

# Google Data Analytics Mind-Map Summary

By Olisa H. Ezeh

mind map of the broad concepts of using data analytics framework to solve business problems successfully by making data driven decisions.

# FOUNDATIONS: DATA, DATA, EVERYWHERE

## INTRODUCING DATA ANALYTICS AND ANALYTICAL THINKING

### Ask

Understand the challenges to be solved by asking variety of questions.

### PREPARE

Find and collect data to answer questions. Verify data sources and accuracy to ensure it can be useful.

### PROCESS

Clean and organise data, i.e removing inconsistencies, to ensure data is ready for analysis

TRANSFORM DATA  
INTO INSIGHTS

## 6 PHASES of DATA Analytics

### ANALYSE

Doing the necessary data analysis to uncover answers & solutions for example calculating averages or counting items to review trends & patterns.

### SHARE

Presenting findings to decision-makers through reports, presentation or visualisations. Using tools like Tableau, Python(seaborn), R, to make charts.

### ACT

Putting insights from data into action such as implementing new business strategy.

# FOUNDATIONS: DATA, DATA, EVERYWHERE

## INTRODUCING DATA ANALYTICS AND ANALYTICAL THINKING

### ITERATIVE PROCESS

The data analysis process is designed to build on itself, so the results from each step are the inputs for the next step. However, you might not always move through the steps linearly. Depending on the circumstance you may have to go back to an earlier stage and work through the process with new, better information. For example, if you're in analyze phase and realise you're working with a wrong or incomplete dataset or you could learn while cleaning data that your initial question didn't accurately define the problem.

### TRANSFORM DATA INTO INSIGHTS

### DATA VS GUT INSTINCT

Data analysts and detectives share a similar approach to problem-solving, both relying on evidence and facts to make decisions. Gut instinct is an intuitive understanding with little to no explanation, and can cause biased narratives and judgement. Data-driven decision-making is essential for analysts, but gut instinct can also play a role in identifying patterns and connections. Balancing data and gut instinct is crucial for making informed decisions, and the right mix depends on the project's goals and time constraints.

# FOUNDATIONS: DATA, DATA, EVERYWHERE

## WONDERFUL WORLD OF DATA

### PLAN

Decide what kind of data is needed, how it will be managed, and who will be responsible for it.

### CAPTURE

Collect or bring in data from a variety of different sources

### MANAGE

Care for and maintain the data. This includes determining how and where it is stored and the tools used to do so.

## FOLLOW THE DATA LIFE CYCLE

## 6 STAGES OF DATA LIFECYCLE

### ANALYSE

Use the data to solve problems, make decisions, and support business goals

### ARCHIVE

Keep relevant data stored for long-term and future reference

### DESTROY

Remove data from storage and delete any shared copies of the data

# FOUNDATIONS: DATA, DATA, EVERYWHERE

## WONDERFUL WORLD OF DATA

### SPREADSHEET

Two popular spreadsheet applications you will probably use a lot in your future role as a data analyst are Microsoft Excel and Google Sheets. They let you collect, store, organize, and sort information. Also, you identify patterns and piece the data together in a way that works for each specific data project. Lastly, Create excellent data visualizations, like graphs and charts.

### THE DATA ANALYSIS TOOLBOX

### DATABASE AND QUERY LANGUAGES

Some popular Structured Query Language (SQL) programs include MySQL, Microsoft SQL Server, and BigQuery. Query languages allow analysts to isolate specific information from a database(s) Make it easier for you to learn and understand the requests made to databases.

### VISUALISATION TOOLS

Data analysts use a number of visualization tools, like graphs, maps, tables, charts, and more. Two popular visualization tools are Tableau and Looker. These tools turn complex numbers into a story that people can understand. They help stakeholders come up with conclusions that lead to informed decisions and effective business strategies.

**Further Comment:** the career as a data analyst also involves using programming languages, like R and Python, which are used a lot for statistical analysis, visualization, and other data analysis. You will usually have to decide which program or solution is right for the particular project you are working on.

# FOUNDATIONS: DATA, DATA, EVERYWHERE

## SET UP YOUR DATA ANALYST TOOLBOX

In the spirit of lifelong learning, it is good to have resources to turn to when you want to know more about using spreadsheets and queries as listed below

ITEM	Link
Google sheets training and help	<a href="https://support.google.com/a/users/answer/9282959?visit_id=637361702049227170-1815413770&amp;rd=1">https://support.google.com/a/users/answer/9282959?visit_id=637361702049227170-1815413770&amp;rd=1</a>
Google sheets tips and tricks	<a href="https://support.google.com/a/users/answer/9300022">https://support.google.com/a/users/answer/9300022</a>
Microsoft excel for windows training	<a href="https://support.microsoft.com/en-us/excel">https://support.microsoft.com/en-us/excel</a>
W3Schools SQL Tutorials	<a href="https://www.sqltutorial.org/sql-cheat-sheet/">https://www.sqltutorial.org/sql-cheat-sheet/</a>
SQL Cheat Sheet	<a href="https://www.w3schools.com/sql/default.asp">https://www.w3schools.com/sql/default.asp</a>

MASTER  
SPREADSHEET  
BASICS



# FOUNDATIONS: DATA, DATA, EVERYWHERE

## SET UP YOUR DATA ANALYST TOOLBOX

### CAPITALISATION, INDENTATION, SEMI COLONS

You can write your SQL queries in all lowercase and don't have to worry about extra spaces between words. Using capitalization and indentation can help you read the information more easily.

```
1  SELECT field1
2  FROM table
3  WHERE field1 = condition;
```

## SQL POSSIBILITIES

### WHERE SELECT FROM

In the query shown by the top right, the SELECT clause identifies the column you want to pull data from by name, field1, and the FROM clause identifies the table where the column is located by name, table. the WHERE clause narrows your query so that the database returns only the data with an exact value match or the data that matches a certain condition that you want to satisfy

### COMMENTS & ALIAS

Comments are text placed between certain characters, /\* and \*/, or after two dashes --) .  
You can assign a new name or alias to the column or table names to make them easier to work with. This is done with a SQL AS clause.

```
1  ∨ SELECT
2  |   my_table_alias.actual_column_name AS my_column_alias
3  ∨ FROM
4  |   actual_table_name AS my_table_alias
```

# FOUNDATIONS: DATA, DATA, EVERYWHERE

## SET UP YOUR DATA ANALYST TOOLBOX

It is established that part of a data professional's responsibility is to make certain that their analysis is fair. Fairness means ensuring your analysis doesn't create or reinforce bias

### THE IMPORTANCE OF FAIR BUSINESS DECISION

Best practice	Explanation
Consider all of the available data	Often there will be data that isn't relevant to what you're focusing on or doesn't seem to align with your expectations. But you can't just ignore it; it's critical to consider all of the available data so that your analysis reflects the truth and not just your own expectations.
Identify surrounding factors	As you'll learn throughout these courses, context is key for you and your stakeholders to understand the final conclusions of any analysis. you also must understand surrounding factors that could influence the insights you're gaining.
Include self-reported data	Self-reporting is a data collection technique where participants provide information about themselves. Self-reported data can be a great way to introduce fairness in your data collection process
Use oversampling effectively	Oversampling is the process of increasing the sample size of nondominant groups in a population. This can help you better represent them and address imbalanced datasets.
Think about fairness from beginning to end	This means that data collection, cleaning, processing, and analysis are all performed with fairness in mind.



# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

## ASK EFFECTIVE QUESTIONS

**Ask:** Define the problem you are trying to solve

**Prepare:** Use business task to decide what metrics to measure

**Process:** Clean data requires getting rid of errors, inaccuracies or inconsistencies

**Analyze:** thinking analytically about data requires performing calculations, combining data from multiple sources, etc

**Share:** Summarising results with clear and enticing visuals using data via graphs or dashboards

**Act:** Providing stakeholders with recommendations based on findings so they can make data-driven decisions

6 steps helps in breaking projects into manageable tasks are considered as '**Structured thinking**'.

