Course 1 - Foundations: Data, Data, Everywhere

- The **data life cycle** deals with the stages that data goes through during its useful life; **data analysis** involves following a process to analyze data.

1. **Ask** questions and define the problem
   - Make sure we fully understand stakeholder expectations, determine who they are; keep focus on the problem itself, ask effective questions
   - **Stakeholders -** people who have invested time and resources into a project and are interested in the outcome
     - **Questions** to ask yourself in this step:
       - What are my stakeholders saying their problems are?
       - Now that I've identified the issues, how can I help the stakeholders resolve their questions?

2. **Prepare** data by collecting and storing the information (what type of data and how)
   - identify which kinds of data are most useful for solving a particular problem
   - needs to be objective and unbiased
   - quantitative vs qualitative; longitudinal vs cross-sectional - what metrics to measure
   - new data vs data that already exists and working with data owners - locate data in your database
   - create security measures to protect the data
   - We need to be thinking about the type of data we need in order to answer the questions that we've set out to answer based on what we learned when we asked the right questions.
     - **Questions** to ask yourself in this step:
       - What do I need to figure out how to solve this problem?
       - What research do I need to do?

3. **Process** data by cleaning and checking the information, getting details right
   - find and eliminate any errors and inaccuracies, transforming it into a more useful format, combining datasets to make info more complete and removing outliers
   - gain strategies for verifying and sharing data cleansing with stakeholders
   - see if we're missing data or missing in a systematic way (coded correctly); quality assurance checks
   - use spreadsheet functions to find incorrectly entered data, SQL functions to check for extra spaces, check for bias, remove repeated entries
     - **Questions** to ask yourself in this step:
       - What data errors or inaccuracies might get in my way of getting the best possible answer to the problem i'm trying to solve?
       - How can I clean my data so the information I have is more consistent?

4. **Analyse** data to find patterns, relationships, and trends.
   - transform, organise that info to draw useful conclusions, make predictions and drive informed decision-making
   - tools - spreadsheets, structured query language (SQL - sequel)
   - let the data speak for itself, we amplify it in an unbiased and objective way
   - perform calculations, combine data from multiple sources, create tables with results
     - **Questions** to ask yourself in this step:
       - What story is my data telling me?
       - How will my data help me solve this problem?

           - Who needs my company's product or service?

                  - What type of person is most likely to use it?

5. **Share** data with your audience.
   - clear and enticing visualisation is essential to getting others to understand what your data is telling you
   - show stakeholders that you solved their problem and how you got there
   - R is a popular tool for data manipulation, calculation, and visualisation
   - sharing will help the team make better and more informed decisions, lead to stronger outcomes

       Questions to ask yourself in this step:
       - How to make presentation to stakeholders engaging and easy to understand
       - What would help me understand this if I were the listener?

6. **Act** on the data and use the analysis results.
   - business takes all of the insights you have provided and puts them into work to solve the original business problem

   Questions to ask yourself in this step:
   -How can I use the feedback I received during the share phase (step 5) to actually meet the stakeholder's needs and expectations?

- **Data analysis life cycle** - process of going from data to decision
    - Create, consume, test, process, reuse
- As described in the Harvard Business Review video below, business analytics is "the use of math and statistics to derive meaning from data in order to make better business decisions."
-
- Data ecosystems - elements that interact with one another in order to produce, manage, store, organise, analyze, and share data
- Data science - creating new ways of modeling and understanding the unknown by using raw data
- Data analysts - find answers to existing questions by creating insights from data sources
- Data analysis - collection, transformation and organisation of data to draw conclusions, make predictions, and drive informed decisions
- Data analytics - the science of data - everything fits under this umbrella term
- **Subject matter experts** - ability to look at the results of data analysis and identify inconsistencies, make sense of grey areas, validate choices
- Data + business knowledge = solution
    - 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?
- **data driven decision making -** Collecting, observing and analysing facts to guide business strategy; includings insights from ppl who are familiar with the business problem
    - Using the data discovered through data analysis
    - Decision making - it's a choice between consequences

