Select another course

VIEW FAQS

1. Foundations
2. Ask
3. Prepare
4. Process
5. Analyze
6. Share
7. Act
8. Capstone

What you will learn:

- Real-life roles and responsibilities of a junior data analyst
- How businesses transform data into actionable insights
- Spreadsheet basics
- Database and query basics
- Data visualization basics

Skill sets you will build:

- Using data in everyday life
- Thinking analytically
- Applying tools from the data analytics toolkit
- Showing trends and patterns with data visualizations
- Ensuring your data analysis is fair

Select another course

VIEW FAQS

1. Foundations
2. Ask
3. Prepare
4. Process
5. Analyze
6. Share
7. Act
8. Capstone

What you will learn:

- How data analysts solve problems with data
- The use of analytics for making data-driven decisions
- Spreadsheet formulas and functions
- Dashboard basics, including an introduction to Tableau
- Data reporting basics

Skill sets you will build:

- Asking SMART and effective questions
- Structuring how you think
- Summarizing data
- Putting things into context
- Managing team and stakeholder expectations
- Problem-solving and conflict-resolution

Select another course

VIEW FAQS

1. Foundations
2. Ask
3. Prepare
4. Process
5. Analyze
6. Share
7. Act
8. Capstone

What you will learn:

- How data is generated
- Features of different data types, fields, and values
- Database structures
- The function of metadata in data analytics
- Structured Query Language (SQL) functions

Skill sets you will build:

- Ensuring ethical data analysis practices
- Addressing issues of bias and credibility
- Accessing databases and importing data
- Writing simple queries
- Organizing and protecting data
- Connecting with the data community (optional)

Select another course

VIEW FAQS

1. Foundations
2. Ask
3. Prepare
4. Process
5. Analyze
6. Share
7. Act
8. Capstone

What you will learn:

- Data integrity and the importance of clean data
- The tools and processes used by data analysts to clean data
- Data-cleaning verification and reports
- Statistics, hypothesis testing, and margin of error
- Resume building and interpretation of job postings (optional)

Skill sets you will build:

- Connecting business objectives to data analysis
- Identifying clean and dirty data
- Cleaning small datasets using spreadsheet tools
- Cleaning large datasets by writing SQL queries
- Documenting data-cleaning processes

Select another course

VIEW FAQS

1. Foundations
2. Ask
3. Prepare
4. Process
5. Analyze
6. Share
7. Act
8. Capstone

What you will learn:

- Steps data analysts take to organize data
- How to combine data from multiple sources
- Spreadsheet calculations and pivot tables
- SQL calculations
- Temporary tables
- Data validation

Skill sets you will build:

- Sorting data in spreadsheets and by writing SQL queries
- Filtering data in spreadsheets and by writing SQL queries
- Converting data
- Formatting data
- Substantiating data analysis processes
- Seeking feedback and support from others during data analysis

