WEEK 3

**Follow the data life cycle**

* The data analysis process consists of six phases: ask, prepare, process, analyze, share, and act.
* This program is designed around these six phases to teach you the data analysis process.
* To understand how data analysts work with data, it's essential to grasp the data life cycle.
* Data analysts use various data analysis tools, including spreadsheets, databases, query languages, and visualization software.
* Learning these tools is a gradual process, and everyone starts with little to no knowledge.
* The speaker shares a personal experience of learning and becoming proficient with spreadsheets during an internship.
* Gaining expertise in data analysis tools allows analysts to bring data to life and make valuable contributions in their roles.
* The program will guide you through the learning process to help you become proficient in data analysis.
* Let's continue exploring the data analysis process and the data life cycle.

----------------------------------------------------------------------------------------------------------

* The data life cycle consists of six phases: plan, capture, manage, analyze, archive, and destroy.
* The first phase is planning, where a business defines the type of data needed, how it will be managed, who is responsible, and the desired outcomes.
* Data capture involves collecting data from various sources, including external datasets and internal databases.
* Databases are collections of data stored in computer systems, often used for managing and storing company information.
* Data integrity, credibility, and privacy are essential considerations during data management.
* Data analysis is the phase where data is used to solve problems, make decisions, and support business goals.
* Archiving involves storing data in a retrievable but less actively used state, reducing clutter during analysis.
* The destroy phase ensures secure data erasure to protect sensitive information and maintain data privacy.
* Understanding the data life cycle helps in approaching the data analysis process effectively.
* Each phase of the data life cycle plays a critical role in managing and leveraging data for business purposes.

## Outlining the data analysis process

* The data analysis process consists of several steps: ask, prepare, process, analyze, share, and act.
* In the "ask" phase, data analysts define the problem and understand stakeholder expectations.
* Stakeholders are individuals or groups who have invested in the project and have specific goals and expectations.
* Effective communication with stakeholders is crucial in understanding their needs and keeping the project on track.
* The "prepare" phase involves collecting and storing data for analysis, identifying useful data types, and ensuring data objectivity and impartiality.
* Data analysts also focus on cleaning, transforming, and validating data in the "prepare" phase.
* The "process" phase includes error elimination, data transformation, data combination, and outlier removal.
* Data cleansing ensures that the data is complete and correct, fixing any typos, inconsistencies, or inaccuracies.
* In the "analyze" phase, data analysts use tools to transform and organize data to draw conclusions and make informed decisions.
* Spreadsheets and SQL (Structured Query Language) are tools commonly used in data analysis.
* The "share" phase focuses on interpreting and presenting results to stakeholders effectively, emphasizing the importance of data visualization.
* Data visualization helps convey complex concepts and data insights to others.
* A course on using the programming language R is provided to enhance data manipulation, calculation, and visualization skills.
* The "act" phase is when the business applies the insights gained from data analysis to solve the original problem and make data-driven decisions.
* Completing a case study project during the "act" phase is an opportunity to showcase skills and stand out in job interviews.

----------------------------------------------------------------------------------------------------

The data analysis process involves several key steps, which can be applied to various types of data analysis projects. Here's an overview of the process using an example of an employee engagement survey:

1. **Ask**: In this initial phase, data analysts define the problem they aim to solve and gather information about stakeholder expectations. Asking the right questions is crucial to understanding the scope and purpose of the analysis.
2. **Prepare**: During the preparation phase, analysts determine the type of data needed to address the key questions identified in the "ask" phase. This includes considering whether quantitative or qualitative data is required and whether data collection is necessary. It's important to plan how data will be collected if needed.
3. **Process**: Data processing begins with data cleaning, which involves understanding the structure of the data, checking for missing or inconsistent data, and addressing outliers. This phase also includes running quality assurance checks to ensure data integrity.
4. **Analyze**: In the analysis phase, data analysts objectively examine the data to draw insights. It's essential to avoid biases and let the data speak for itself. Analysts perform planned analyses based on the questions defined in the "ask" phase.
5. **Share**: Data and insights are shared with stakeholders, often starting with high-level findings presented to executives. Effective data visualization and presentation skills play a significant role in conveying insights to others.
6. **Act**: The final phase involves taking action based on the insights gained from the analysis. It's critical to use data-driven insights to make informed decisions and introduce interventions if necessary. This phase is where the real work of implementing changes begins.