TAKE ACTION WITH DATA

# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

ASK EFFECTIVE QUESTIONS

MAKING PREDICTIONS

CATEGORISING THINGS

SPOTTING UNUSUAL  
THINGS

SOLVE PROBLEMS WITH  
DATA

6 COMMON PROBLEM  
TYPES

IDENTIFYING THEMES

DISCOVERING  
CONNECTIONS

FINDING PATTERNS

# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

## ASK EFFECTIVE QUESTIONS

### SPECIFIC

Is the question specific? Does it have context? Does it address the problem?

### MEASURABLE

Will the question give you answers you can measure?

### ACTION-ORIENTED

Will the answers provide information that helps you devise some type of plan?

### RELEVANT

Is the question about the particular problem you are trying to solve?

### TIME-BOUND

Are the answers relevant to the specific time being studied?

## CRAFT EFFECTIVE QUESTIONS

Questions should be open-ended. Avoid the following; leading questions, closed-ended questions, vague questions.

# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

## MAKE DATA-DRIVEN DECISIONS

### QUALITATIVE DATA TOOLS

Focus groups  
Social media text analysis  
In-person interviews

### QUANTITATIVE DATA TOOLS

Structured interviews  
Surveys  
Polls

UNDERSTAND THE  
POWER OF DATA

# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

## MAKE DATA-DRIVEN DECISIONS

The following summarises benefits of using dashboards for both data analysts and stakeholders.

### FOLLOW THE EVIDENCE

BENEFITS	FOR DATA ANALYSTS	FOR STAKEHOLDERS
Centralization	Share a single source of data with all stakeholders	Work with a comprehensive view of data, initiatives, objectives, projects, processes, and more
Visualisation	Show and update live, incoming data in real time*	Spot changing trends and patterns more quickly
Insightfulness	Pull relevant information from different datasets	Understand the story behind the numbers to keep track of goals and make data-driven decisions
Customisation	Create custom views dedicated to a specific person, project, or presentation of the data	Drill down to more specific areas of specialized interest or concern

# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

## MAKE DATA-DRIVEN DECISIONS

Advanced data analytics communication resources

### FOLLOW THE EVIDENCE

ITEM	LINK
This gallery is full of great examples that were created using real data;	<a href="https://www.tableau.com/data-insights/dashboard-showcase">https://www.tableau.com/data-insights/dashboard-showcase</a>
These are visualizations created by Tableau users and are a great way to learn more about how other data analysts are using data visualization tools	<a href="https://public.tableau.com/app/discover">https://public.tableau.com/app/discover</a>
Filters show certain data while hiding the rest of the data in a dashboard. This can be a big help to identify patterns while keeping the original data intact	<a href="https://help.tableau.com/current/pro/desktop/en-us/actions_filter.htm">https://help.tableau.com/current/pro/desktop/en-us/actions_filter.htm</a>



# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

CONNECT THE DATA  
DOTS

SMALL DATA		BIG DATA	
Describes a dataset made up of specific metrics over a short, well-defined time period		Describes large, less-specific datasets 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	
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

# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

## SPREADSHEET MAGIC

### WORK WITH SPREADSHEETS

#### 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.

#### CAPTURE

Capture data by the source by connecting spreadsheets to other data sources, such as an online survey application or a database.

#### MANAGE

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

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

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

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.

# ASK QUESTIONS TO MAKE DATA-DRIVEN DECISIONS

## ALWAYS REMEMBER THE STAKEHOLDERS

### EXECUTIVE TEAM

The executive team might include vice presidents, the chief marketing officer, and senior-level professionals who help plan and direct the company's work. These stakeholders think about decisions at a very high level and they are looking for the major themes about your project first. They are less interested in the details. Time is very limited with them, so make the most of it by leading your presentations with the answers to their questions.

### BALANCE TEAM AND STAKEHOLDERS NEEDS

### CUSTOMER FACING TEAM

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. It is important to let the data tell the story and not be swayed by asks from your stakeholders to find certain patterns that might not exist.

### DATA SCIENCE TEAM

Organizing data within a company takes teamwork. There's a good chance you'll find yourself working with other data analysts, data scientists, and data engineers. A big part of your job will be collaborating with other data team members to find new angles of the data to explore.

# PREPARE DATA FOR EXPLORATION

## DATA TYPES AND STRUCTURES

### DATA SOURCES

Second party data is collected directly by another group and then sold.

Third party data is sold by a provider that didn't collect the data themselves.

### SOLVING BUSINESS PROBLEMS

Choose data that solve your problem question.

### HOW MUCH DATA TO COLLECT

Reasonable decision about sample size. Strategic collection for certain criteria.

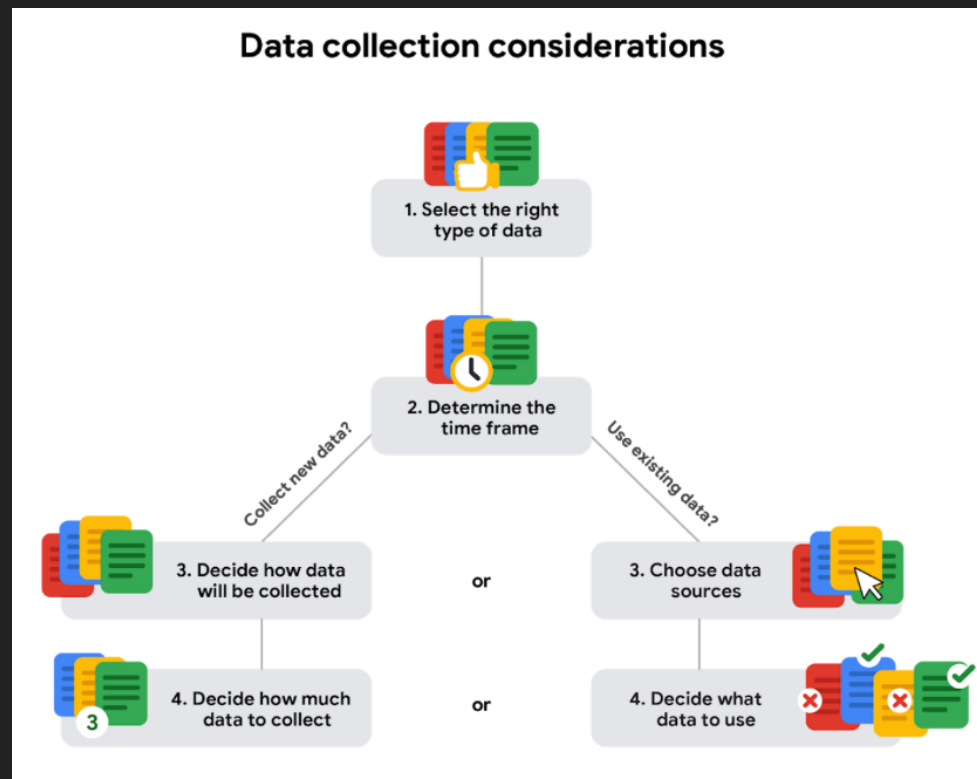
### TIME FRAME

Factor if you are tracking trends over a long time periods.

## COLLECT DATA

# PREPARE DATA FOR EXPLORATION

Flow chart for data collection if time consideration is vital



# PREPARE DATA FOR EXPLORATION

## DATA TYPES AND STRUCTURES

### PRIMARY DATA VS SECONDARY DATA

Collected by a researcher from first hand sources  
**vs** Gathered by other people or from other research

### INTERNAL DATA VS EXTERNAL DATA

Data that is stored inside a company's own systems  
**vs** Data that is stored outside of a company or organization

### CONTINUOUS DATA VS DISCRETE DATA

Data that is measured and can have almost any numeric value  
**vs** Data that is counted and has a limited number of values

DIFFERENTIATE DATA  
FORMATS AND  
STRUCTURES



# PREPARE DATA FOR EXPLORATION

## DATA TYPES AND STRUCTURES

### QUALITATIVE DATA VS QUANTITATIVE DATA

A subjective and explanatory measure of a quality or characteristic vs A specific and objective measure such as number, quantity or range

## DIFFERENTIATE DATA FORMATS AND STRUCTURES

### NOMINAL DATA VS ORDINAL DATA

A type of qualitative data without a set order vs A type of qualitative data with a set order or scale

### STRUCTURED DATA VS UNSTRUCTURED DATA

Data organised in certain formats like rows and columns vs Data that cannot be stored as rows and columns in a relational database

# PREPARE DATA FOR EXPLORATION

## DATA TYPES AND STRUCTURES

### STRUCTURED DATA

Organized in a certain format, such as rows and columns.

