mind throughout the entire project.

4. **Correlation -** like a relationship. You can find all kinds of correlations in 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**
    - **Big-picture thinking** - This means being able to see the big picture as well as the details. Big-picture thinking is like looking at a complete puzzle. It helps you zoom out and see possibilities and opportunities. This leads to exciting new ideas or innovations.
    - **Detail-oriented thinking** - all about figuring out all of the aspects that will help you execute a plan. In other words, the pieces that make up your puzzle.

Exploring Core Analytical Skills

Questions Data Analysts Ask ✧.*

1. *What is the root of the problem?*
    - **Root cause** - reason why problems occur. (ask five whys)
2. *Where are the gaps in our process?*
    - **Gap analysis** - lets you examine and evaluate how a process works currently in order to get where you want to be in the future.

- 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.

3. *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.

Using Data to Drive Successful Outcomes

Data-driven decision making ✧.*

With data, they can gain valuable insights, verify their theories or assumptions, better understand opportunities and challenges, support an objective, help make a plan, and much more.

Real-world data magic

Quartile ✧.*
Divides data points into four equal parts or quarters

Nonprofits ✧.*
Organizations dedicated to advancing a social cause or advocating for a particular effort such as food security, education, or the arts.

Week III

Data Phases and Tools

Data analysis tool ✧.*
These include spreadsheets, databases, query languages, and visualization software.

Stages of the Data Life Cycle

Data life cycle ✧.*
This includes plan, capture, manage, archive, and destroy.

1. **Planning** - happens well before starting an analysis project. 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. **Capturing** - where data is collected from a variety of different sources and brought into the organization.

Database ✧.*
A collection of data stored in a computer system.

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. This phase is very important to data cleansing.
4. **Analyze** - This is where data analysts really shine. In this phase, the data is used to solve problems, make great decisions, and support business goals.
5. **Archive** - means storing data in a place where it's still available, but may not be used again.

6. **Destroy** - This is important for protecting a company's private information, as well as private data about its customers.

Variations of the Data Life Cycle

1. **Plan:** Decide what kind of data is needed, how it will be managed, and who will be responsible for it.
2. **Capture:** Collect or bring in data from a variety of different sources.
3. **Manage:** Care for and maintain the data. This includes determining how and where it is stored and the tools used to do so.
4. **Analyze:** Use the data to solve problems, make decisions, and support business goals.
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.

U.S. Fish and Wildlife Service

The U.S. Fish and Wildlife Service uses the following data life cycle:

1. Plan
2. Acquire
3. Maintain
4. Access
5. Evaluate
6. Archive

The U.S. Geological Survey (USGS)

The USGS uses the data life cycle below:

1. Plan
2. Acquire
3. Process
4. Analyze
5. Preserve
6. Publish/Share

Several cross-cutting or overarching activities are also performed during each stage of their life cycle:

- Describe (metadata and documentation)
- Manage Quality
- Backup and Secure

Financial institutions

1. Capture
2. Qualify
3. Transform
4. Utilize
5. Report
6. Archive
7. Purge

Harvard Business School (HBS)

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

Six Phases of Data Analysis

1. **Ask Phase ✧.***

- In this phase, we do two things. We define the problem to be solved and we make sure that we fully understand stakeholder expectations.

Stakeholders ✧.*

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

- **Defining a problem** - means you look at the current state and identify how it's different from the ideal state. Usually there's an obstacle we need to get rid of or something wrong that needs to be fixed.
- **Understanding stakeholder expectations** - determine who the stakeholders are. What they all have in common is that they help make decisions, influence actions and strategies, and have specific goals they want to meet. They also care about the project.

2. **Prepare Phase ✧.***

- This is where data analysts collect and store data they'll use for the upcoming analysis process. Any decisions made from your analysis should always be based on facts and be fair and impartial.

3. **Process Phase ✧.***
- Data analysts find and eliminate any errors and inaccuracies that can get in the way of results. This usually means cleaning data, transforming it into a more useful format, combining two or more datasets to make information more complete and removing outliers, which are any data points that could skew the information. This phase is all about getting the details right.

