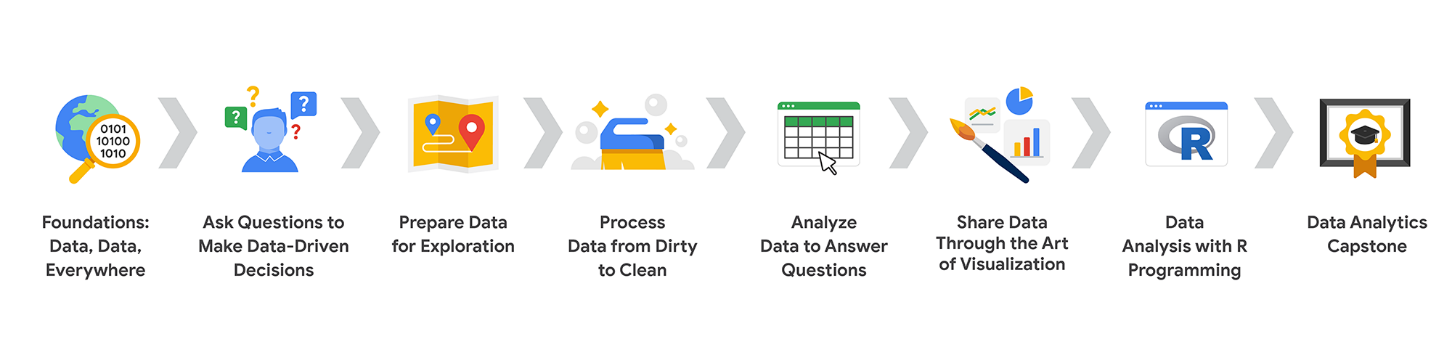
**Data analysis** is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

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



Foundations

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

Ask

**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

Prepare

**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)

Process

**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

Analyze

**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

Share

**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

Act

**Skill sets you will build:**

* Coding in R
* Writing functions in R
* Accessing data in R
* Cleaning data in R
* Generating data visualizations in R
* Reporting on data analysis to stakeholders

Capstone

**Skill sets you will build:**

* Building a portfolio
* Increasing your employability
* Showcasing your data analytics knowledge, skill, and technical expertise
* Sharing your work during an interview
* Communicating your unique value proposition to a potential employer

Businesses need a way to control all that data so

they can use it to **improve processes**, **identify opportunities** and **trends**,

**launch new products**, **serve customer**s, and **make thoughtful decisions**.

The six steps of the data analysis process that you have been learning in this program are: **ask, prepare, process, analyze, share,** and **act**.

Ask

First up, the analysts 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. These were the kinds of questions they asked:

* What do you think new employees need to learn to be successful in their first year on the job?
* Have you gathered data from new employees before? If so, may we have access to the historical data?
* Do you believe managers with higher retention rates offer new employees something extra or unique?
* What do you suspect is a leading cause of dissatisfaction among new employees?
* By what percentage would you like employee retention to increase in the next fiscal year?

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 an online survey of new employees. These were the things they did to prepare:

* They developed specific questions to ask about employee satisfaction with different business processes, such as hiring and onboarding, and their overall compensation.
* They established rules for who would have access to the data collected - in this case, anyone outside the group wouldn't have access to the raw data, but could view summarized or aggregated data. For example, an individual's compensation wouldn't be available, but salary ranges for groups of individuals would be viewable.
* They finalized 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.

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**. Collecting and using data ethically is one of the responsibilities of data analysts. In order to maintain confidentiality and protect and store the data effectively, these were the steps they took:

* They restricted access to the data to a limited number of analysts.
* They cleaned the data to make sure it was complete, correct, and relevant. Certain data was aggregated and summarized without revealing individual responses.
* They uploaded raw data to an internal data warehouse for an additional layer of security.

Analyze

Then, the analysts did what they do best: analyze! From the completed surveys, the data analysts **discovered** that an employee’s experience with certain processes was a key indicator of overall job satisfaction. These were their findings:

* Employees who experienced a long and complicated hiring process were most likely to leave the company.
* Employees who experienced an efficient and transparent evaluation and feedback process were most likely to remain with 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 diminish trust in the survey process and reduce their ability to collect truthful data from employees in the future.

Share

Just as they made sure the data was carefully protected, the analysts were also careful **sharing the report**. This is how they shared their findings:

* They shared the report with managers who met or exceeded the minimum number of direct reports with submitted responses to the survey.
* They presented the results to the managers to make sure they had the full picture.
* They asked the managers to personally deliver the results to their teams.

This process gave managers an opportunity to **communicate the results** with the right context. As a result, they could have productive team conversations about next steps to improve employee engagement.

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. These were their recommendations:

* Standardize the hiring and evaluation process for employees based on the most efficient and transparent practices.
* Conduct the same survey annually and compare results with those from the previous year.