### UNSTRUCTURED DATA

Not organized in any easy-to-identify way.

## CHARACTERISTICS:

### DIFFERENTIATE DATA FORMATS AND STRUCTURES

- Defined data type
- Most often quantitative data
- Easy to organise
- Easy to search
- Easy to analyse
- Stored in relational databases and warehouses
- Contained in rows and columns
- Examples: Google sheets, Excel, SQL, customer data, phone records, transaction history

- Varied data types
- Most often qualitative data
- Difficult to search
- Provides more freedom for analysis
- Stored in data lakes, data warehouses, and NoSQL databases
- Cannot be put in rows and columns
- Examples: Text messages, social media comments, phone call transcriptions, various log files, images, audio, video

# PREPARE DATA FOR EXPLORATION

## DATA TYPES AND STRUCTURES

### DATA MODELLING

Data modeling is the process of creating diagrams that visually represent how data is organized and structured. These visual representations are called data models.

### CONCEPTUAL DATA MODELLING

Conceptual data modeling gives a high-level view of the data structure, such as how data interacts across an organization. For example, a conceptual data model may be used to define the business requirements for a new database

### LOGICAL DATA MODELLING

Logical data modeling focuses on the technical details of a database such as relationships, attributes, and entities. For example, a logical data model defines how individual records are uniquely identified in a database.

### PHYSICAL DATA MODELLING

Physical data modeling depicts how a database operates. A physical data model defines all entities and attributes used; for example, it includes table names, column names, and data types for the database.

## DIFFERENTIATE DATA FORMATS AND STRUCTURES

# PREPARE DATA FOR EXPLORATION

## DATA TYPES AND STRUCTURES

### DATA TRANSFORMATION

**Data transformation** is the process of changing the data's format, structure, or values. **Long data** is data where each row contains a single data point for a particular item. **Wide data** is data where each row contains multiple data points for the particular items identified in the columns.

EXPLORE DATA TYPES,  
FIELDS AND VALUES

#### GOALS OF DATA TRANSFORMATION:

**Data organization:** better organized data is easier to use.

**Data compatibility:** different applications or systems can then use the same data.

**Data migration:** data with matching formats can be moved from one system to another.

**Data merging:** data with the same organization can be merged together.

**Data enhancement:** data can be displayed with more detailed fields.

**Data comparison:** apples-to-apples comparisons of the data can then be made.

# PREPARE DATA FOR EXPLORATION

## DATA RESPONSIBILITY

### DATA ANONYMIZATION

Data anonymization is the process of protecting people's private or sensitive data by eliminating that kind of information. Typically, data anonymization involves blanking, hashing, or masking personal information, often by using fixed-length codes to represent data columns, or hiding data with altered values.

**Data anonymization** is used in every industry but healthcare and finance are sectors where it's use is critical due to sensitive nature of the data and operations.

The following types of data are examples of data that are usually anonymized: Telephone numbers, Names, License plates and license numbers, Social security numbers, IP addresses, Medical records, Email addresses, Photographs, Account numbers.

## DATA ETHICS AND PRIVACY

# PREPARE DATA FOR EXPLORATION

## DATABASE ESSENTIALS

### RELATIONAL DATABASES

#### WORK WITH DATABASES

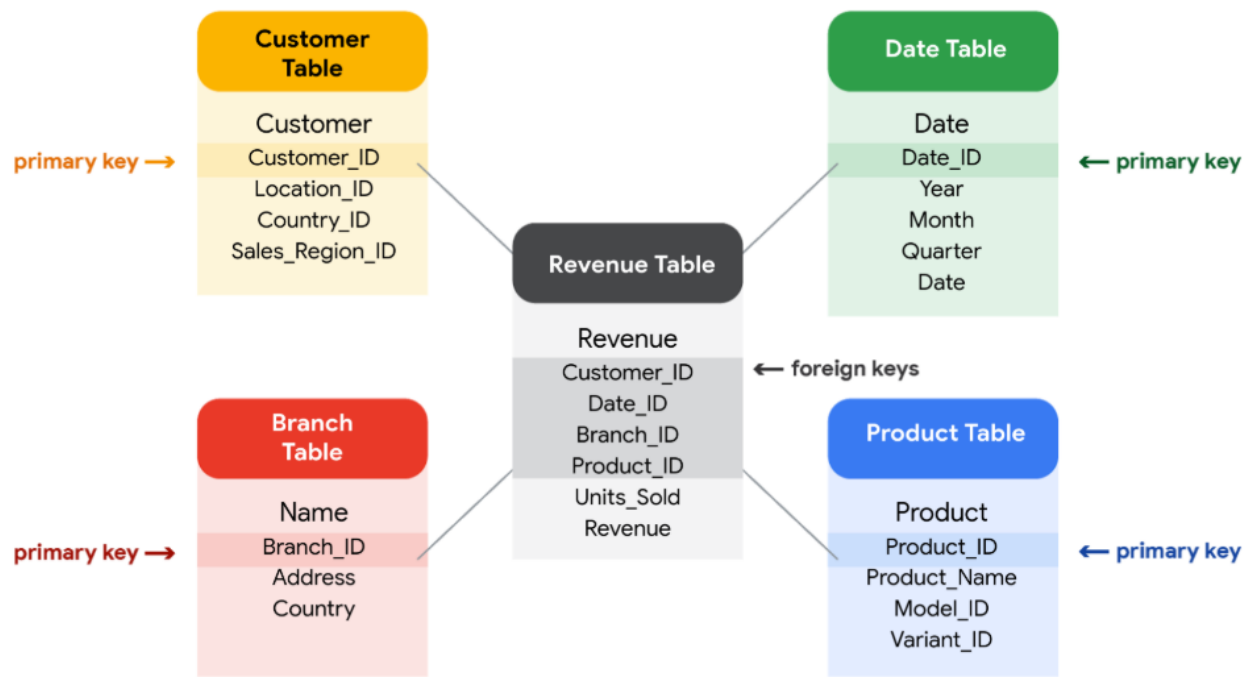
A **relational database** is a database that contains a series of tables that can be connected to form relationships. Basically, they allow data analysts to organize and link data based on what the data has in common.

**Normalization** is a process of organizing data in a relational database. For example, creating tables and establishing relationships between those tables. It is applied to eliminate data redundancy, increase data integrity, and reduce complexity in a database.



# PREPARE DATA FOR EXPLORATION

## RELATIONAL DATABASES



a **primary key** is an identifier that references a column in which each value is unique. In other words, it's a column of a table that is used to uniquely identify each record within that table. The value assigned to the primary key in a particular row must be unique within the entire table. For example, if `customer_id` is the primary key for the customer table, no two customers will ever have the same `customer_id`.

By contrast, a **foreign key** is a field within a table that is a primary key in another table. A table can have only one primary key, but it can have multiple foreign keys. These keys are what create the relationships between tables in a relational database, which helps organize and connect data across multiple tables in the database.

# PREPARE DATA FOR EXPLORATION

## DATABASE ESSENTIALS

**metadata** is data about data. In database management, metadata provides information about other data and helps data analysts interpret the contents of the data within a database.