4. **Analyze Phase ✧.***
- Analyzing the data you've collected involves using tools to transform and organize that information so that you can draw useful conclusions, make predictions, and drive informed decision-making.

5. **Share Phase ✧.***
- how data analysts interpret results and share them with others to help stakeholders make effective data-driven decisions. This phase, visualization is a data analyst's best friend.

R ✧.*
is a popular tool for data manipulation, calculation, and visualization.

6. **Act Phase ✧.***
- This is the exciting moment when the business takes all of the insights you, the data analyst, have provided and puts them to work in order to solve the original business problem.

The Data Analysis Process and This Program

Example of Data Process

Ask ✧.*

What is the problem we're trying to solve? What is the purpose of this analysis? What are we hoping to learn from it?

Prepare ✧.*

think about what type of data we need to answer those key questions. This could be anything from quantitative data or qualitative data. Think about how we're going to collect that data or if we need to collect that data.

Process ✧.*

It begins with cleaning. This is where you get a chance to understand its structure, its quirks, its nuances, and you really get a chance to understand deeply what type of data you're going to be working with and understand what potential that data has to answer all of your questions. This is running all of our quality assurance checks.

Analyze ✧.*
This is the point where we analyze our data, making sure to do so in as objective and unbiased a way as possible. To do this, the first thing we do is run through a series of analyses that we've already planned ahead of time based on the questions that we know we want to answer from the very, very beginning of the process.

Share ✧.*
share all of the data and insights that you've generated from your analyses.

Act ✧.*
"All of this work from asking the right questions to collecting your data, to analyzing and sharing, doesn't mean much of anything if we aren't taking action on what we've just learned. " This is where we use all of those data-driven insights to decide what types of interventions we want to introduce, not only at the organizational level, but also at the team level as well.

Data Analyst Tools

Spreadsheets ✧.*
Two popular options are Microsoft Excel and Google Sheets. It 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. Spreadsheets also have some really useful features called formulas and functions.

- **Formulas** - a set of instructions that performs a specific calculation using the data in a spreadsheet. Formulas can do basic things like add, subtract, multiply and divide.
- **Function** - a preset command that automatically performs a specific process or task using the data in a spreadsheet.

Spreadsheets help 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 languages for databases ✧.*

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

- **SQL** - Structured Query Language or Sequel. a language that lets data analysts communicate with a database. You can ask it to do a lot of different things such as insert, delete, select or update data. Some popular Structured Query Language (SQL) programs include MySQL, Microsoft SQL Server, and BigQuery.

Query languages help you:

- 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 ✧.*

*Includes graphs, maps, and tables. Some popular visualization tools are Tableau and Looker. Data analysts like using **Tableau** because it helps them create visuals that are very easy to understand. **Looker** is also popular with data analysts because it gives them an easy way to create visuals based on the results of a query.*

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.

The Right Tool for the Job

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

Week IV

Columns and Rows and Cells, Oh my!

	A	B	C	D
1			This	
2	This is a cell.		is	
3	This	is	a	row.
4			column.	
5				

Columns ✧.*
Organized vertically in a spreadsheet and are ordered by letter.

Rows ✧.*
Organized horizontally in a spreadsheet and are ordered by numbers. In a dataset, it is also called observation.

Cells ✧.*
Combination of column letter and row number.

Attribute ✧.*
a characteristic or quality of data used to label a column in a table. More commonly, attributes are referred to as column names, column labels, headers, or the header row.

