



</talentlabs>

CHAPTER 3

Data Wrangling with Google Sheets

Learning Objectives



Understand how to prepare a small dataset for analysis



Leverage Google Sheets as a data preparation tool



Complete the entire data wrangling workflow on real datasets in Google Sheets



Agenda

- Loading Data into Google Sheets
- Inspecting Data
- Data Cleaning
- Enriching Data
- Verification
- Version History & Exporting Data
- Chapter Summary & Assignment



Agenda

- Step 1: Loading Data into Google Sheets
- Step 2: Inspecting Data
- Step 3: Data Cleaning
- Step 4: Enriching Data
- Step 5: Verification
- Step 6: Documentation & Exporting Data
- Chapter Summary & Assignment



Loading Data into Google Sheets



Setting up Google Sheets



Organising data



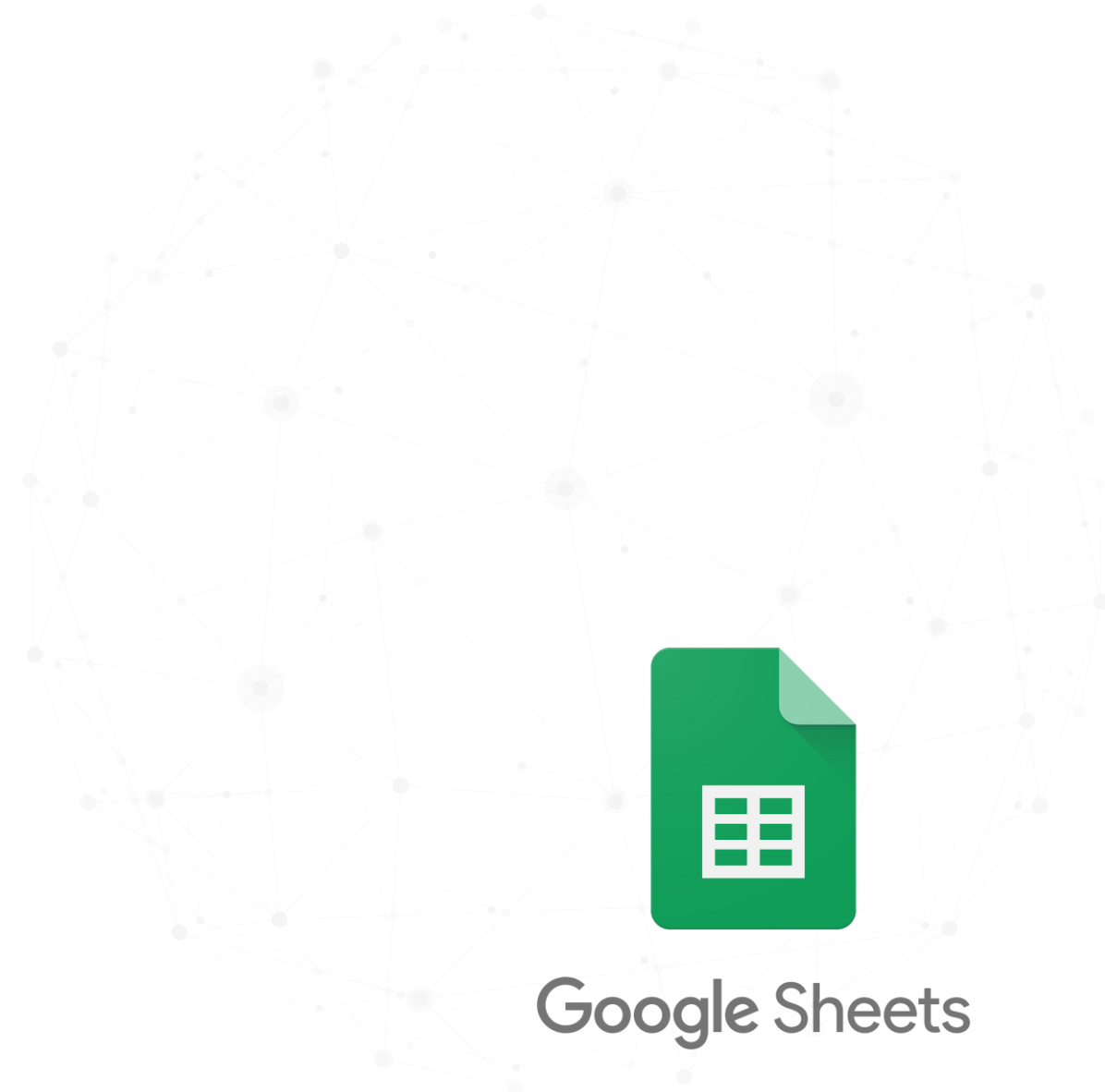
Google Sheets – Getting Started

Go to: sheets.google.com

Login or create an account

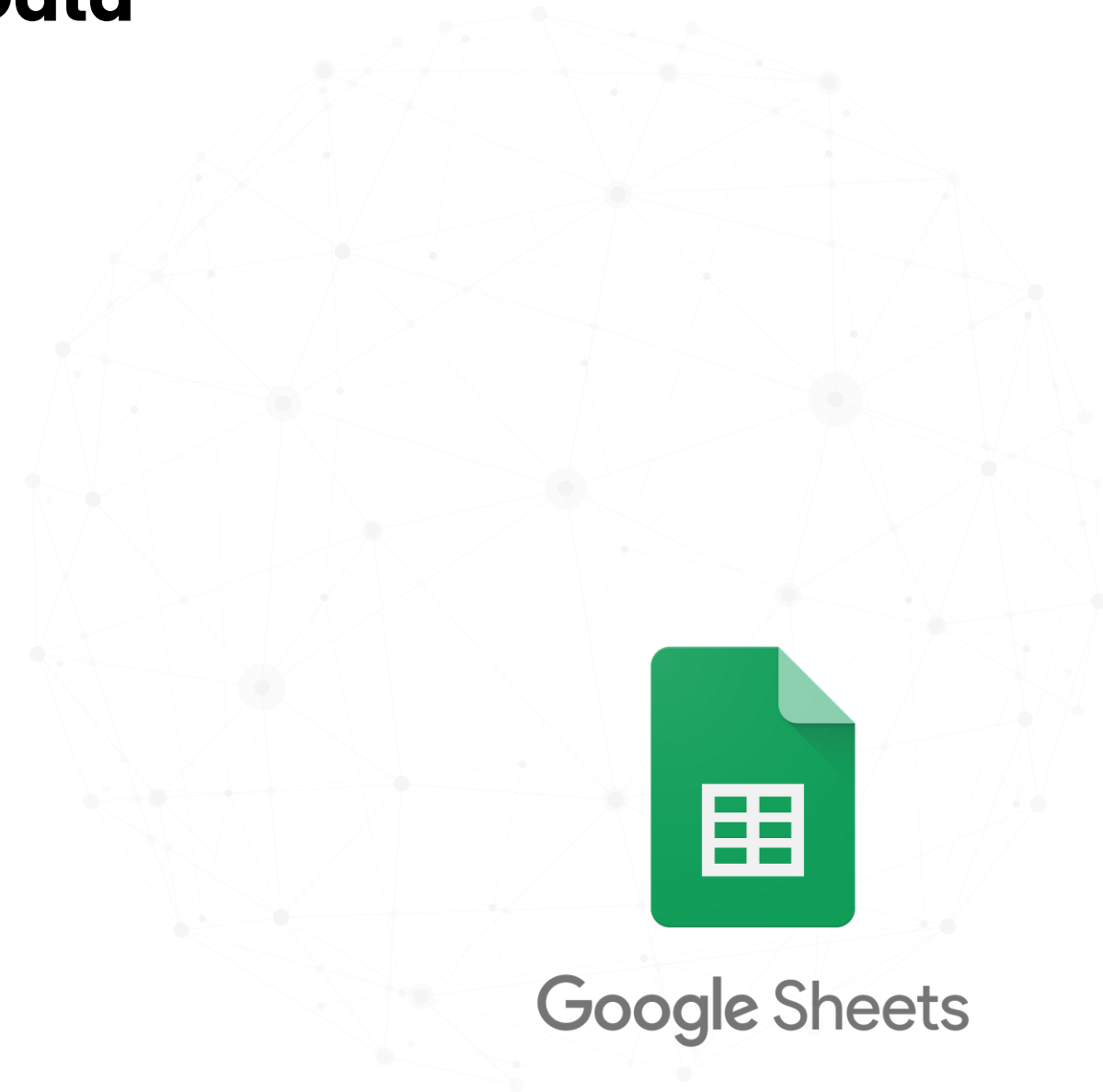
Start a new spreadsheet +

Look up Google Sheets cheat sheet!