**File or document type:** What type of file or document are you examining?

**Date, time, and creator:** When was it created? Who created it? When was it last modified?

**Title and description:** What is the name of the item you are examining? What type of content does it contain?

**Geolocation:** If you're examining a photo, where was it taken?

**Tags and categories:** What is the general overview of the item that you have? Is it indexed or described in a specific way?

**Who last modified it and when:** Were any changes made to the file? If yes, when were the most recent modifications made?

**Who can access or update it:** If you're examining a dataset, is it public? Are special permissions needed to customize or modify it?

MANAGE DATA WITH  
METADATA

# PREPARE DATA FOR EXPLORATION

## DATABASE ESSENTIALS

Open data helps create a lot of public datasets that you can access to make data-driven decisions. Here are some resources you can use to start searching for public datasets on your own.

ACCESS DIFFERENT  
DATA SOURCES

ITEM	LINK
The Google Cloud Public Dataset	<a href="https://cloud.google.com/datasets?hl=en">https://cloud.google.com/datasets?hl=en</a>
The dataset search	<a href="https://datasetsearch.research.google.com/">https://datasetsearch.research.google.com/</a>
BigQuery	<a href="https://cloud.google.com/bigquery/public-data">https://cloud.google.com/bigquery/public-data</a>
Global health observatory data	<a href="https://www.who.int/data/collections">https://www.who.int/data/collections</a>
1000 Genomes	<a href="https://cloud.google.com/life-sciences/docs/resources/public-datasets/1000-genomes">https://cloud.google.com/life-sciences/docs/resources/public-datasets/1000-genomes</a>
NOAA public dataset library	<a href="https://www.climate.gov/maps-data/all?listingMain=datasetgallery">https://www.climate.gov/maps-data/all?listingMain=datasetgallery</a>

# PREPARE DATA FOR EXPLORATION

## ORGANISE AND PROTECT DATA

### FILE ORGANISATION GUIDELINES:

**Name** Giving a file a meaningful name to describe its contents makes searching for it straightforward. It also makes it easy to understand the type of data the file contains.

**Creation date** Knowing when a file was created can help you understand if it is relevant to your current analysis. For example, you might want to analyze only data from 2023.

**Consistent order and style** Make sure the information you include in a file name follows a consistent order.

**Ensure team consistency** To ensure all team members use the agreed-upon file naming conventions, create a text file as a sample that includes all of the naming conventions on a project.

## BRING DATA TO ORDER

# PROCESS DATA FROM DIRTY TO CLEAN

## IMPORTANCE OF INTEGRITY

### Well-aligned objectives and data

You can gain powerful insights and make accurate conclusions when data is well-aligned to business objectives. As a data analyst, alignment is something you will need to judge.

### Data constraints for new variables

When there is clean data and good alignment, you can get accurate insights and make conclusions the data supports.

If there is good alignment but the data needs to be cleaned, clean the data before you perform your analysis.

If the data only partially aligns with an objective, think about how you could modify the objective, or use data constraints to make sure that the subset of data better aligns with the business objective.

## DATA INTEGRITY AND ANALYTICS OBJECTIVES

# PROCESS DATA FROM DIRTY TO CLEAN

## IMPORTANCE OF INTEGRITY

### WHEN YOU FIND ISSUES WITH YOUR DATA:

#### DATA ISSUE 1: NO DATA

Gather the data on a small scale to perform a preliminary analysis and then request additional time to complete the analysis after you have collected more data. If there isn't time to collect data, perform the analysis using proxy data from other datasets. This is the most common workaround.

#### DATA ISSUE 2: TOO LITTLE DATA

Do the analysis using proxy data along with actual data. Adjust your analysis to align with the data you already have.

#### DATA ISSUE 3: WRONG DATA, INCLUDING DATA WITH ERROR?

If you have the wrong data because requirements were misunderstood, communicate the requirements again. Identify errors in the data and, if possible, correct them at the source by looking for a pattern in the errors.

### OVERCOME ISSUES OF INSUFFICIENT DATA



# PROCESS DATA FROM DIRTY TO CLEAN

OVERCOME CHALLENGES  
OF INSUFFICIENT DATA

## IMPORTANCE OF INTEGRITY

### TERMINOLOGIES RELATED TO SAMPLE SIZE:

#### POPULATION

The entire group that you are interested in for your study.

#### SAMPLE

A subset of your population.

#### MARGIN OF ERROR

Since a sample is used to represent a population, the sample's results are expected to differ from what the result would have been if you had surveyed the entire population. This difference is called the margin of error.

#### CONFIDENCE LEVEL

How confident you are in the survey results. For example, a 95% confidence level means that if you were to run the same survey 100 times, you would get similar results 95 of those 100 times.

#### CONFIDENCE INTERVAL

The range of possible values that the population's result would be at the confidence level of the study. This range is the sample result +/- the margin of error.

#### STATISTICAL SIGNIFICANCE

The determination of whether your result could be due to random chance or not. The greater the significance, the less due to chance.

# PROCESS DATA FROM DIRTY TO CLEAN

## CLEAN DATA FOR MORE ACCURATE INSIGHTS

**Dirty data** is data that is incomplete, incorrect, or irrelevant to the problem you are trying to solve.

Any data that is old which should be replaced with newer and more accurate information

DATA CLEANING IS A  
MUST

Any data record that shows up more than once

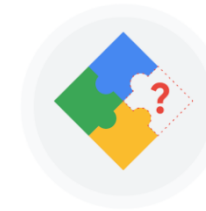
Types of dirty data



Duplicate data



Outdated data



Incomplete data

Any data that is missing important fields

Any data that is complete but inaccurate



Incorrect/inaccurate data



Inconsistent data

Any data that uses different formats to represent the same thing

# PROCESS DATA FROM DIRTY TO CLEAN

CLEAN DATA FOR MORE ACCURATE  
INSIGHTS

COMMON DATA CLEANING PITFALLS:



Not checking for  
spelling errors



Forgetting  
to document  
errors



Not checking for  
misfielded  
values



Overlooking  
missing values



Looking at  
a subset of data  
and not the whole  
picture



Losing track  
of the business  
objectives



Not fixing  
the source  
of the error



Not analyzing the  
system prior to  
data cleaning



Not backing up  
your data prior  
to data  
cleansing



Not accounting  
for data  
cleaning in your  
deadlines/process

FIRST STEPS  
TOWARDS CLEAN  
DATA

# PROCESS DATA FROM DIRTY TO CLEAN

## CLEAN DATA FOR MORE ACCURATE INSIGHTS

**Workflow automation** is the process of automating parts of your work. That could mean creating an event trigger that sends a notification when a system is updated. Or it could mean automating parts of the data cleaning process. As you can probably imagine, automating different parts of your work can save you tons of time, increase productivity, and give you more bandwidth to focus on other important aspects of the job.

TASK	CAN IT BE AUTOMATED?	WHY?
Communicating with your team and stakeholders	No	Communication is key to understanding the needs of your team and stakeholders as you complete the tasks you are working on. There is no replacement for person-to-person communications.
Presenting your findings	No	Presenting your data is a big part of your job as a data analyst. Making data accessible and understandable to stakeholders and creating data visualizations can't be automated for the same reasons that communications can't be automated.
Preparing and cleaning data	Partially	Some tasks in data preparation and cleaning can be automated by setting up specific processes, like using a programming script to automatically detect missing values.
Modelling the data	yes	Data modelling is a difficult process that involves lots of different factors; luckily there are tools that can completely automate the different stages.

CONTINUE  
CLEANING DATA IN  
SPREADSHEET

# PROCESS DATA FROM DIRTY TO CLEAN

## DATA CLEANING WITH SQL

As a data analyst, you will be tasked with handling a lot of data, and SQL is one of the tools that can help make your work a lot easier. SQL is the primary way data analysts extract data from databases.