1. **Ask**: business challenge, objective, or question
2. **Prepare**: data generation, collection, storage, and data management
3. **Process**: data cleaning and data integrity
4. **Analyze**: data exploration, visualization, and analysis
5. **Share**: communicating and interpreting results
6. **Act**:  putting insights to work to solve the problem

**EMC's data analysis process**

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

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

**SAS's iterative process**

An iterative data analysis process 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.

**Project-based data analytics process**

A project-based data analytics process has five simple steps:

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

This data analytics project process was developed by Vignesh Prajapati. It doesn’t include the sixth phase, or the act phase.

**data analysis** is the collection, transformation, and organization of data in order to draw conclusions,

make predictions, and drive informed decision-making.

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

Analytical thinking involves identifying

and defining a problem and then

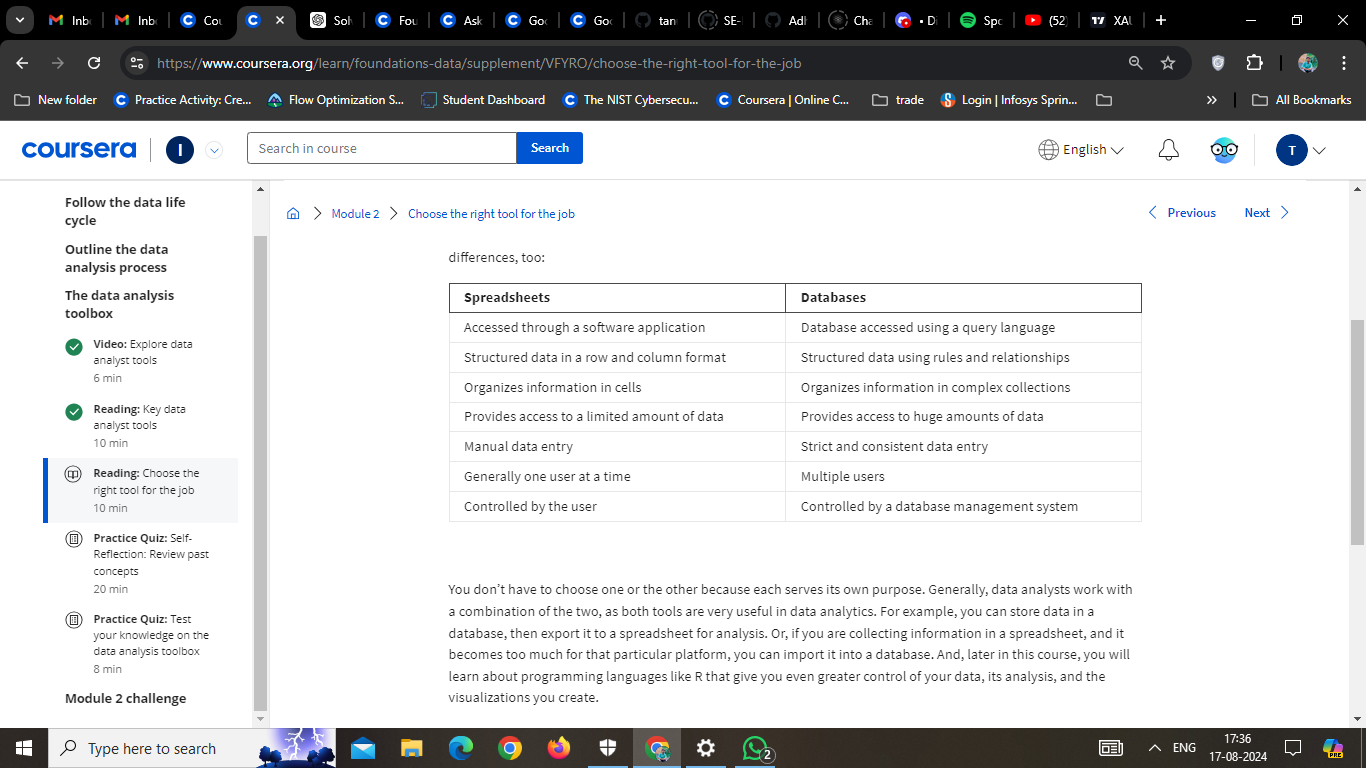
solving it by using data in

an organized, step-by-step manner

**five whys for root cause analysis**

Think of it like this:

* **You start with a problem:** For example, "Customers are complaining about damaged grocery deliveries."
* **You ask "Why?"** "Why are customers complaining about damaged grocery deliveries?" The answer might be: "Because the products are arriving damaged."
* **You ask "Why?" again:** "Why are the products arriving damaged?" Maybe the answer is: "Because they are not packaged properly."
* **You keep asking "Why?"** You continue this process of asking "Why?" for each answer until you get to the root cause.



**Data analysis process:**

* Ask
* Prepare
* Process
* Analyze
* Share
* Act

**Data life cycle:**

* Plan
* Capture
* Manage
* Analyze
* Archive
* Destroy