Week 1

- Structured thinking: Structured thinking is the process of recognizing the current problem or situation, organizing available information, revealing gaps and opportunities, and identifying the options.
- Problem types:
    1) making predictions (using data to make an informed decision about how things may be in the future),
    2) categorizing things (assigning information to different groups or clusters based on common features),
    3) spotting something unusual (identifying data that is different from the norm),
    4) identifying themes (grouping categorized information into broader concepts),
    5) discovering connections (finding similar challenges faced by different entities and combining data and insights to address theme), and
    6) finding patterns, using historical data to understand what happened in the past and is therefore likely to happen again
- SMART methodology: specific, measurable, action-oriented, relevant and time –bound.
    1) Specific questions are simple, significant, and focused on a single topic or a few closely related ideas.
    2) Measurable question can be quantified and assessed.
    3) Action-oriented questions encourage changes.
    4) Relevant questions matter, are important, and have significance to the problem you're trying to solve

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?

Week 2

- Data-inspired decision making: explores different data sources to find out what they have in common

- Algorithm: a process or set of rules to be followed for a specific task

- Quantitative data: specific and objective measures of numerical facts (the what, how many, how many)
- qualitative data describes subjective or explanatory measures of qualities and characteristics or things that can't be measured with numerical data, like your hair color

- Report: static collection of data given to stakeholders periodically

Reports

Pros

- High-level historical data
- Easy to design
- Pre-cleaned and sorted data

Cons

- Continual maintenance
- Less visually appealing
- Static

- Dashboard: monitors live, incoming data.

Dashboards

Pros

- Dynamic, automatic, and interactive
- More stakeholder access
- Low maintenance

Cons

- Labor-intensive design
- Can be confusing
- Potentially uncleaned data

- A pivot table is a data summarization tool that is used in data processing. Pivot tables are used to summarize, sort, re-organize, group, count, total, or average data stored in a database. It allows its users to transform columns into rows and rows into columns.
- Metric: single, quantifiable type of data that can be used for measurement

The following table summarizes the benefits of using a dashboard for both data analysts and their stakeholders.

Benefits	For Data Analysts	For Stakeholders
Centralization	Sharing a single source of data with all stakeholders	Working with a comprehensive view of data, initiatives, objectives, projects, processes, and more
Visualization	Showing and updating live, incoming data in real time*	Spotting changing trends and patterns more quickly
Insightfulness	Pulling relevant information from different datasets	Understanding the story behind the numbers to keep track of goals and make data-driven decisions
Customization	Creating custom views dedicated to a specific person, project, or presentation of the data	Drilling down to more specific areas of specialized interest or concern

Small data   Big data

- Specific
- Short time-period
- Day-to-day decisions

- Large and less-specific
- Long time-period
- Big decisions

Small data	Big data
Describes a data set made up of specific metrics over a short, well-defined time period	Describes large, less-specific data sets that cover a long time period
Usually organized and analyzed in spreadsheets	Usually kept in a database and queried
Likely to be used by small and midsize businesses	Likely to be used by large organizations
Simple to collect, store, manage, sort, and visually represent	Takes a lot of effort to collect, store, manage, sort, and visually represent
Usually already a manageable size for analysis	Usually needs to be broken into smaller pieces in order to be organized and analyzed effectively for decision-making

The three (or four) V words for big data

When thinking about the benefits and challenges of big data, it helps to think about the three Vs: **volume, variety,** and **velocity.** Volume describes the amount of data. Variety describes the different kinds of data. Velocity describes how fast the data can be processed. Some data analysts also consider a fourth V: **veracity.** Veracity refers to the quality and reliability of the data. These are all important considerations related to processing huge, complex data sets.

Volume	Variety	Velocity	Veracity
The amount of data	The different kinds of data	How fast the data can be processed	The quality and reliability of the data

Week 3

- **Plan** for the users who will work within a spreadsheet by developing organizational standards. This can mean formatting your cells, the headings you choose to highlight, the color scheme, and the way you order your data points. When you take the time to set these standards, you will improve communication, ensure consistency, and help people be more efficient with their time.

- **Capture** data by the source by connecting spreadsheets to other data sources, such as an online survey application or a database. This data will automatically be updated in the spreadsheet. That way, the information is always as current and accurate as possible.