### SQL FOR CLEAN DATA

FEATURES OF SPREADSHEETS	FEATURES OF SQL DATABASES
Smaller data sets	Larger datasets
Enter data manually	Access tables across a database
Create graphs and visualizations in the same program	Prepare data for further analysis in another software
Built-in spell check and other useful functions	Fast and powerful functionality
Best when working solo on a project Great for collaborative work and tracking queries run by all users	Great for collaborative work and tracking queries run by all users

# PROCESS DATA FROM DIRTY TO CLEAN

## VERIFY AND REPORT ON CLEANING RESULTS

Make sure you identified the most common problems and corrected them, including:

**Sources of errors:** Did you use the right tools and functions to find the source of the errors in your dataset?

**Null data:** Did you search for NULLs using conditional formatting and filters?

**Misspelled words:** Did you locate all misspellings?

**Mistyped numbers:** Did you double-check that your numeric data has been entered correctly?

**Extra spaces and characters:** Did you remove any extra spaces or characters using the TRIM function?

**Duplicates:** Did you remove duplicates in spreadsheets using the Remove Duplicates function or DISTINCT in SQL?

**Mismatched data types:** Did you check that numeric, date, and string data are typecast correctly?

**Messy (inconsistent) strings:** Did you make sure that all of your strings are consistent and meaningful?

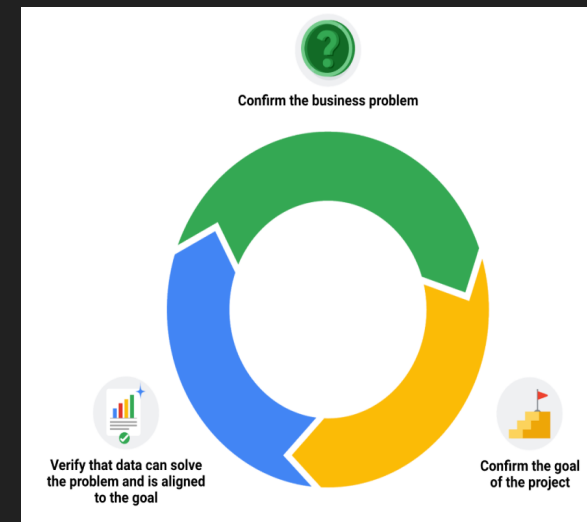
**Messy (inconsistent) date formats:** Did you format the dates consistently throughout your dataset?

**Misleading variable labels (columns):** Did you name your columns meaningfully?

**Truncated data:** Did you check for truncated or missing data that needs correction?

**Business Logic:** Did you check that the data makes sense given your knowledge of the business?

## MANUALLY CLEANING DATA



# PROCESS DATA FROM DIRTY TO CLEAN

## VERIFY AND REPORT ON CLEANING RESULTS

Data analysts use **changelogs** to keep track of data transformation and cleaning. A changelog can build on your automated version history by giving you an even more detailed record of your work. This is where data analysts record all the changes they make to the data.

Typically, a changelog records:

- Data, file, formula, query, or any other component that changed
- Description of what changed
- Date of the change
- Person who made the change
- Person who approved the change
- Version number
- Reason for the change

## DOCUMENT THE CLEANING PROCESS



# ANALYSE DATA TO ANSWER QUESTIONS

## FORMAT AND ADJUST DATA

### SORT DATA

Sorting is the process of arranging data into a meaningful order to make it easier to understand, analyze, and visualize. It ranks your data based on a specific metric you choose.

## ORGANISE DATA FOR ANALYSIS

### FILTER DATA

Sometimes, an analysis may require only a subset of the data in your dataset. You can use a filter to show only the data that meets a specified criteria while hiding the rest. Filtering is useful when you have lots of data.

The first two phases of data analysis, **Organize data** and **Format and adjust data**, are important for data analysts because they can use these phases to manipulate their data in ways that make important patterns and trends more obvious.

# ANALYSE DATA TO ANSWER QUESTIONS

## ORGANISE DATA FOR MORE EFFECTIVE ANALYSIS

### SORT DATASETS IN SPREADSHEETS

### ORGANISE DATA FOR ANALYSIS

**Sort a sheet with the menu** → Select Data from the menu. Select Sort sheet.

**Sort data in a specific column with the menu** → Select cell A to highlight all of column A. Select Data from the menu. Select Sort range. Select Sort range by column A (A to Z).

`SORT(array, [sort_index], [sort_order], [by_col])`

`FILTER(array, include, [if_empty])`

**Sorting tips:** <https://support.microsoft.com/en-us/office/sort-data-in-a-range-or-table-62d0b95d-2a90-4610-a6ae-2e545c4a4654>

**Sort and filter tips:** <https://support.google.com/docs/answer/3540681>  
<https://www.sheetgo.com/blog/google-sheets-formulas/sort-formula-google-sheets/>

# ANALYSE DATA TO ANSWER QUESTIONS

## ORGANISE DATA FOR MORE EFFECTIVE ANALYSIS

### Sort data by one column

The ORDER BY command sorts data by column in a database

### Sort data in descending order

to use the ORDER BY command to sort data by descending order, specify DESC at the end of the ORDER BY command.

### Filter and sort data in descending order

Use WHERE and ORDER BY together to filter, then sort, data.

### Filter on two conditions, then sort data in descending order:

Use WHERE, AND, and ORDER BY to filter data on two conditions and then sort it.

## SORT DATA USING SQL

# ANALYSE DATA TO ANSWER QUESTIONS

## FORMAT AND ADJUST DATA

### CONVERT DATA IN SPREADSHEETS

As a data analyst, there are lots of scenarios when you might need to convert data in a spreadsheet.

#### Conversion scenarios:

- String to date
- String to numbers
- Combining columns
- Number to percentage

### CONVERT DATA IN SQL

#### Conversion scenarios:

CAST FUNCTION

Numeric to Integer, Floating integer, String  
String to boolean, Integer, Bytes, Date, Time, Timestamp  
Date to strings, Datetime

*And others.*

**Further resources: CAST & Convert:** <https://learn.microsoft.com/en-us/sql/t-sql/functions/cast-and-convert-transact-sql?view=sql-server-ver15>

**MySQL CAST FUNCTIONS & Operators:** <https://dev.mysql.com/doc/refman/8.0/en/cast-functions.html>

## FORMATTING FOR BETTER ANALYSIS

# ANALYSE DATA TO ANSWER QUESTIONS

## FORMAT AND ADJUST DATA

As a data analyst, there are many occasions where you will need to import and combine data from one file or location to another.

### COMBINING MULTIPLE DATASETS

#### IMPORT & COMBINE DATA IN SQL

Import INTO with SELECT Statement

INSERT INTO along with the SELECT and WHERE commands

CONCAT FUNCTION

#### MERGE TEXT STRINGS IN SPREADSHEET

The LEN function

the FIND function

The RIGHT function

The LEFT function

#### MANIPULATE STRINGS IN SQL

CONCAT function

CONCAT\_WS function

|| function

Example:

```
1 SELECT CONCAT_WS('.', 'www', 'your_company', 'com') as website FROM web_data;
```

# ANALYSE DATA TO ANSWER QUESTIONS

ITEM	Link
Keyboard shortcuts for Google sheets	<a href="https://support.google.com/docs/answer/181110#zippy=%2Cpc-shortcuts">https://support.google.com/docs/answer/181110#zippy=%2Cpc-shortcuts</a>
List of Google sheets functions	<a href="https://support.google.com/docs/table/25273?hl=en">https://support.google.com/docs/table/25273?hl=en</a>
23 Must-know Google Sheet Formulas	<a href="https://golayer.io/blog/google-sheets/google-sheets-formulas">https://golayer.io/blog/google-sheets/google-sheets-formulas</a>
18 Google sheets formulas tips and techniques	<a href="https://www.benlcollins.com/spreadsheets/google-sheets-formulas-techniques/">https://www.benlcollins.com/spreadsheets/google-sheets-formulas-techniques/</a>
Keyboard shortcuts in Excel	<a href="https://support.microsoft.com/en-us/office/keyboard-shortcuts-in-excel-1798d9d5-842a-42b8-9c99-9b7213f0040f?ui=en-US&amp;rs=en-US&amp;ad=US">https://support.microsoft.com/en-us/office/keyboard-shortcuts-in-excel-1798d9d5-842a-42b8-9c99-9b7213f0040f?ui=en-US&amp;rs=en-US&amp;ad=US</a>
222 Excel shortcuts	<a href="https://exceljet.net/shortcuts">https://exceljet.net/shortcuts</a>
List of spreadsheet functions	<a href="https://exceljet.net/functions">https://exceljet.net/functions</a>
List of spreadsheet formulas	<a href="https://exceljet.net/formulas">https://exceljet.net/formulas</a>
Essential Excel Skills for Analyzing data	<a href="https://learntocodewith.me/posts/excel-skills/">https://learntocodewith.me/posts/excel-skills/</a>
Advanced spreadsheet skills	<a href="https://www.slideshare.net/slideshow/advanced-spreadsheet-skills/84038471">https://www.slideshare.net/slideshow/advanced-spreadsheet-skills/84038471</a>

