* Accessing different data sources :→
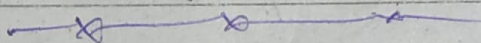
Internal Data!

Data that lives within company's own system.
Also called as <u>Primary Data</u>.

External Data:

Data that lives and is generated outside an organisation.
Also called as Secondary Data.

External Data is required when internal data doesn't give the full picture.

————————————

Importing data in spreadsheet :→

① Import from other sheet :

Google sheets :→

= ImportRange ("Address", "Sheet 1 | A1 : C10 ")

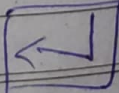② Importing cvs: Google Sheets → file > import > Select

③

=Import Data ("URL")   =ImportHTML ("URL", 1)

In Excel, Data > Get data > select   or
Data > from Text/CSV

Data > Get data > from other Sources
→ from web "URL" ⟵

* CSV: Comma Separated Values
1) CSV stores data in table format

Downloads a dataset from WHO website.

> Import into Sheets
> Gives title to the sheet.

—x—————x—————x—

week 4/5
3
14/01/23

Sorting Data:→
Arranging data into a meaningful order to make
it easier to understand, analyze, and visualize

ASC | DESC | Alphabetically / Numerically

Across all spreadsheet or just single column

Can sort by multiple variables : (first city, then state)

Anytime you're sorting, freeze & increase size of header Row
by   > View > freeze > 1 row

Anytime we scroll the header ! Row stays visible

Sort   Select column right click & sort in your choosen way!
                        or
                Click on ▽
                Drop Down

Sort Range : Select desired range > Data > Sort Range

Columns will be sorted in order of $L \rightarrow R$

Advanced Range sorting options
↳
for full control over sorting

———×——————×————×————×——→

* **Filtering :**
Showing only the data that meets a specific criteria while hiding the rest.

Select column headers and > Data > Create filter

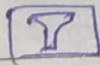Click 🔽 > Clear > [🔍]

enter the field you wanna filter wt
In this case Eims.

[GREAT]

To view all data turn off of the filter !
By Data > Remove filter
     OR
     [🔽] in second left
       ↓
You can also create a filter from here!

* Converting data →

Select desired column

Here menu ; Go to Edit > Find & Replace

Put the values get the results
Here : Find none & replace with 0
Check match entire cell contents.

Clicks Replace All
Repeat to change all occurances.
Repeat for fedu column!

————————×——————

And a lot of information on sheets & SQL

* Question	Spreadsheet	Database
Store data	cells	tables
Interact	calculation, graph report (GUI)	store, search & sort (CLI)
Powerful	less	more
Sorting Pros & cons	small dataset easy due to inbuilt function	great power, requires knowledge.
filtering pros and cons	easy to use filter option limitation on size	can handle complex filtering easily with advanced knowledge

——— ×———→

Big Query :-

(A) Sandbox

Free

Anyone can login with
a account & use it!

> Max 12 projects at a time

DM2 { > cannot insert new records to DB
not   { > Cannot update field values of
Supported   existing records
       > No payment info

(B) Free Trial

> 300 $ Credit during first
  90 days
> Can easily practice within
  budget
> Update post limit
> not auto charge
> payment option

Project transfer: Trial to Sandbox not allowed

＊ > Big Query landing Page
   > SQL workplace

Goal
: search public pataset
  through data explores
: run query
: upload own data.

＊ Pin Public: data > explorer > search public >
                        > Broaden search
                        > Pin "Bigquery-public-data"
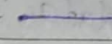
＊ Explore > [+ Add Data]

> Explore public datasets
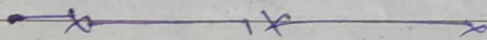> Opens market place

> select the appropriate dataset & table

Here : > Bigquery-public-data
        > austin-311
           > 311-service-requests
   ↓
           Click Query

SELECT * .... ;

congratulation you ran your first query in Bigquery.

———*————*————>

Create Dataset

: Select your database project
Here: GDAC Getting started (⋮)
   ↑
                    Click Here
                    Create the dataset
                    Give name hit create

Now go to that the dataset, click
on the window hit [+] Creat table Icon

Give Table Name   DAMN NICE
select source
schema ! Auto detect   I created an empty
Click create Table   Table
                                        only !