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

- **Analytical skill sets**: qualities and characteristics associated with solving problems using facts
    - **Curiosity** - wanting to learn something, seeking new challenges and experiences - leads to knowledge
        - Start to see patterns and relationships in everyday life;
    - **Context** - condition in which something exists or happens (structure, environment) helps group things into categories
        - use context to make predictions, research answers, draw conclusions
        - Identifying the motivation behind data collection
    - **Technical skills -** ability to break things down into smaller steps and work with them in an orderly and logical way
        - Build on gut instinct and use a more technical approach to explore them, alway seek out the facts, use insights to make decisions
    - **Data design -** how you organise information
        - designing your data so that it is organised in a logical way makes it easy for data analysts to access, understand, and make the most of available information
        - If you make decisions that are informed by data, you are more likely to make more informed and effective (data-driven)decisions
    - **Data strategy** - management of the people, processes and tools used in data analysis; make sure the people involved know how to use the data
        - Gives you high-level view of the path needed to achieve your goals
        - Make sure everyone is onboard and on the same page - need specific procedure in place; technology used is aligned

- **Analytical thinking -** identifying and defining a problem and then solving it by using data in an organized, step by step manner
    - **Visualisation** - graphical representation of info
    - **Strategy** - help stay focused and on track - see what they want toa chive and how to get there, improves the quality and usefulness of data
    - **Problem-orientation** - keeping problem at top of mind to identify, describe, and solve problems; improves communication and save time
    - **Correlation** - identify between two or more pieces of data; correlation doesn't mean causation - just b/c two pieces of info are trending in teh same direction, doesn't mean they are related
    - **Big picture thinking** - looking at the complete puzzle; can't only focus on individual pieces - zoom out, and see possibilities and opportunities
        - **Detail oriented thinking** - figuring out all aspects htat will help execute a plan

- in data analysis, solutions are almost never right in front of you. You need to think critically to find out the right questions to ask.
    - But you also need to think creatively to get new and unexpected answers
- Common questions:
    - What is the root cause of a problem - ask why five times to reveal the root cause - final anser should give you some useful and surprising insights
    - Where are the gaps in our process - **gap analysis** - method for examining and evaluating how a process works currently to get where you want to be in the future - understand where you are now and compare
    - What did we not consider before - what info or procedure might be missing
    - **Quartile** - divides data poitns into 4 equal quarters

- **Data life cycle -**
    - **Plan -** what kind of data it needs, how it will be managed and by who
    - **Capture** - data is collected from variety of sources and brought into organization; finding the best way ; public data base vs internal
        - database is a collection of data stored in a computer system
    - **Manage** - how we take car of our data, how and where it's stored, the tools to keep it safe and secure and actions to make sure it's properly maintained
    - **Analyse -** the data is used to solve problems, make great decisions, and support business goals
    - **Archive** - storing data in a place where it's still available but may not be used again - good way to store data thats no longer relevant
    - **Destroy** - data erasure software, shredding papers - protects company's and customers' private info

- **Query language -** computer programming language that allows you to retrieve and manipulate data from a database
    - **SQL** - A language that lets data analysts communicate with a database
        - Requesting that a database do something for you
    - Allow analysts to isolate specific info from a database
    - Crate, select, add, or download data from a database for analysis
- **Data visualisation** - graphical representation of information
    - Tableau - create visuals that are easy to understand even by non-technical users, has simple drag and drop feature; interactive graphs in dashboards and worksheets
    - looker - easy way to create visuals based on query; communicates with database directly, can connect your data right to the visual tool you choose
    - Helps stakeholders come up with conclusions that lead to informed decisions and effective business strategies

- Function - predefined operation in a spreadsheet
- **Formula** - a set of instructions used to perform a specific calculation

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

- Analysts work with a combo of the two - store in database and export it to spreadsheet for analysis
- Or if collecting info in a spreadsheet and it becomes too much, you can import it to a database
- In a table, an attribute is a characteristic or quality of data used to label a column.
- In a data table, a row is called an observation. An observation includes all of the attributes for what is contained in the row. An attribute is a quality or characteristic of data.
-
- **Structured Query Language (SQL) 'sequel' -** enables data analysts to talk to their databases
- **Query** - a request for data or info from a database
- **Syntax** - predetermined structure of a language that includes all required words, symbols, and punctuation - every programming language follows syntax
    - **Select** - choose **columns** you want to return
    - **From** - choose the **tables** where the columns you want are located
    - **Where** - filter for an exact value match or data that matches a certain condition
        - *Need to use underscore in lieu of space if more than one word*