# ANALYSE DATA TO ANSWER QUESTIONS

## AGGREGATE DATA FOR ANALYSIS

### VLOOKUP

Syntax: VLOOKUP(search\_key, range, index, is\_sorted)

### VLOOKUP & DATA AGGREGATION

ITEM	Link
How to use VLOOKUP in Excel	<a href="https://support.microsoft.com/en-us/office/vlookup-function-0bbc8083-26fe-4963-8ab8-93a18ad188a1">https://support.microsoft.com/en-us/office/vlookup-function-0bbc8083-26fe-4963-8ab8-93a18ad188a1</a>
VLOOKUP in Excel tutorials	<a href="https://www.youtube.com/watch?v=d3BYVQ6xIE4&amp;ab_channel=ExcelCampus-Jon">https://www.youtube.com/watch?v=d3BYVQ6xIE4&amp;ab_channel=ExcelCampus-Jon</a>
23 you should know about VLOOKUP in Excel	<a href="https://exceljet.net/articles/things-you-should-know-about-vlookup">https://exceljet.net/articles/things-you-should-know-about-vlookup</a>
How to use Excel's VLOOKUP function	<a href="https://edu.gcfglobal.org/en/excel-tips/how-to-use-excels-vlookup-function/1/">https://edu.gcfglobal.org/en/excel-tips/how-to-use-excels-vlookup-function/1/</a>
VLOOKUP in Excel vs Google Sheet	<a href="https://infoinspired.com/sheets-vs-excel-formula/vlookup-formula-in-excel-and-google-sheets/">https://infoinspired.com/sheets-vs-excel-formula/vlookup-formula-in-excel-and-google-sheets/</a>

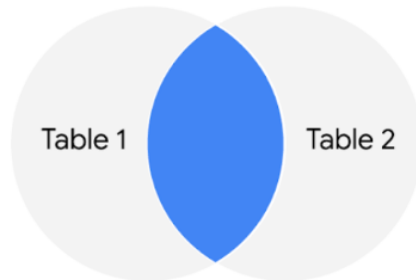


# ANALYSE DATA TO ANSWER QUESTIONS

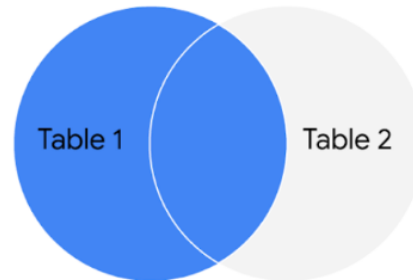
## AGGREGATE DATA FOR ANALYSIS

USE JOINS TO  
AGGREGATE DATA IN  
SQL

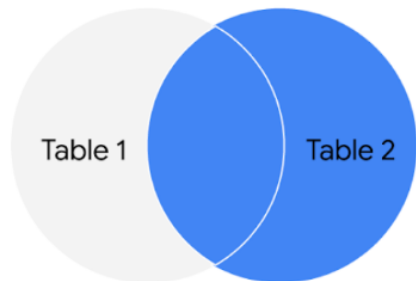
INNER JOIN



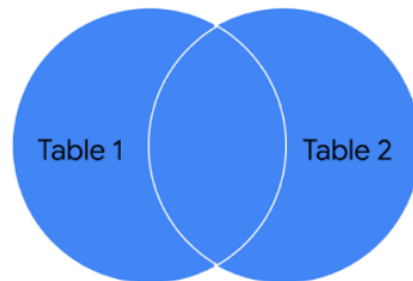
LEFT JOIN



RIGHT JOIN



FULL OUTER JOIN



# ANALYSE DATA TO ANSWER QUESTIONS

AGGREGATE DATA FOR ANALYSIS

USE JOINS TO  
AGGREGATE DATA IN  
SQL

ITEM	Link
SQL JOINS	<a href="https://www.w3schools.com/sql/sql_join.asp">https://www.w3schools.com/sql/sql_join.asp</a>
Database JOINS	<a href="https://www.essentialsql.com/sql-joins/">https://www.essentialsql.com/sql-joins/</a>
SQL JOIN Types Explained in Visuals	<a href="https://dataschool.com/how-to-teach-people-sql/sql-join-types-explained-visually/">https://dataschool.com/how-to-teach-people-sql/sql-join-types-explained-visually/</a>
SQL JOINS: Bringing data together one JOIN at a time	<a href="https://towardsdatascience.com/sql-join-8212e3eb9fde">https://towardsdatascience.com/sql-join-8212e3eb9fde</a>
SQL JOIN	<a href="https://www.dofactory.com/sql/join">https://www.dofactory.com/sql/join</a>

# ANALYSE DATA TO ANSWER QUESTIONS

## PERFORM DATA CALCULATIONS

### SUMIF to SUMIFS

Syntax; =SUMIFS(sum\_range, criteria\_range1, criterion1, [criterion\_range2, criterion2,...])

### COUNTIF to COUNTIFS

Syntax; =COUNTIFS(sum\_range, criteria\_range1, criterion1, [criterion\_range2, criterion2,...])

GET STARTED WITH  
DATA CALCULATIONS

ITEM	Link
How to use Excel IFS function	<a href="https://exceljet.net/functions/ifs-function">https://exceljet.net/functions/ifs-function</a>
VLOOKUP in Excel with multiple criteria	<a href="https://exceljet.net/formulas/vlookup-with-multiple-criteria">https://exceljet.net/formulas/vlookup-with-multiple-criteria</a>
Index and MATCH in Excel with multiple criteria	<a href="https://exceljet.net/formulas/index-and-match-with-multiple-criteria">https://exceljet.net/formulas/index-and-match-with-multiple-criteria</a>
Using IF with AND, OR, and NOT functions in Excel	<a href="https://support.microsoft.com/en-us/office/using-if-with-and-or-and-not-functions-in-excel-d895f58c-b36c-419e-b1f2-5c193a236d97">https://support.microsoft.com/en-us/office/using-if-with-and-or-and-not-functions-in-excel-d895f58c-b36c-419e-b1f2-5c193a236d97</a>

# SHARE DATA THROUGH THE ART OF VISUALIZATION

## VISUALISE DATA

Framework for organising thoughts about visualisation - The David McCandless Method