* Runs query on BigQuery

```
SELECT *
FROM D.t-name
       ↓
Database-Name    T-name
```

```
SELECT *
FROM bigquery-publicdata . sunroof-solar . censustract
LIMIT 25;
```

Specify column to get those results.

WHERE : where to look for information.

```
SELECT * FROM big-roof
WHERE state_name = 'Pennsylvannia'
```

; or not doesn't matter on BigQuery.

Task :

```
SELECT * FROM company
WHERE country = 'Germany';
```

* Effectively organise data

Best practices when organizing data
- ⓞ Naming conventions
- ⓞ Foldering
- ⓞ Archiving older files

Naming Conventions:
Consistent guidlines that describe the content, date or version of a file in its name.

> use logical & descriptive names for your files to make them easier to find and use

• Organize your files into folders (foldering)

• Break folders down into subfoldess.

Move old projects to a separate location to create an archieve and cut down on clutter

ⓞ Align your naming and storage practices with your team.

ⓞ Develop metadata practices.

Think about how often you'll making copies of data and storing it in different places.

Avoiding Spaces makes it easier to work with SQL

* File naming DO's :→
  - Work out your conventions early
  - Align file naming with your team
  - make sure file names are meaningful
  - keep file names short and sweet
  - Format dates YYYY mmdd : SalesReport2020 1125
  - lead revision numbers with 0 :
    Sales Report 20201125 v02
  - Use underscores, hyphens or Capatalized letters:
    SalesReport_2020_11_25_v02

✱ Create a text file that lays out all your naming conventions on a project.

Metadata

————— ✗ ————— ✗ —————

17/V 23

week 4
2/2

Data security:
Protecting data from unauthorized access or corruption by adopting safety measures.

Access control :→ Password protection
                  User permission

Excel :→ Encrypt files & worksheets with passwords.

Google sheets :→ sharing options,
                 hiding tabs (can be Unhidden by others)

Encryption uses a unique algorithm to alter data and make it unusable by users and applications that don't know the algorithm.

This algorithm is saved as a "key" which can be used to reverse encryptions, so if you have the key you can still use data in its original form.

Tokenization: replaces the data elements you want to protect with randomly generated data referred to as a "token".

The original data is stored in a separate location and mapped to tokens. To access the complete original data, the user or application needs to have permission to use tokenized data and tokenized mapping.

→ This means that even if tokenized data is hacked, original data is safe & secure in separate location.

Kaggle : version feature , Github version Control

private data & notebooks

Managing presence as a Data Analyst.

* Why online presence is important?
  > Helps potential employers find you
  > make connections with other analysts
  > learn & share data findings
  > Participate in community events.

Build Network

○ LinkedIn : make connections, follow industry, job oppo.
○ GitHub : part code-sharing / part Social media

"You never know you might end up with someone working with someone again"

GitHub: • Share insights & resources
        • Read forums & wikis
        • Manage team projects   • Host comm. events

"I've been using these already but I don't want to miss on something just in case!!!"

∼∼∼∼∼∼∼∼

* Would I be okay with potential employers and colleagues seeing your social media profiles.?

* A potential employer ask himself if you are the right person!

* Check the privacy settings on your accounts.

Maya Angelou ⟶ Oprah Winfrey

Mentor Steven Spielberg ⟶ JJ Abrams

Warren Buffet ⟶ Bill gates

Damn! I need <u>this</u>

" ⋆ Filter out your Socials ⋆ "

───────────────

⋆ Networking : Professional Relationship Building

Digital Networking can take you <u>anywhere!</u>

<u>Podcasts :</u>   Partially Derivative
      O'Reily Data show.

{ O'Reily
  kaggle
  KDNuggets
  GitHub
  medium

───────────────

⋆ <u>mentor :</u>
A professional who shares their knowledge, skills
and experience to help you develop & grow

SCORE org
micromentor.org
mentorship

⋆ Sponsor :→
A professional advocate who's committed to moving
a sponsee's career forward within an org.

A mentor helps you __Skill Up__

A sponsor helps you __more up__

* You don't get to choose sponsor, the sponsor always chooses you!

__Rachel__

Its really important to have someone in your corner as you're navigating all of the different ins and outs of your career

Most import thing to look for in a mentor is somebody who you will get along with and somebody that you trust.