The data analysis process is rigorous and requires attention to detail, objectivity, and effective communication with stakeholders. It's important to follow each step of the process to ensure that meaningful insights are derived from the data.

## The data analysis toolbox

Understanding the tools used in data analysis is crucial for a data analyst. Here's a brief overview of the key tools mentioned in the video:

1. **Spreadsheets (e.g., Microsoft Excel, Google Sheets)**: Spreadsheets are digital worksheets used to store, organize, and manipulate data. They are essential for structuring data, performing calculations, and creating visualizations. Formulas and functions in spreadsheets allow you to automate calculations and data manipulation.
2. **Structured Query Language (SQL)**: SQL is a programming language used to communicate with databases. Databases store large volumes of structured data, and SQL allows data analysts to retrieve, modify, and analyze data within these databases. SQL queries are used to request specific data or perform operations like data retrieval, insertion, updating, or deletion.
3. **Data Visualization Tools (e.g., Tableau, Looker)**: Data visualization tools enable data analysts to create graphical representations of data, such as charts, graphs, maps, and tables. Visualizations make it easier to understand complex data and convey insights to stakeholders effectively. These tools help in presenting data in a visually appealing and informative way.

Each of these tools plays a unique role in the data analysis process. Spreadsheets help in data preparation and basic analysis, SQL is used for querying databases and performing advanced data manipulations, and data visualization tools assist in presenting data insights to a broader audience.

As you progress in your data analysis journey, you will dive deeper into these tools, learn how to use them effectively, and understand when to apply each one to various data analysis tasks. These tools will empower you to work with data efficiently and communicate your findings clearly to stakeholders.

key data analyst tools and their importance:

**Spreadsheets (e.g., Microsoft Excel, Google Sheets):**

* Used for collecting, organizing, and storing data.
* Aid in identifying patterns and creating data visualizations.
* Facilitate data structuring for specific projects.

**Databases and Query Languages (e.g., SQL, MySQL, Microsoft SQL Server, BigQuery):**

* Databases store structured data.
* Query languages like SQL allow analysts to:
  + Isolate specific data from databases.
  + Understand and communicate with databases.
  + Select, create, add, or download data for analysis.

**Visualization Tools (e.g., Tableau, Looker):**

* Transform complex data into understandable stories.
* Assist stakeholders in making informed decisions and forming effective strategies.
* Offer features like interactive graph creation (Tableau) and direct data connectivity (Looker).

**Programming Languages (e.g., R, Python):**

* Used for statistical analysis, data visualization, and other data-related tasks.
* Commonly employed by data analysts in their work.

These tools empower data analysts to explore and interpret data effectively, supporting informed decision-making and business strategies.

-------------------------------------------------------------------------------------------------------------------------------

choosing the right tool for the job as a data analyst:

**Spreadsheets:**

* Suitable for organizing, cleaning, and analyzing data.
* Structured in a row and column format.
* Often used for manual data entry.
* Generally limited to one user at a time.
* Controlled by the user.

**Databases:**

* Ideal for storing and managing large volumes of data.
* Structured using rules and relationships.
* Require strict and consistent data entry.
* Allow multiple users simultaneously.
* Controlled by a database management system (DBMS).

Data analysts often use a combination of spreadsheets and databases, depending on the specific needs of their projects. Spreadsheets are valuable for initial data entry and analysis, while databases are better suited for handling large-scale data storage and retrieval. Additionally, programming languages like R offer further capabilities for data control, analysis, and visualization.