- **Manage** different kinds of data with a spreadsheet. This can involve storing, organizing, filtering, and updating information. Spreadsheets also let you decide who can access the data, how the information is shared, and how to keep your data safe and secure.

- **Analyze** data in a spreadsheet to help make better decisions. Some of the most common spreadsheet analysis tools include formulas to aggregate data or create reports, and pivot tables for clear, easy-to-understand visuals.

- **Archive** any spreadsheet that you don't use often, but might need to reference later with built-in tools. This is especially useful if you want to store historical data before it gets updated.

- **Destroy** your spreadsheet when you are certain that you will never need it again, if you have better backup copies, or for legal or security reasons. Keep in mind, lots of businesses are required to follow certain rules or have measures in place to make sure data is destroyed properly.

Error	Description	Example
#DIV/0!	A formula is trying to divide a value in a cell by 0 (or an empty cell with no value)	=B2/B3, when the cell B3 contains the value 0
#ERROR!	(Google Sheets only) Something can't be interpreted as it has been input. This is also known as a parsing error.	=COUNT(B1:D1 C1:C10) is invalid because the cell ranges aren't separated by a comma
#N/A	A formula can't find the data	The cell being referenced can't be found
#NAME?	The name of a formula or function used isn't recognized	The name of a function is misspelled
#NUM!	The spreadsheet can't perform a formula calculation because a cell has an invalid numeric value	=DATEDIF(A4, B4, "M") is unable to calculate the number of months between two dates because the date in cell A4 falls after the date in cell B4
#REF!	A formula is referencing a cell that isn't valid	A cell used in a formula was in a column that was deleted
#VALUE!	A general error indicating a problem with a formula or with referenced cells	There could be problems with spaces or text, or with referenced cells in a formula; you may have additional work to find the source of the problem.

- Structured thinking: the process of recognizing the current problem or situation, organizing available information, revealing gaps and opportunities, and identifying the options.
- A scope of work or SOW is an agreed- upon outline of the work you're going to perform on a project
- A statement of work is a document that clearly identifies the products and services a vendor or contractor will provide to an organization. It includes objectives, guidelines, deliverables, schedule, and costs.
- What is a good SOW?
    1) Deliverables: What work is being done, and what things are being created as a result of this project? When the project is complete, what are you expected to deliver to the stakeholders? Be specific here. Will you collect data for this project? How much, or for how long?
    2) Milestones: This is closely related to your timeline. What are the major milestones for progress in your project? How do you know when a given part of the project is considered complete?
    3) Timeline: Your timeline will be closely tied to the milestones you create for your project. The timeline is a way of mapping expectations for how long each step of the process should take. The timeline should be specific enough to help all involved decide if a project is on schedule. When will the deliverables be completed? How long do you expect the project will take to complete?
    4) Reports: Good SOWs also set boundaries for how and when you'll give status updates to stakeholders. How will you communicate progress with stakeholders and sponsors, and how often? Will progress be reported weekly? Monthly? When milestones are completed? What information will status reports contain?
- Context is the condition in which something exists or happens.

Context can turn raw data into meaningful information. It is very important for data analysts to contextualize their data. This means giving the data perspective by defining it. To do this, you need to identify:

- Who: The person or organization that created, collected, and/or funded the data collection
- What: The things in the world that data could have an impact on
- Where: The origin of the data
- When: The time when the data was created or collected
- Why: The motivation behind the creation or collection
- How: The method used to create or collect it

Week 4
- Stakeholders: people that have invested time, interest, and resources into the projects you'll be working on as a data analyst.
- there are three common stakeholder groups that you might find yourself working with: the executive team, the customer-facing team, and the data science team
    1) The executive team provides strategic and operational leadership to the company. They set goals, develop strategy, and make sure that strategy is executed effectively. The

executive team might include vice presidents, the chief marketing officer, and senior-level professionals who help plan and direct the company's work.

2) The customer-facing team includes anyone in an organization who has some level of interaction with customers and potential customers. Typically they compile information, set expectations, and communicate customer feedback to other parts of the internal organization. These stakeholders have their own objectives and may come to you with specific asks.

3) Data science team

- Focus on what matters
    1) Who are the primary and secondary stakeholders
    2) Who is managing the data
    3) Where can you go for help

Before you communicate, think about

1. Who your audience is

2. What they already know

3. What they need to know

4. How you can communicate that effectively to them

-
-