Select another course

VIEW FAQS

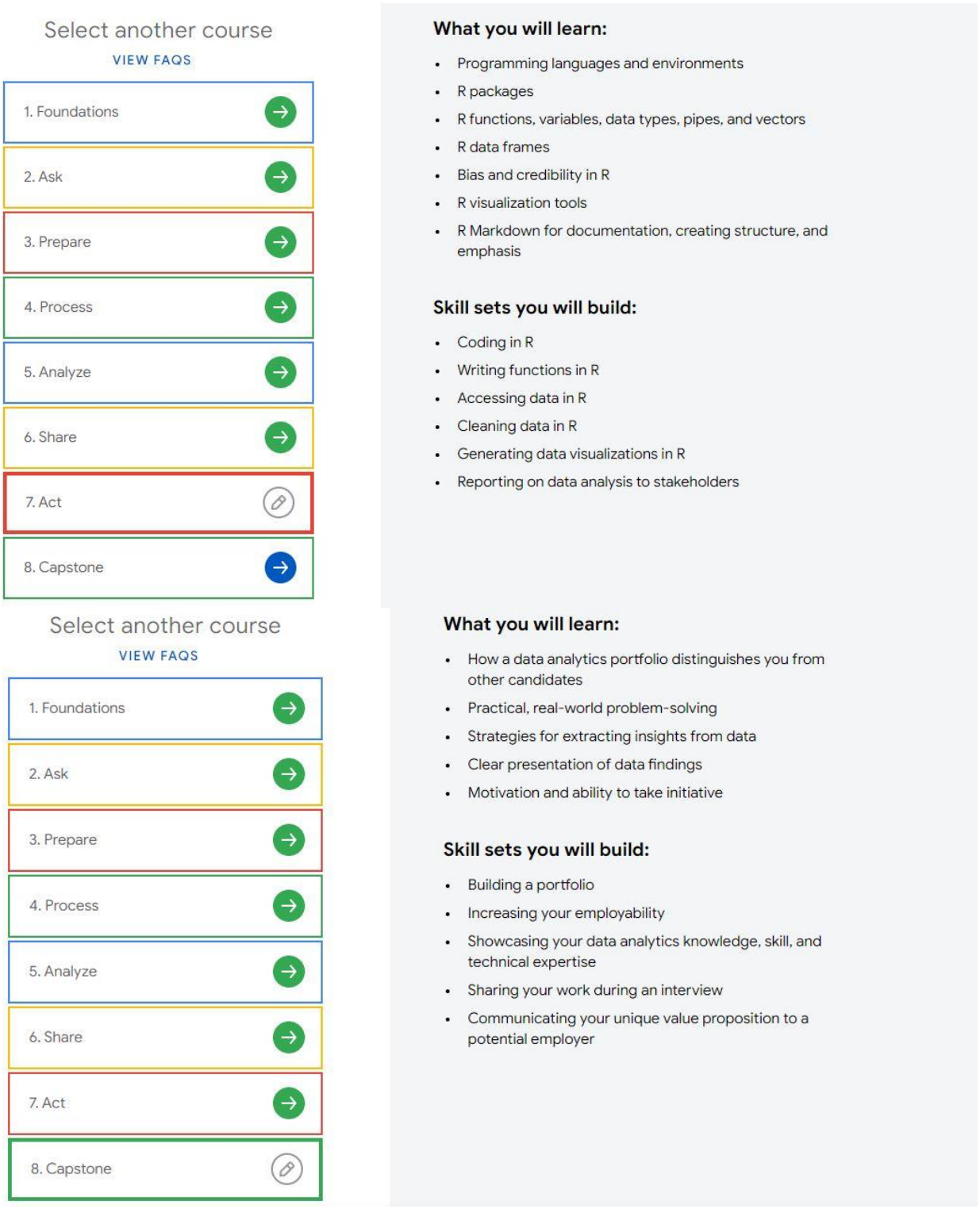
1. Foundations
2. Ask
3. Prepare
4. Process
5. Analyze
6. Share
7. Act
8. Capstone

What you will learn:

- Design thinking
- How data analysts use visualizations to communicate about data
- The benefits of Tableau for presenting data analysis findings
- Data-driven storytelling
- Dashboards and dashboard filters
- Strategies for creating an effective data presentation

Skill sets you will build:

- Creating visualizations and dashboards in Tableau
- Addressing accessibility issues when communicating about data
- Understanding the purpose of different business communication tools
- Telling a data-driven story
- Presenting to others about data
- Answering questions about data



Course 1 - Foundations: Data, Data, Everywhere

five essential aspects of analytical skills:

curiosity, understanding context, having a technical mindset, data design, and data strategy.

- **Curiosity:** a desire to know more about something, asking the right questions
- **Understanding context:** understanding where information fits into the "big picture"
- **Having a technical mindset:** breaking big things into smaller steps
- **Data design:** thinking about how to organize data and information
- **Data strategy:** thinking about the people, processes, and tools used in data analysis

Phases of data analysis

- **Ask:** Define the problem and confirm stakeholder expectations
- **Prepare:** Collect and store data for analysis
- **Process:** Clean and transform data to ensure integrity
- **Analyze:** Use data analysis tools to draw conclusions
- **Share:** Interpret and communicate results to others to make data-driven decisions
- **Act:** Put your insights to work in order to solve the original problem

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.