- SQL query is like filling in a template

```
SELECT
      field1
FROM
      table
WHERE
      field1 = condition;
```

- Semicolon - a statement terminator, some may not have it
- **LIKE -** tells database to look for a certain pattern
- **Percent sign % -** used as a wildcard to match on or more characters

WHERE field1 LIKE 'Ch%'

vs

WHERE field1 = 'Chavez'

- **SELECT*** - you would select all fo the columns instead of just filed1 like in the example above
- **Comments** - can add them to help you understand what names that are unclear represents - they are placed between /* and */ or after two dashes –

```
SELECT
      field1 /* this is the last name column */
FROM
      table -- this is the customer data table
WHERE
      field1 LIKE 'Ch%';
```

- Be consistent

- **alias** - good for the duration of the query only and doesn't change the actual name, just makes it easier for yourself to work with
    - **AS** clause

```
field1 AS last_name -- Alias to make my work easier
table AS customers -- Alias to make my work easier

SELECT
      last_name
FROM
      customers
WHERE
      last_name LIKE 'Ch%';
```

- **AND** clause - tests for two conditions

```
SELECT
        *
FROM
        Employee
WHERE
        jobCode <> 'INT'
        AND salary <= 30000;
```

- Data visualisation - to explain complex data quickly, reinforce data analysis ; make data easy to understand and interesting to look at
- Planning visualisation
    1. Explore the data for patterns - get access to current sales records, website analytics reports
    - customers behavior on website - who visited, who bought, how much
    - can see patterns like who visited most frequently based on geography and larger amounts spent
    - can explain why sales are so strong right now and help company find ways to make them even stronger through the new website
    2. Plan your visuals - refine the data so you can combine your data from different tables into one that explain findings quickly and effectively (decide what info and insights you want to display)
    - if audience is sales oriented, should use - sales numbers over time, show sales to location relationship, also between sales and website use, which customers fuel growth
    3. Create your visuals - a process that involves trying diff visualisation formats, making adjustments
    - a mix would best communicate your findings and turn analysis into the most compelling story
    - line charts - track sales over time; maps - connect sales to location; donut charts - show customer segments; bar charts - compare total visitors and visitors that purchases

- Data visualisation toolkit
    1. spreadsheets - simple line and bar charts
    - pie charts, some advanced like maps, waterfall and funnel diagrams
    2. Tableau - can integrate data into dashboard-style visualisations
    - can pull data from almost any system and turn into compelling visuals or actionable insights
    - has built in visual best practices - makes analysing and sharing data fast, easy and useful

3. Programming language R - use visualisation tools in Rstrudio - integrated developer environment (IDE)

- Data determines a website's reliability and accuracy to make sure you get the most useful results in a search you make on google
- Data analyst help predict rush hours at hospitals, weather patterns to help the zoo staff according to traffic
- Data analysts like problem solving - taking a complex problem, mystery, riddle and being able to find the answers and come up with a solution, also being able to work with people and help people
- First starting off - generalist; later on go into something more specific
    - SQL expert - manipulate large data sets for financial analysis purposes
    - Operations functions - business analyst, data analysts
    - Government industry - use population data to determine which communities need federal funding
    - Marketing - use demographic data to target ads for a new product

- **Issue** - a topic or subject to investigate
- **Question** - designed to discover information
- **Problem** - an obstacle or complication that needs to be worked out
- **Business task** - question or problem data analysis answers for a business

- Data helps us see the whole picture, vs using observation and gut feelings
    - You can see all of the problems, its causes and find surprising new solutions
    - Makes businesses make better decisions
    - Better understand customer buying habits
    - Create more effective social media messaging

- Analyzing historical data such as weather patterns, social media comments, and past sales would provide useful insights into the increase in sales/subscription

- responsibility as a data analyst - make sure you're gathering, analyzing, and presenting data in a way that's fair to the people being represented by that data.
- **Fairness** - ensuring that your analysis doesn't create or reinforce bias
    - Think about this from the moment you start collecting data to the time you present
- Beneficence - improving the lives of people using data
- Are the people represented in this data going to be benefited by this , you have a responsibility to these people