Google Sheets – Organising the Data

- Back up and load the data
- Setup folder structure
- Clear existing formatting
- Resize columns
- Format headers



Inspecting Data



How to inspect a dataset



Google Sheets – Data Inspection

- Have metadata file on hand
- Make an outstanding data cleaning action list
- Check every column of data
- Look for dirty data



Google Sheets

Google Sheets – Whole Dataset

➤ Duplicates:

Data > Data clean-up > Remove duplicates

➤ Missing data:

- Highlight cells: Format > Conditional formatting
- Count missing: COUNTBLANK or COUNTIF after column header

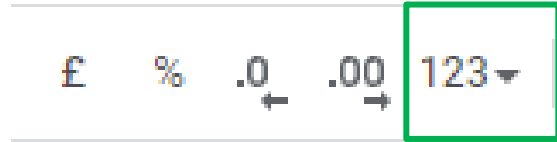
TRANSPOSE



Google Sheets

Google Sheets – Individual Columns

- Data type



- Check for errors and inconsistencies: Data > Column stats

Use Pivot table: Insert > Pivot Table

- Spell check: Tools > Spelling > Spell check



Google Sheets

Data Cleaning



Implementing outlined data cleaning operations



Data Cleaning Pointers

➤ Remove duplicates:

Data > Data clean-up > Remove duplicates

➤ Handle missing values:

- Drop
- Impute
- Flag

➤ Fix inconsistent data entry: Edit > Find and replace

➤ Data > Data clean-up > Clean-up suggestions



Google Sheets

Enriching Data



Applying different data
wrangling methods



Google Sheets – Enriching data

- Sort columns: Data > Sort Sheet
- Filter columns: Data > Filter views > Create new filter view
- Split text columns: Data > Split text to columns
OR use LEFT and RIGHT functions
- Join text columns: CONCATENATE function
- Extract from text: REGEXMATCH/REGEXEXTRACT



Google Sheets








Verification



Validating data quality



Data Cleaning Checklist

| | | |
|----|---|-------------------------|
| 01 |  | Irrelevant columns |
| 02 |  | Duplicates |
| 03 |  | Data types |
| 04 |  | Missing data |
| 05 |  | Inconsistent data entry |
| 06 |  | Errors/misspellings |
| 07 |  | Format |