Course 2 - Ask Questions to Make Data-Driven Decisions

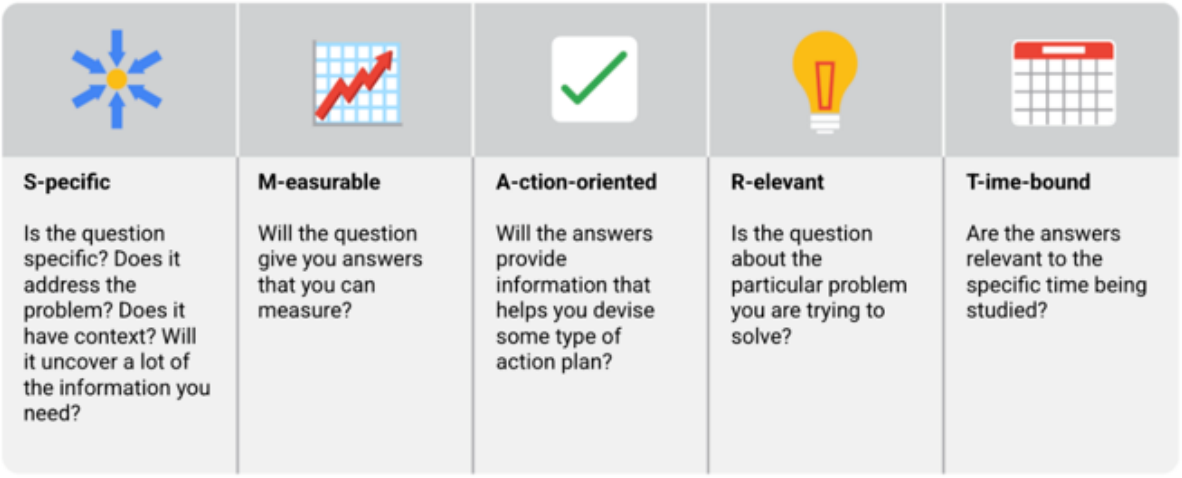
Data analysts typically work with six problem types

Data analysts typically work with six problem types

1. Making predictions	2. Categorizing things	3. Spotting something unusual
		
4. Identifying themes	**5. Discovering connections**	**6. Finding patterns**
		

Highly effective questions are SMART questions:

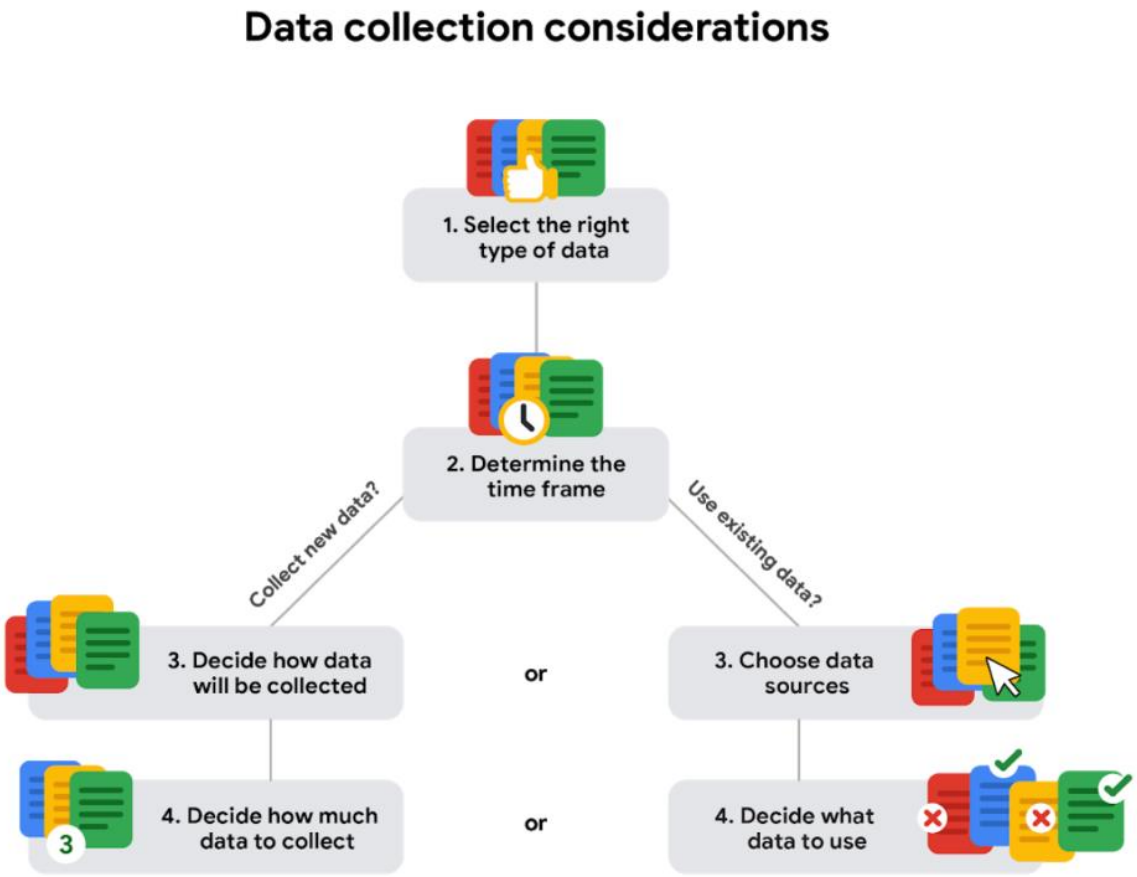
S M A R T



S-pecific	M-easurable	A-ction-oriented	R-elevant	T-ime-bound
Is the question specific? Does it address the problem? Does it have context? Will it uncover a lot of the information you need?	Will the question give you answers that you can measure?	Will the answers provide information that helps you devise some type of action plan?	Is the question about the particular problem you are trying to solve?	Are the answers relevant to the specific time being studied?

Course-3-Prepare Data for Exploration

Data Collection



BigQuery

Two account types: sanbox and free trail

- A **free sandbox account** doesn't ask for a method of payment. It does, however, limit you to 12 projects. It also doesn't allow you to insert new records to a database or update the field values of existing records. These data manipulation language (DML) operations aren't supported in the sandbox.
- A **free trial account** requires a method of payment to establish a billable account, but offers full functionality during the trial period.

https://www.coursera.org/learn/data-preparation/supplement/DYOQK/using-bigquery

sample size calculator

A **sample size calculator** tells you how many people you need to interview (or things you need to test) to get results that represent the target population. Let's review some terms you will come across when using a sample size calculator:

- **Confidence level**: The probability that your sample size accurately reflects the greater population.

- **Margin of error**: The maximum amount that the sample results are expected to differ from those of the actual population.
- **Population**: This is the total number you hope to pull your sample from.
- **Sample**: A part of a population that is representative of the population.
- **Estimated response rate**: If you are running a survey of individuals, this is the percentage of people you expect will complete your survey out of those who received the survey.

Margin of error in marketing

The margin of error is also important in marketing. Let's use A/B testing as an example. **A/B testing** (or split testing) tests two variations of the same web page to determine which page is more successful in attracting user traffic and generating revenue. User traffic that gets monetized is known as the **conversion rate**. A/B testing allows marketers to test emails, ads, and landing pages to find the data behind what is working and what isn't working. Marketers use the **confidence interval** (determined by the conversion rate and the margin of error) to understand the results.

For example, suppose you are conducting an A/B test to compare the effectiveness of two different email subject lines to entice people to open the email. You find that subject line A: "Special offer just for you" resulted in a 5% open rate compared to subject line B: "Don't miss this opportunity" at 3%.