- Keen in mind how to keep aspects of the data private and protected
- Need to ensure that there's actionable ways people can consent to giving those data and ways they can ask for it to be revoked or removed - give consumers more control over their data - data are people

- Think about interest during job search - employers want to know why you're interested in their company and how you can address their needs - speak about your motivation to work in data analytics during interviews

- **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
- Many jobs may combine or look for candidates with overlapping skills

	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

- 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 - think about a time where you used data to solve a problem - professional or personal
    - Prepare questions for the interviewer - questions that help you understand the team and the job better
    - If case study - expect to be given a business problem, with sample data set
    - Have to analyse and come up with solution ; make sure you are coming up with solutions that relate back to data (interviewers are looking to see your thought process and the way you get to your solution)

Course 2 - Ask Questions to Make Data-Driven Decisions

- Every single action we do is creating some kind of data, most of it is meaningless until someone adds interpretation around it
- the goal of all data analysts is to use data to draw accurate conclusions and make good recommendations - need complete, correct and relevant data
- Need to interpret data correctly:
    - **When data is interpreted incorrectly, it can lead to huge losses**
- **Structured thinking**:
    1. Recognizing the current problem or situation
    2. Organising available information
    3. Revealting gaps and opportunities
    4. Identifying our options
- Example situation:
- Things to think about when advertising - target audience, budget
- Ask questions to define the problem (not knowing target audience)
- Collab with stakeholders to understand their needs
- Prepare - collect data on different ad methods to determine which was the most popular among people with video games
- Process - cleaned the data to eliminate errors or inaccuracies - transform it to a more useful format
- Analyse - who's most likely to own gaming system, where these ppl would likely see ads - conclusions people 18-34 - tv commercials and podcast
Podcast would be more cost effective
- Share with clear and compelling visuals

- **6 common types of problems data analysts try to solve:**
    1. **Making Predictions** - using historical data to make an informed decision about how things may be in the future

Eg. predict future health problems using data from patient monitoring combined with risk factors, age, etc. may reduce hospitalizations

2. **Categorising things** - assigning info to diff groups or clusters based on common features
Eg. create groups for employees who are most and least effective at engineering, repair and maintenance, assembly
- Categorising key words that come up during calls - politeness, satisfaction, etc.

3. **Spotting something unusual** - identify data that is diff from the norm
Eg. a sudden increase in student registration at a school - analyst may discover that new apartment complexes had been built

4. **Identify themes -** grouping categorised info into broader concepts
Eg. from earlier, group into broader concept of low and high productivity - help business reward top performers

5. **Discovering connections** - find similar challenges faced by diff entities and combining data and insights to address them
Eg. scooter company has shortage of wheels, wheel supplier has problem with the rubber, rubber supplier also short of materials
- third party logistics - view each other's timelines to see what causes wait times

6. **Finding patterns** - using historical data to understand what happened in the past and likely to happen again
Eg. look at transaction data to understand customer buying habits - buying more canned goods before hurricane, fewer winter accessories bought in winter months

- Asking effective questions lead to key insights to solve problems
- Should be open-ended, avoid leading questions, close-ended, and vague
    - What price range would you consider purchasing this product vs is this too expensive ; yes or no, doesn't encourage elaboration; no context
- **SMART questions**
    - **Specific** - simple, significant, and focused on a single topic or few closely related ideas
    - **Measurable** - can be quantified and assessed - how many times was our video shared the first week; can also include yes or no questions, rank, scale
    - **Action-Oriented** - encourage change, devise an action plan- What design features will make our packaging easier to recycle
    - **Relevant, Time-bound -** what environmental factors changed in Durham between **1983-2004** that could cause tree frogs to disappear
        - Limits the range of possibilities

- **Data-inspired decision-making**-explores diff 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 measure - number, quantity, range; answers what, how many, how often

- **Qualitative data -** subjective or explanatory measures of qualities and characteristics - answers why questions