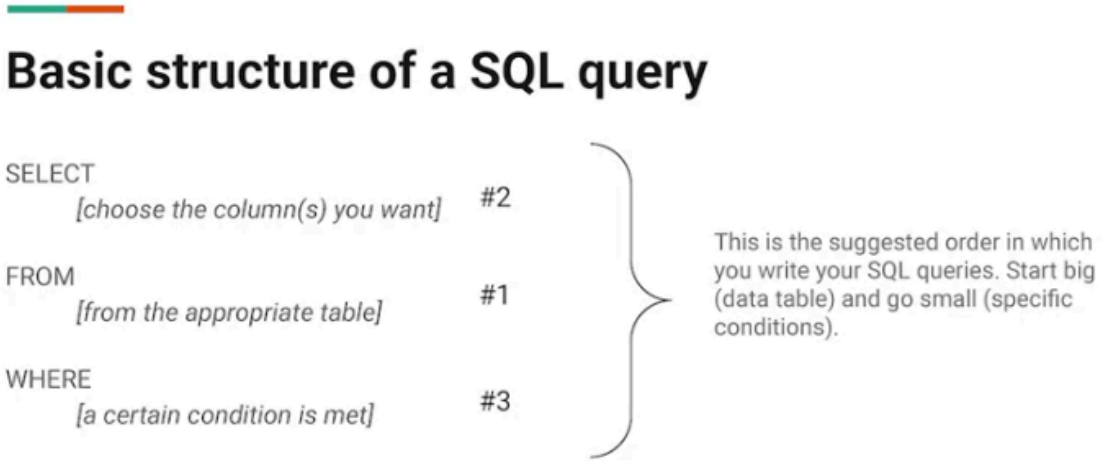
Observation ✧.*
includes all of the attributes for something contained in a row of a data table.

SQL in Action

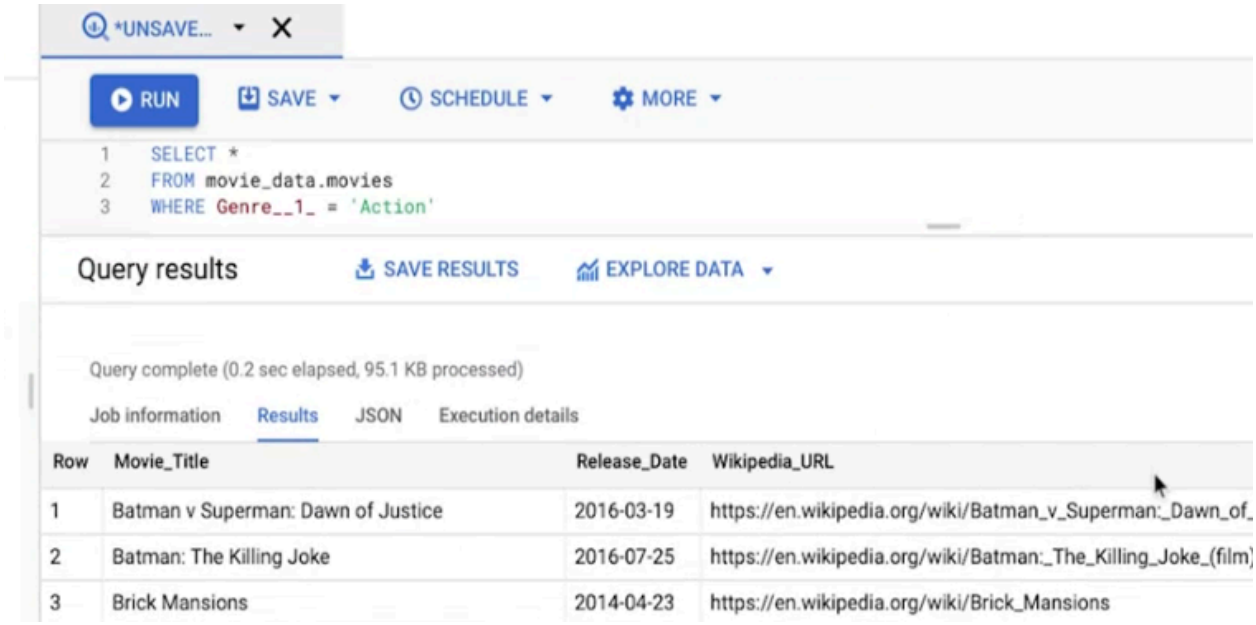
SQL ✧.*
can do lots of the same things with data that spreadsheets can do. You can use it to store, organize and analyze your data, among other things. Supersized spreadsheets. No matter which database you use, SQL basically works for the same in each. Queries are universal.