### UNDERSTAND DATA VISUALISATIONS

#### INFORMATION

The data with which you're working

#### STORY

A clear and compelling narrative or concept

#### GOAL

A specific objective or function for the visual

#### VISUAL FORM

An effective use of metaphors or visual expression

# SHARE DATA THROUGH THE ART OF VISUALIZATION

## VISUALISE DATA

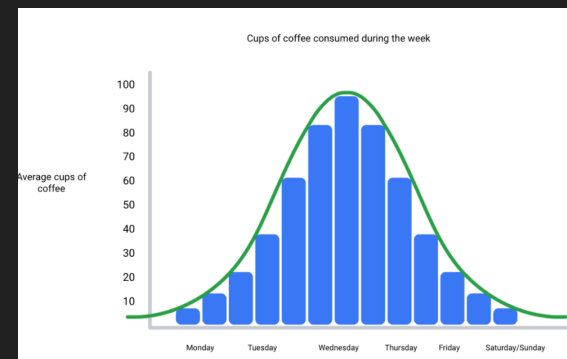
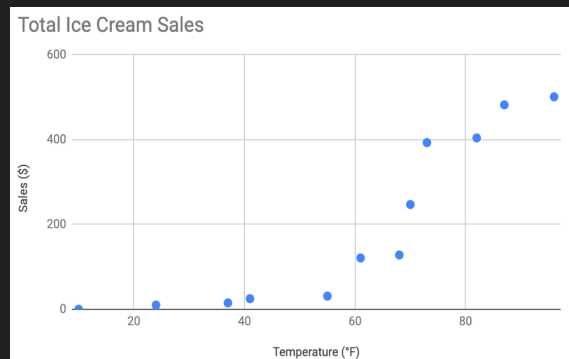
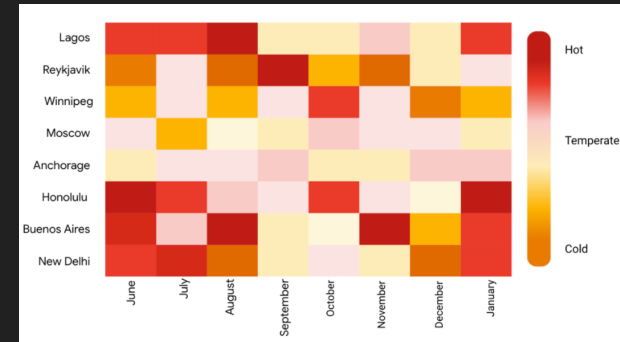
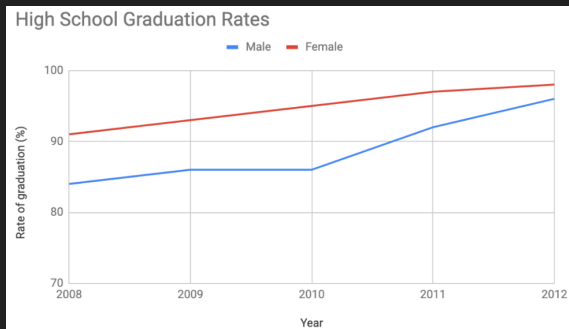
You will find that organizing your data and communicating your results are significant parts of a data analyst's role

### UNDERSTAND DATA VISUALISATIONS

ITEM	Link
The data visualisation catalogue	<a href="https://datavizcatalogue.com/#google_vignette">https://datavizcatalogue.com/#google_vignette</a>
The 25 best data visualisations	<a href="https://visme.co/blog/best-data-visualizations/">https://visme.co/blog/best-data-visualizations/</a>
10 data visualisation blogs	<a href="https://www.tableau.com/learn/articles/best-data-visualization-blogs">https://www.tableau.com/learn/articles/best-data-visualization-blogs</a>
Information is beautiful	<a href="https://informationisbeautiful.net/wdvp/gallery-2019/">https://informationisbeautiful.net/wdvp/gallery-2019/</a>
Data Studio Gallery	<a href="https://lookerstudio.google.com/gallery?category=visualization">https://lookerstudio.google.com/gallery?category=visualization</a>

# SHARE DATA THROUGH THE ART OF VISUALIZATION

## VISUALISE DATA



# SHARE DATA THROUGH THE ART OF VISUALIZATION

## VISUALISE DATA

Reviewing each of these visual examples, where do you notice that they fit in relation to your type of data? One way to answer this is by evaluating patterns in data. Meaningful patterns can take many forms, such as:

**Change:** This is a trend or instance of observations that become different over time. A great way to measure change in data is through a line or column chart.

**Clustering:** A collection of data points with similar or different values. This is best represented through a distribution graph.

**Relativity:** These are observations considered in relation or in proportion to something else. You have probably seen examples of relativity data in a pie chart.

**Ranking:** This is a position in a scale of achievement or status. Data that requires ranking is best represented by a column chart.

**Correlation:** This shows a mutual relationship or connection between two or more things. A scatterplot is an excellent way to represent this type of data pattern.

## UNDERSTAND DATA VISUALISATIONS



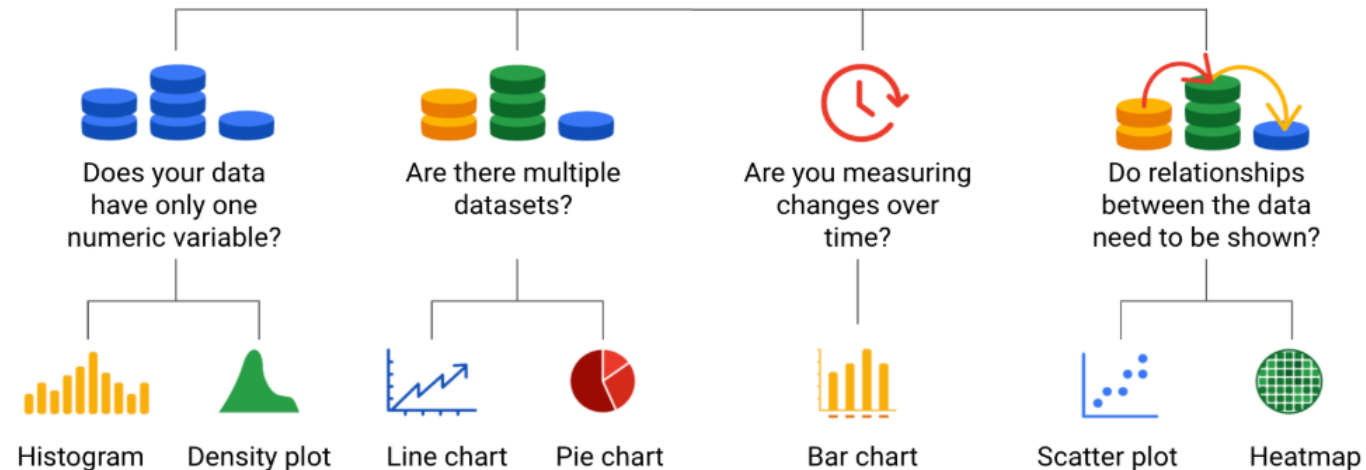
# SHARE DATA THROUGH THE ART OF VISUALIZATION

## VISUALISE DATA

### UNDERSTAND DATA VISUALISATIONS

### Decision tree example

Which story would you like to tell?



# SHARE DATA THROUGH THE ART OF VISUALIZATION

## VISUALISE DATA

### DESIGN DATA VISUALISATIONS

There are nine basic **principles of design** that data analysts should think about when building their visualizations.



Balance



Emphasis



Movement



Pattern



Repetition



Proportion



Rhythm



Variety



Unity

# SHARE DATA THROUGH THE ART OF VISUALIZATION

CREATE DATA VISUALISATIONS WITH  
TABLEAU

GET STARTED WITH  
TABLEAU

ITEM	Link
The ultimate cheat sheet on Tableau	<a href="https://medium.com/data-science/the-ultimate-cheat-sheet-on-tableau-charts-642bca94dde5">https://medium.com/data-science/the-ultimate-cheat-sheet-on-tableau-charts-642bca94dde5</a>
Which chart is right for you?	<a href="https://www.tableau.com/sites/default/files/media/which_chart_v6_final_0.pdf">https://www.tableau.com/sites/default/files/media/which_chart_v6_final_0.pdf</a>
Heat maps	<a href="https://help.tableau.com/current/pro/desktop/en-us/buildexamples_highlight.htm">https://help.tableau.com/current/pro/desktop/en-us/buildexamples_highlight.htm</a>
Highlight tables	<a href="https://help.tableau.com/current/pro/desktop/en-us/buildexamples_highlight.htm">https://help.tableau.com/current/pro/desktop/en-us/buildexamples_highlight.htm</a>
Packed bubble charts	<a href="https://help.tableau.com/current/pro/desktop/en-us/buildexamples_bubbles.htm">https://help.tableau.com/current/pro/desktop/en-us/buildexamples_bubbles.htm</a>
Filled Maps	<a href="https://interworks.com/blog/2014/09/23/tableau-essentials-chart-types-filled-map/">https://interworks.com/blog/2014/09/23/tableau-essentials-chart-types-filled-map/</a>
Circle views	<a href="https://interworks.com/blog/2014/10/17/tableau-essentials-chart-types-circle-view/">https://interworks.com/blog/2014/10/17/tableau-essentials-chart-types-circle-view/</a>

# SHARE DATA THROUGH THE ART OF VISUALIZATION

## CREATE DATA VISUALISATIONS IN TABLEAU

These design principles are important to your work as a data analyst because they help ensure that you're creating visualizations that communicate your data effectively to your audience.