- **Report** - static collection of data given to stakeholders periodically
    - Pros - high-level historical data, easy to design, pre-cleaned and sorted
    - Cons - continual maintenance, less visually appealing, static
- **Dashboard -** monitors live, incoming data
    - Organises info from multiple datasets into one central location (tableau)
    - Pros - dynamic, automatic and interactive, more stakeholder access, low maintenance
    - Cons - labor-intensive design, can be confusing, potentially uncleaned data
    Strategic: focuses on long term goals and strategies at the highest level of metrics - quarterly or annual financials
    Operational: short-term performance tracking and intermediate goals - days, weeks, months
    Analytical: consists of the datasets and the mathematics used in these sets - details involved in usage, analysis, predictions - msotly used b data scientists

- **Pivot table** - a data summarization tool used in data processing
    - Used to summarise, sort, reorganise, group, coin, total, or average data stored in a database
    - Insert - create pivot table

- **Metric** - single, quantifiable type of data that can be used for measurement; set and evaluate goals
    - Data starts as a collection of raw facts, until we organise them into individual metrics that represent a single type of data
    - Can be combined into formulas
    - Eg. revenue - 2 metrics used - number of sales x sales price
    ROI - return on investment - a formula designed using metrics that let a business know how well an investment is doing - metrics used - net profit over a period of time /cost of investment
    - Marketing - metrics can be used to calculator customer retention rates - compare beginning and end period to see how effective marketing is
    - All trying to meet a specific goal by measuring data - meet monthly sales or a certain percentage of repeat customers

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

- **Mathematical thinking** - looking at a problem and logically breaking it down step-by-step to see the relationship of patterns in your data
- **Small data** - specific, short-time period - useful for day-to-day decisions - spreadsheet
- **Big data** - large and less specific, long- time period - large scale questions - use SQL

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

- benefits and challenges of big data
- **Volume** - the amount of data
- **Variety -** the different kinds of data
- **Velocity** - how fast the data can be processed
- **Veracity** - quality and reliability of the data

- **Spreadsheet tasks** - organise your data in a pivot table, sort and filter; calculate your data - formulas and functions
    - Plan for the users who will work within it - how to format cells, headings, the way you order data points
    - Capture data by the source by connecting spreadsheets to other data sources - online survey or a database - data will be updated automatically
    - Maange different kinds of data with a spreadsheet - storing, organising, filtering, updating; who can access, how info is shared
    - Analyze the data to make better decisions - formulas to aggregate data or create reports, pivot tables for clear easy to understand viduals
    - Archive - spreadsheets you don't use often but need to reference later; if you want to store historical data before it gets updated
    - Destroy - when you are certain you will enver need it again
- Titles should be short, clear and state exactly what the data int eh spreadsheet is about
- **Data range** - the set of cells a data analysis selects to include in a formula
- **Operator -** a symbol that names the type of operation or calculation to be performed - plus sign
- **Absolute referencing** - $ - will not change when you copy and paste the formula
- **Relative referencing** - will change anytime the formula is copied - use **F4** to change between absolute and relative
- **Mixed references** (cells partially referenced with a dollar sign, like $A2 or A$2
- COUNTIF() is a formula and a function. - the function runs based on criteria set by the formula - count is the formula and it will be executed IF the conditions you crate are true
    - =COUNTIF(A1:A16, "7") to count only the cells that contained the number 7
      Range, criterion
- DATEDIF(B6,C6, "M") - calculate the number of months between months
    - Start date first, end date, M = months

Spreadsheet errors:

- #DIV/0! - a formula is trying to divid a value in a cell by 0 or an empty cell
    - Fix - IFERROR(B4/A4, "Not applicable")
- #ERROR! (google sheets only) - formula can't be interpreted as input (parsing error)
    - Fix - delimiter -adding a comma between ranges
- #N/A - data in formula can't be found by the spreadsheet, most often occurs when using functions such as VLOOKUP - searches for a certain value in a column to return to a corresponding piece of info
    - FIx - could be typo (Almond vs Almonds)
- #NAME? - a formula or function name isn't understood
    - Fix - could be typo in the formula
- #NUM! - formula or function calculator can't be performed as specified
    - Fix - for DATEDIF - if the months were reversed
- #VALUE! - A general error that could indicate a problem with a formula or referenced cells