Does that mean subject line A is better than subject line B? It depends on your margin of error. If the margin of error was 2%, then subject line A's actual open rate or confidence interval is somewhere between 3% and 7%. Since the lower end of the interval overlaps with subject line B's results at 3%, you can't conclude that there is a statistically significant difference between subject line A and B. Examining the margin of error is important when making conclusions based on your test results.

Course 4 - Process Data from Dirty to Clean

Validity

Definition

The concept of using data integrity principles to ensure measures conform to defined business rules or constraints

Example

Data collected five years ago used technology that is not approved or supported by the business

Accuracy

Definition

The degree of conformity of a measure to a standard or a true value

Example

Addresses in the business database are identified as incorrect when compared to the public postal service database

Completeness

Definition

The degree to which all required measures are known

Example

NULL/missing value for the item "Number of employees per store"

Consistency

Definition

The degree to which a set of measures is equivalent across systems

Example

Date of store opening stored in both MM/DD/YYYY and MM/YY formats

Course 5 - Analyze Data to Answer Questions

VLOOKUP syntax

A VLOOKUP function is available in both Microsoft Excel and Google Sheets. You will be introduced to the general syntax in Google Sheets. (You can refer to the resources at the end of this reading for more information about VLOOKUP in Microsoft Excel.)

```
VLOOKUP(10003, A2:B26, 2, FALSE)
```

Here is the syntax.

```
VLOOKUP(search_key, range, index, [is_sorted])
```

search_key

- The value to search for.
- For example, 42, "Cats", or I24.

range

- The range to consider for the search.
- The first column in the range is searched to locate data matching the value specified by search_key.

index

- The column index of the value to be returned, where the first column in range is numbered 1.
- If index is not between 1 and the number of columns in range, #VALUE! is returned.

is_sorted

- Indicates whether the column to be searched (the first column of the specified range) is sorted. TRUE by default.
- It's recommended to set is_sorted to FALSE. If set to FALSE, an exact match is returned. If there are multiple matching values, the content of the cell corresponding to the first value found is returned, and #N/A is returned if no such value is found.
- If is_sorted is TRUE or omitted, the nearest match (less than or equal to the search key) is returned. If all values in the search column are greater than the search key, #N/A is returned.

Course 6 - Share Data Through the Art of Visualization

Correlation and causation

Guidelines and pro tips

Refer to the following table for recommended guidelines and style checks for headlines, subtitles, labels, and annotations in your data visualizations. Think of these guidelines as guardrails. Sometimes data visualizations can become too crowded or busy. When this happens, the audience can get confused or distracted by elements that aren't really necessary. The guidelines

will help keep your data visualizations simple, and the style checks will help make your data visualizations more elegant.

Visualization components	Guidelines	Style checks
Headlines	- **Content**: Briefly describe the data - **Length**: Usually the width of the data frame - **Position**: Above the data	- Use brief language - Don't use all caps - Don't use italic - Don't use acronyms - Don't use abbreviations - Don't use humor or sarcasm
Subtitles	- **Content**: Clarify context for the data - **Length**: Same as or shorter than headline - **Position**: Directly below the headline	- Use smaller font size than headline - Don't use undefined words - Don't use all caps, bold, or italic - Don't use acronyms - Don't use abbreviations
Labels	- **Content**: Replace the need for legends - **Length**: Usually fewer than 30 characters - **Position**: Next to data or below or beside axes	- Use a few words only - Use thoughtful color-coding - Use callouts to point to the data - Don't use all caps, bold, or italic
Annotations	- **Content**: Draw attention to certain data - **Length**: Varies, limited by open space - **Position**: Immediately next to data annotated	- Don't use all caps, bold, or italic - Don't use rotated text - Don't distract viewers from the data

Interview Warmup is a tool that helps you practice answering questions to get more confident and comfortable with interviewing.

Big Interview, an online interview preparation platform, to create interactive interview tools specifically for data analytics learners like you.

We're excited to be able to offer you 12 months of free access (originally a $79/month value) as part of the Google Data Analytics Certificate! You've earned it.