Examples of databases that use SQL: Oracle, MySQL, PostgreSQL, Microsoft SQL Server, etc.

Query ✧.*
A request for data or information from a database.



Basic Query in Sequel ✧.*



SQL Guide

Syntax ✧.*
It is the predetermined structure of a language that includes all required words, symbols, and punctuation, as well as their proper placement.

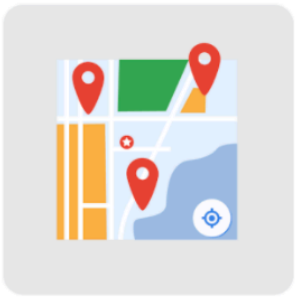
The syntax of every SQL query is the same:
Use **SELECT** to choose the columns you want to return.
Use **FROM** to choose the tables where the columns you want are located.
Use **WHERE** to filter for certain information.

SQL Query ✧.*
is like filling in a template.

Planning a data visualization

Steps to plan a data visualization
Step 1: Explore the data for patterns
Step 2: Plan your visuals
Step 3: Create your visuals

Line charts can track sales over time



Maps can connect sales to locations

Donut charts can show customer segments

Bar charts can compare total visitors and visitors that make a purchase

Spreadsheets (Microsoft Excel or Google Sheets) ✧.*

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.

Programming language (R with RStudio) ✧.*

A lot of data analysts work with a programming language called R. Most people who work with R end up also using RStudio, an integrated developer environment (IDE), for their data visualization needs. As with Tableau, you can create dashboard-style data visualizations using RStudio.

Week V

The Power of Data in Business

Issue ✧.*
A topic or subject to investigate

Question ✧.*
Designed to discover information

Problem ✧.*
An obstacle or complication that need to be worked out

Business task ✧.*
The question or problem data analysis answers for a business

Understanding Data and Fairness

Fairness ✧.*
Ensuring that your analysis doesn't create or reinforce bias.

Data Analysts in Different Industries

Important factors to think about when searching for dream job ✧.*
Industry, tools, location, travel and culture.

Think about what you're interested in ✧.*
The key is to think about your interests early in your job search. That'll lead you in the right direction, and it will help you in interviews too.

Decisions where you want to live ✧.*
Does your preferred industry have opportunities in your area? Are you trying to stay local or would you be happy relocating? How long are you willing to commute to work every day? Will you drive to work, walk, take public transport? Is that possible year-round? How do you feel about working remotely? Does working from home excite you or bore you?

Think about your values and what company culture is a good fit for you. ✧.*
Do you work best in a team or by yourself? Do you like to have a set routine or do you enjoy taking a new project and trying new things? Do your values match the company's values?

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

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 specialization by industry

- **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

Interview Best Practices

“Think about a time where you've used data to solve a problem, whether it's in your professional or personal projects.:

Increase professional work ✧.*

- One of them is to increase your online footprint, reach out to other analysts on LinkedIn, join local meet-ups with other data scientists. Sometimes when we're looking for a unique skill set, recruiters are going on websites like LinkedIn, and GitHub, and trying to find that talent themselves. It's really important to have your LinkedIn updated along with websites like GitHub, where you can showcase a lot of the data analyst projects you've done.
- Prepare questions for the interviewer. Make sure they're not broad questions. They should be questions that will help you understand the team and the job better. If you're given a case study in an interview, you should expect to be given a business problem along with the sample data set. Then you'd be asked to take that sample data set, analyze it, and come up with a solution. One of the things you can do to help prepare yourself for this is to ensure you are analyzing the data and coming up with a solution that relates back to that data. Sometimes there is no right answer, and a lot of times interviewers are looking to see your thought process and the way you get to your solution.
- Look for the recruiter. Look for the hiring manager online. See if you can reach out to them and set up a coffee chat or send them your resume directly. Online applications could be a really big black hole where you never hear back from the recruiter or the team. When you reach out directly to a hiring manager or recruiter, it really shows your eagerness for the role and your interests for the role.