- Fix - often not clear right away, could be the cell was in text instead of numerical value
- #REF! - comes up when cells being referenced in a formula have been deleted
    - Avoid by using the SUM function and range of cells instead of adding the cell value by direct reference
- **Conditional formatting** - highlight cells a different color based on their contents.
    - extremely helpful when you want to locate all errors in a large spreadsheet. For example, using conditional formatting, you can highlight in yellow all cells that contain an error, and then work to fix them.
    - Format - conditional formatting - custom formula is - enter =ISERROR(A1), select color, done
- **function -** a preset command that automatically performs a specific process or task using the data - eg. SUM(A1:A6)
    - MIN(select every cell that should be included) same with MAX)
- **Problem domain** - the specific area of analysis tha tencompasses every activity affecting or affected by the problem
    - Need to understand this in order to have a structured approach

- **Scope of work - SOW** - an agreed-upon outline of the work you're going to perform on a project - work details, schedules, reports; data analyst - data preparation, validation, analysis of quantitative and qualitative datasets, initial results,
    - Addresses any confusion, contradictions, or questions about the data upfront
    - **Deliverables** - what work is being done, what things are being created as a result of thai project - be specific - use numbers, aim for measurable goals
    - **Milestones** - closely related to your timeline - major milestones, how do you know when a given part of the project is considered complete
        - Defined by you or stakeholders, project manager - eg. incremental steps ina larger project - collect 50% of required data, complete initial data analysis report, delivered completed dashboard
    - **Timeline** - closely tied ot milestones you create - away of mapping expectations; be specific enough to decide if a project is on schedule
        - When all deliverables will be completed, how long to expect the project to complete, how long each component of project will take
    - **Reports** - good SOWs also set boundaries for hwo and when you'll give status updates - how progress will be communicated, how often - weekly, monthly; when milestones are completed

- Staying objective within the context, consider these questions:
    - Who collected the data
    - What is it about, represent, how it relates to other data
    - When was the data collected - old may have certain limitations
    - Where it was collected
    - How it was collected - a survey might not be as effective as in-person

- Why it was collected - can have a strong relationship with bias - sometimes data is collected or even made up to serve an agenda
- Need to make sure you start with an accurate representation of the population and collect the data in the most appropriate and objective way
- Context is important because it helps yous ift through huge amount of disorganised data and turn it into something meaningful
- Data has little value if not paired with context
- Context turns raw data into meaningful information

3 common stakeholder groups:

1. **Executive team** - provides strategic and operational leadership to the company
   - set goals, develop strategy eg. VP, chief marketing officer, senior level professionals
   - less interested in details, time is limited - balance what info they will need to make informed decisions during your presentation
2. **Customer-facing team** - anyone who has some level of interaction with customers
   - they compile info, set expectations and communicate customer feedback
   - help by collecting and sharing data about consumers; buying behaviour to help inform product features
3. **Data science team -** work with them to boost company engagement to lower rates of employee turnover ; one analyst looks at employee productivity, another loosk at hiring data - then share this with data scientist on the team - they will use it to predict how new processes could boost employee productivity and engagement

- Remember to **discuss goals** - ask about the kind of results
  - Your work needs to line up with their needs
- Be empowered to **say no** when necessary- stakeholders may not realise the time and effort that goes into collecting and analysing data or you may need to prioritize other things - let them understand what needs to be done and what can be
- accomplished and they will be comfortable resetting their expectations
- **Plan for unexpected** - give yourself extra time for problem solving at each stage , make a list of potential roadblocks
- **Know your project -** keep track of discussions over email or reports, be ready to answer questions, know how your project connects to the rest of the company, be involved in providing the most insight possible, understand the why in your analysis
- **Start with words and visuals -** this helps to avoid the illusion of agreement and wrong assumptions
  - The faster everyone agrees, the faster you can perform the first analysis to test the usefulness of the project, measure the feedback, learn from the data, and implement changes.
- **Communicate often** - stakeholders want regular updates - share notes about milestones, setbacks, changes - use these notes to create shareable report
  - **Change log** - a file containing a chronologically ordered list of modifications made to a project - stakeholders can pop in and view updates whenever