Version History & Exporting Data



Version History



Exporting data



Chapter Summary & Assignment



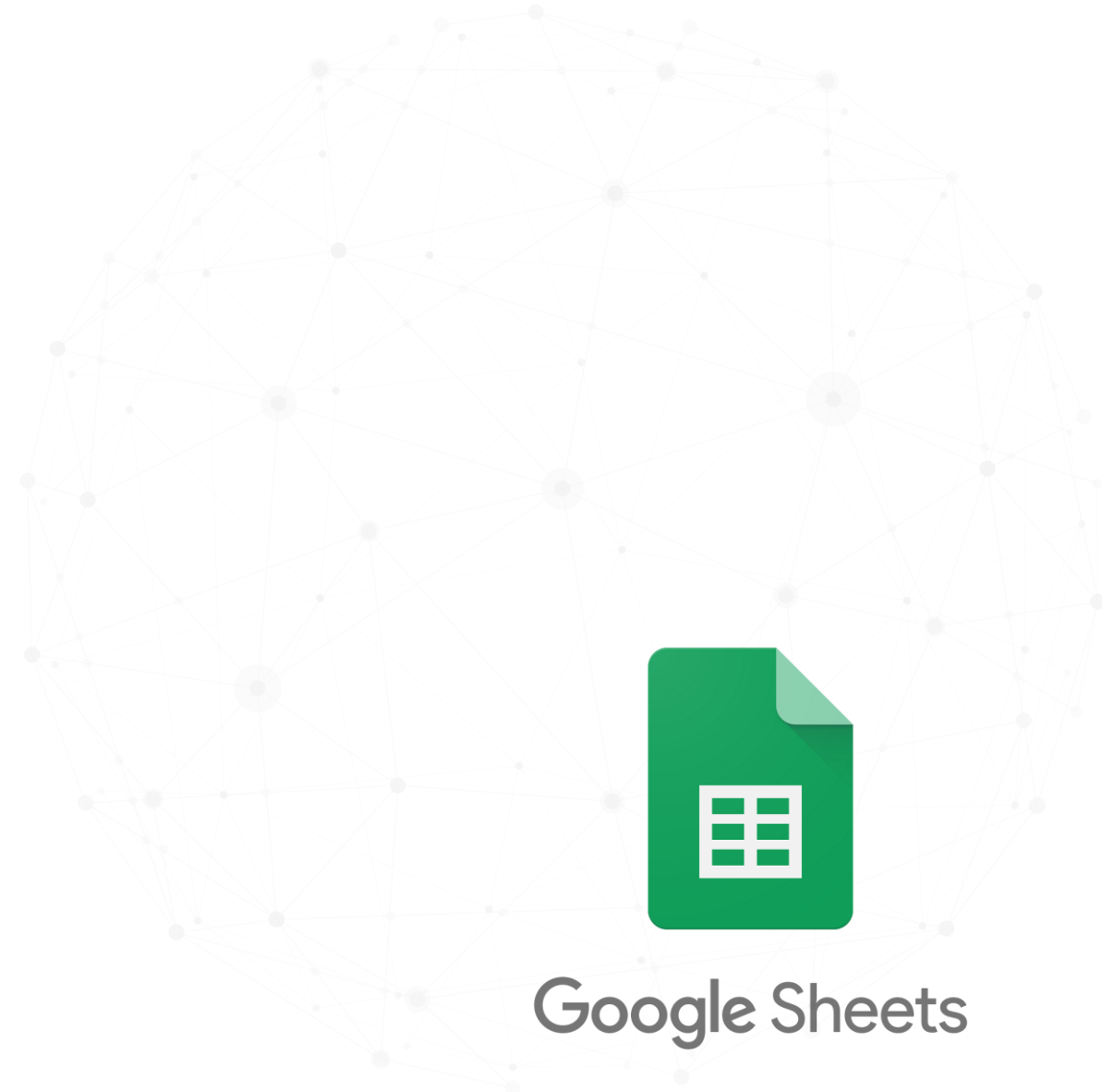
Google Sheets – Getting Started

Go to: sheets.google.com

Login or create an account

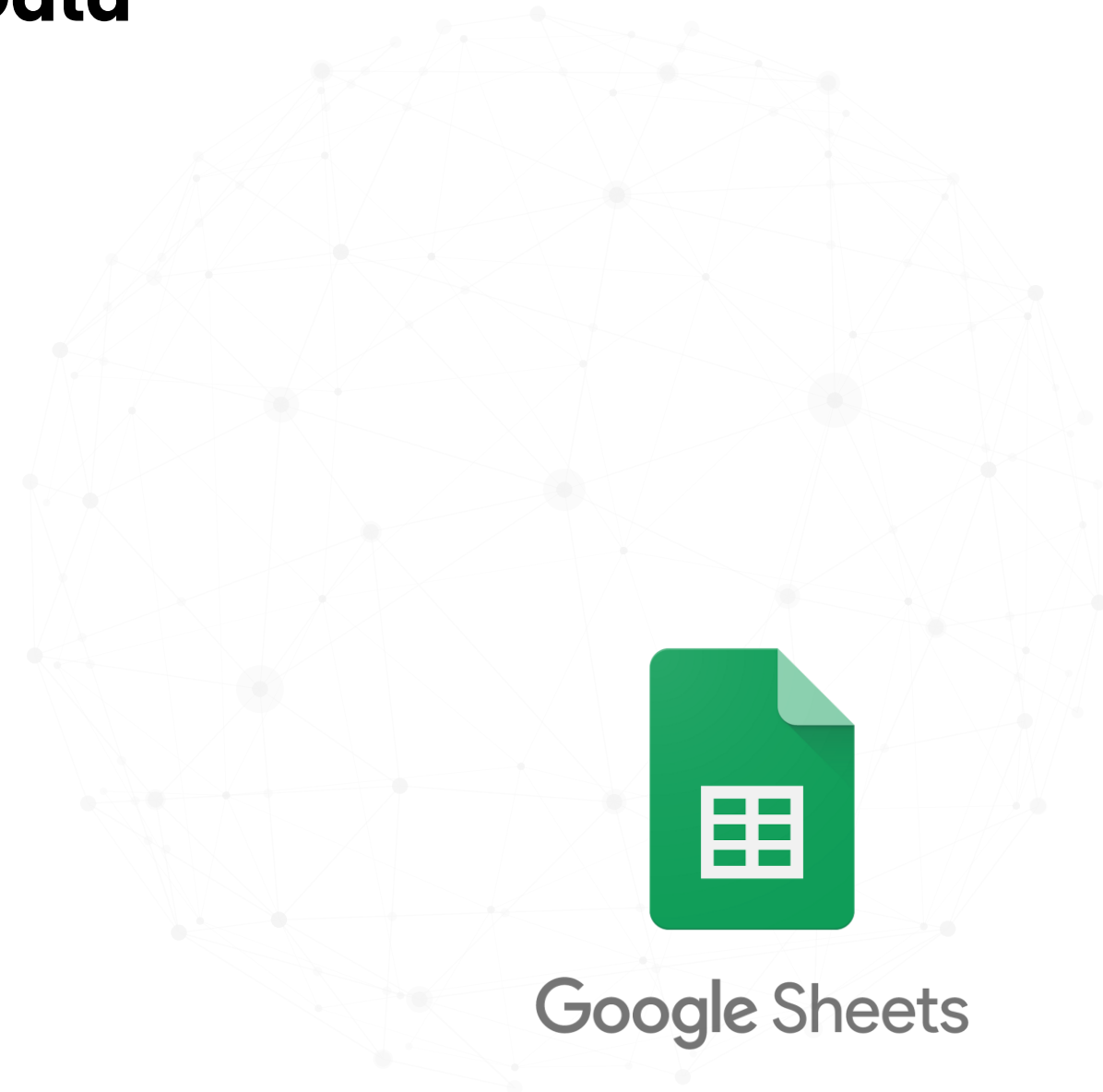
Start a new spreadsheet +

Look up Google Sheets cheat sheet!



Google Sheets – Organising the Data

- Back up and load the data
- Setup folder structure
- Clear existing formatting
- Resize columns
- Format headers



Google Sheets – Data Inspection

- Have metadata file on hand
- Make an outstanding data cleaning action list
- Check every column of data
- Look for dirty data



Google Sheets

Google Sheets – Whole Dataset

➤ Duplicates:

Data > Data clean-up > Remove duplicates

➤ Missing data:

- Highlight cells: Format > Conditional formatting
- Count missing: COUNTBLANK or COUNTIF after column header

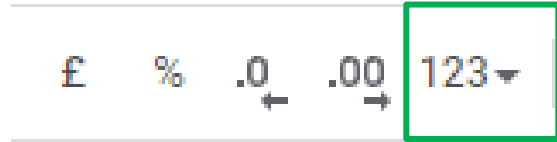
TRANSPOSE



Google Sheets

Google Sheets – Individual Columns

- Data type



- Check for errors and inconsistencies: Data > Column stats

Use Pivot table: Insert > Pivot Table

- Spell check: Tools > Spelling > Spell check



Google Sheets

Data Cleaning Pointers

➤ Remove duplicates:

Data > Data clean-up > Remove duplicates

➤ Handle missing values:

- Drop
- Impute
- Flag

➤ Fix inconsistent data entry: Edit > Find and replace

➤ Data > Data clean-up > Clean-up suggestions



Google Sheets








Google Sheets – Enriching data

- Sort columns: Data > Sort Sheet
- Filter columns: Data > Filter views > Create new filter view
- Split text columns: Data > Split text to columns
OR use LEFT and RIGHT functions
- Join text columns: CONCATENATE function
- Extract from text: REGEXMATCH/REGEXEXTRACT



Google Sheets

Data Cleaning Checklist

| | | |
|----|---|-------------------------|
| 01 |  | Irrelevant columns |
| 02 |  | Duplicates |
| 03 |  | Data types |
| 04 |  | Missing data |
| 05 |  | Inconsistent data entry |
| 06 |  | Errors/misspellings |
| 07 |  | Format |



Assignment Information

Hands-on activity