### DESIGN VISUALISATIONS IN TABLEAU

PRINCIPLE	DESCRIPTION
Choose the right visual	One of the first things you have to decide is which visual will be the most effective for your audience.
Optimize the data-ink ratio	The data-ink entails focusing on the part of the visual that is essential to understanding the point of the chart.
Use orientation effectively	Make sure the written components of the visual, like the labels on a bar chart, are easy to read.
Colour	There are a lot of important considerations when thinking about using color in your visuals. These include using color consciously and meaningfully, staying consistent throughout your visuals, being considerate of what colors mean to different people, and using inclusive color scales that make sense for everyone viewing them.
Numbers of elements	Think about how many elements you include in any visual. If your visualization uses lines, try to plot five or fewer. If that isn't possible, use color or hue to emphasize important lines.

# SHARE DATA THROUGH THE ART OF VISUALIZATION

## CREATE DATA VISUALISATIONS IN TABLEAU

### STATIC DATA

involves providing screenshots or snapshots in presentations or building dashboards using snapshots of data. There are pros and cons to static data.

#### PROS

Can tightly control a point-in-time narrative of the data and insight.  
Allows for complex analysis to be explained in-depth to a larger audience.

#### CONS

Insight immediately begins to lose value and continues to do so the longer the data remains in a static state.  
Snapshots can't keep up with the pace of data change

### LIVE DATA

means that you can build dashboards, reports, and views connected to automatically updated data.

#### PROS

Dashboards can be built to be more dynamic and scalable.  
Gives the most up-to-date data to the people who need it at the time when they need it.  
Allows for up-to-date curated views into data with the ability to build a scalable "single source of truth" for various use cases.  
Allows for immediate action to be taken on data that changes frequently.  
Alleviates time/resources spent on processes for every analysis

#### CONS

Can take engineering resources to keep pipelines live and scalable, which may be outside the scope of some companies' data resource allocation.

Without the ability to interpret data, you can lose control of the narrative, which can cause data chaos (i.e. teams coming to conflicting conclusions based on the same data)

## DESIGN VISUALISATIONS IN TABLEAU

# SHARE DATA THROUGH THE ART OF VISUALIZATION

## CRAFT DATA STORIES

When you're checking over slide presentations, there are some best practices you can check for:

**Include a title, subtitle, and date:** Making sure that your slide deck presentation has a title, subtitle, and date makes sure that your audience knows exactly what you are presenting and when the information was from. That way they know it's relevant and current to them!

**Use a logical sequence of slides:** Organizing your slides in an order that makes sense guides your audience through your narrative, building understanding step by step.

**Provide an agenda with a timeline:** An agenda offers a roadmap of your presentation, allowing your audience to follow along and anticipate key topics.

**Limit the amount of text on slides:** Keeping text brief ensures clarity and retains the audience's attention; aim for your audience to scan it within 5 seconds.

**Start with the business task:** By immediately relating the content to the business task at hand, you contextualize your information, making it relevant and actionable.

**Establish the initial hypothesis:** Presenting an initial hypothesis gives your audience a starting point for what to expect and frames the subsequent analysis.

**Show what business metrics you used:** Clarifying which metrics you're analyzing validates your arguments and helps the audience gauge your presentation's relevance to business outcomes.

**Use visualizations:** Visual aids can illustrate complex data more effectively than text alone, making your message more accessible.

**Introduce the graphic by name:** A brief introduction to each graphic aids in understanding and retaining information.

**Provide a title for each graph:** Titles act as signposts, helping the audience quickly grasp the meaning of each visual.

**Go from the general to the specific:** Starting with a broad overview before diving into details ensures that all audience members are on the same page.

**Use speaker notes to help you remember talking points:** Notes act as your cue cards, enabling a smoother delivery and ensuring no critical point is missed.

**Include key takeaways:** Summarizing the main points at the end of your presentation reinforces the message and ensures the audience leaves with the intended takeaways.

## THE ART AND SCIENCE OF VISUALISATION



# LINKS TO RESOURCES ON R CODING

ITEM	Link
Detailed guide to R Markdown documents	<a href="https://bookdown.org/yihui/rmarkdown/documents.html">https://bookdown.org/yihui/rmarkdown/documents.html</a>
Notebooks for R markdown documents	<a href="https://rmarkdown.rstudio.com/lesson-10.html">https://rmarkdown.rstudio.com/lesson-10.html</a>
Presentations by R Markdown	<a href="https://rmarkdown.rstudio.com/lesson-11.html">https://rmarkdown.rstudio.com/lesson-11.html</a>
Dashboards	<a href="https://pkgs.rstudio.com/flexdashboard/">https://pkgs.rstudio.com/flexdashboard/</a>
Shiny for interactive web apps built in R	<a href="https://shiny.posit.co/r/getstarted/shiny-basics/lesson1/">https://shiny.posit.co/r/getstarted/shiny-basics/lesson1/</a>
Rstudio formats	<a href="https://rmarkdown.rstudio.com/formats.html">https://rmarkdown.rstudio.com/formats.html</a>
Bookdown package	<a href="https://github.com/rstudio/bookdown">https://github.com/rstudio/bookdown</a>
Prettydoc package	<a href="https://github.com/yixuan/prettydoc/">https://github.com/yixuan/prettydoc/</a>
Rticles package	<a href="https://github.com/rstudio/rticles">https://github.com/rstudio/rticles</a>
R Markdown Gallery	<a href="https://rmarkdown.rstudio.com/gallery.html">https://rmarkdown.rstudio.com/gallery.html</a>
R Markdown formats	<a href="https://r4ds.had.co.nz/r-markdown-formats.html">https://r4ds.had.co.nz/r-markdown-formats.html</a>



# ADDITIONAL REFERENCE LITERATURE ON DATA ANALYTICS/SCIENCE

Author	Title
Arthur Zhang	Data Analytics: Practical guide to Leveraging Power of Algorithm
Brian Godsey	Think Like A Data Scientist
Rachel Schutt	Doing Data Science
Mona khalil	Effective Data Analysis
Pablo Saenz de Tejada	Tableau Cookbook for Professionals
George Casella	Statistical Inference
Tarek A Atwan	Time Series Analysis with Python Cookbook
Joel Grus	Data Science from Scratch
Ayodele Oluleye	Exploratory Data Analysis with Python Cookbook
Wes McKinney	Python for Data Analysis

Author	Title
Alex J Gutman	Becoming A Data Head How to Think, Speak like and Understand Data Science and Machine Learning
Bernard Marr	Big Data in Practise How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results
Emily Robinson	Build a Career in Data Science
EMC Education Series	Data Science and Big Data Analytics: Discovering, Analyzing, Visualising and Presenting Results
Nicolas Vandeput	Demand Forecasting Best Practises
Thomas Nield	Essential Math for Data Science
Sheldon M Ross	Introduction to Probability Models
Peter Gedeck	Practical Statistics for Data Scientists
Aileen Nielsen	Practical Time Series Analysis
Sowmya Vajjala	Practical Natural Language Processing