- How to stay focused on objectives
    - Who are the primary and secondary stakeholders - VP; secondary - project manager, team members
    - Who is managing the data - other data analysts
    - Where can you go for help - project managers - help with timeline, guidance, resources, setting up efficient workflows
    -
    - The VP of sales provides strategic and operational direction but is less interested in specific details. Ning prepares questions ahead of time to focus on the key findings that the company expects from an annual sales report.
    - Members of the sales team have direct interactions with customers and are highly attuned to how the company performed over the past year. They can provide detailed information on the types of data that will matter most to the company's customers.
    - The data analysts each have a dataset that they focus on and can help pull the various types of data that Ning needs to satisfy the other stakeholders. Ning collaborates with them to complete the report.
    - The data science managers oversee all of the company's datasets and can help Ning prioritize the types of data and analyses required for the annual report. They can also advise on making an effective presentation.

Clear communication, think about:

- **Who your audience is** - connect with other data analysts working on the project, as well as project manager and VP of sales
- **What they already know** - other data analysts working on this project know all the details about which data-set you're using; project manger knows the timeline
- **What they need to know** - fellow data analysts need to know the details of where you have tried so far and any potential solutions you have come up with ; project manager needs to know the different teams that could be affected
- **How you can communicate effectively to them** - book in a meeting instead of a long e-mail; come up with a planh as a team

- Practice good writing habits for emails - would you be proud of what you written if it got published - make it clear and concise
- Set a reasonable and realistic timeline - stakeholders will know what to expect
- Send a high-level schedule with diff phases and their approximate dates
- The earlier you flag problems, the better
- If they request for changes you can prepare a report on how this change affects the project's timeline, also the pros and cons
- Balance stakeholder expectations and what's possible with the projects, resources and limitations

- A data analyst r**eframes a question**. Then, they **outline the problem**, challenges, potential solutions, and **timeframe** - this is to communicate expectation so stakeholders understand how long it will take to provide accurate info, put data into context and find the story it's telling and balances speed with accuracy (fast answers vs correct answers)
    - More time gains context

Limitations on data
- **incomplete or nonexistent data** - records that only go back 2 years , don't have enough data to reach a conclusion - be upfront with it
- **misaligned data** - if one team defines and measure things ina completely different way - establish how to measure things early on to standardise the data for greater reliability and accuracy
- **dirty data** - ones that contain errors - leads to productivity loss, unnecessary spending, unwise decision-making; make sure you clean the data - removing incorrect, corrupt, duplicate data

- **Tell a clear story:**
    - **Compare the same types of data -** data can get mixed up when you chart it for visualisation - double check if any segments in your chart display diff metrics
    - **Visualise with care -** set Y-axis to 0 to make sure your audience sees the full story clearly
    - **Leave out needless graphs** - if a table can show your story at a glance, stick with it instead of a pie chart or graph
    - **Test for statistical significance -** when two datasets look different - need to test whether the difference is real and important - run test to see how much confidence you can place in that difference
    - **Pay attention to sample size -** gather lots of data - if you have too little, be careful about using it to form judgments
- **Be the judge** - need to make sound judgments as a data analyst - when you know your limitations, you can make judgement calls that help ppl make better decisions
    - Take necessary steps to ensure data is complete and consistent
- **Process and outcome** - variable to consider when sharing data, questions:
    - Does your analysis answer the original question, are there angles you haven't considered, can you answer any questions that may get asked about your data and analysis, how detailed should you be, would a high level analysis be ok

- **Organizing a meeting:**
    - Identify your objective - establish purpose, goals, desired outcomes , questions that needs to be addressed
    - Acknowledge participants, keep them involved w/diff points of view
    - Organize the data to be presented
    - Prepare and distribute an agenda:

- Meeting start and end time, location, objective, background material participants should review
- Share it ahead of time
- **During meeting** - make intros, review key messages, present, discuss observations and implications of data, take notes, determine and summarize next steps
- **After -** prepare and distribute a brief recap of meeting with confirmed next steps and timeline for additional actions, can ask for feedback

- **Conflict resolution** - reframe the problem - how can I help you reach your goal
    - Discussion is key, are there other important things i should be considering
    - Understand the context - what their end goal, big picture is
    - Let's take a step back so i can better understand what you want to do with the data
    - I will consider the necessary steps an get back to you with a time estimate

- **Data analyst** - need attention to detail, communication, collaboration