

# The data preparation journey

DATA PREPARATION IN EXCEL

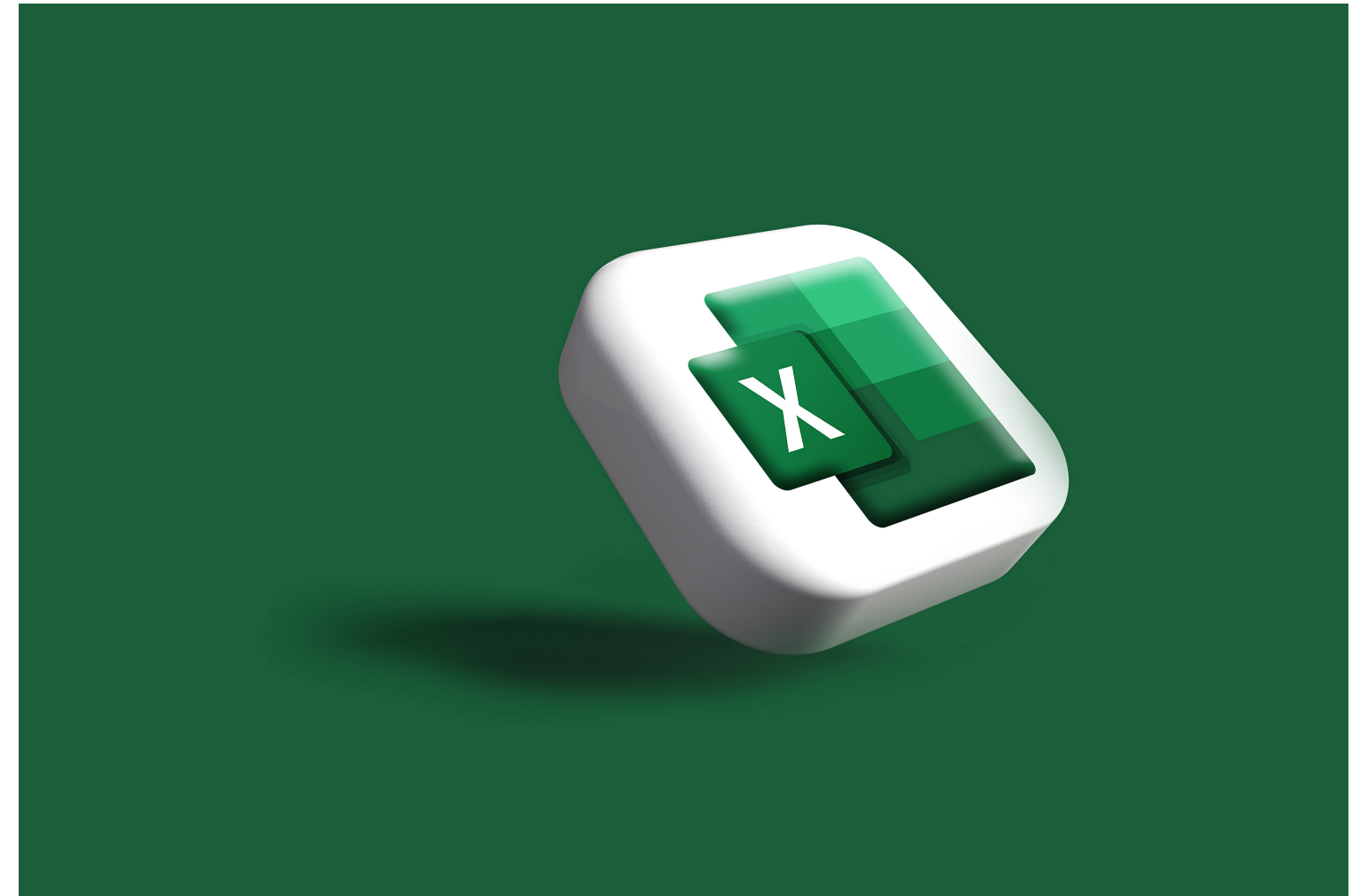


**Iason Prassides**

Content Developer, DataCamp

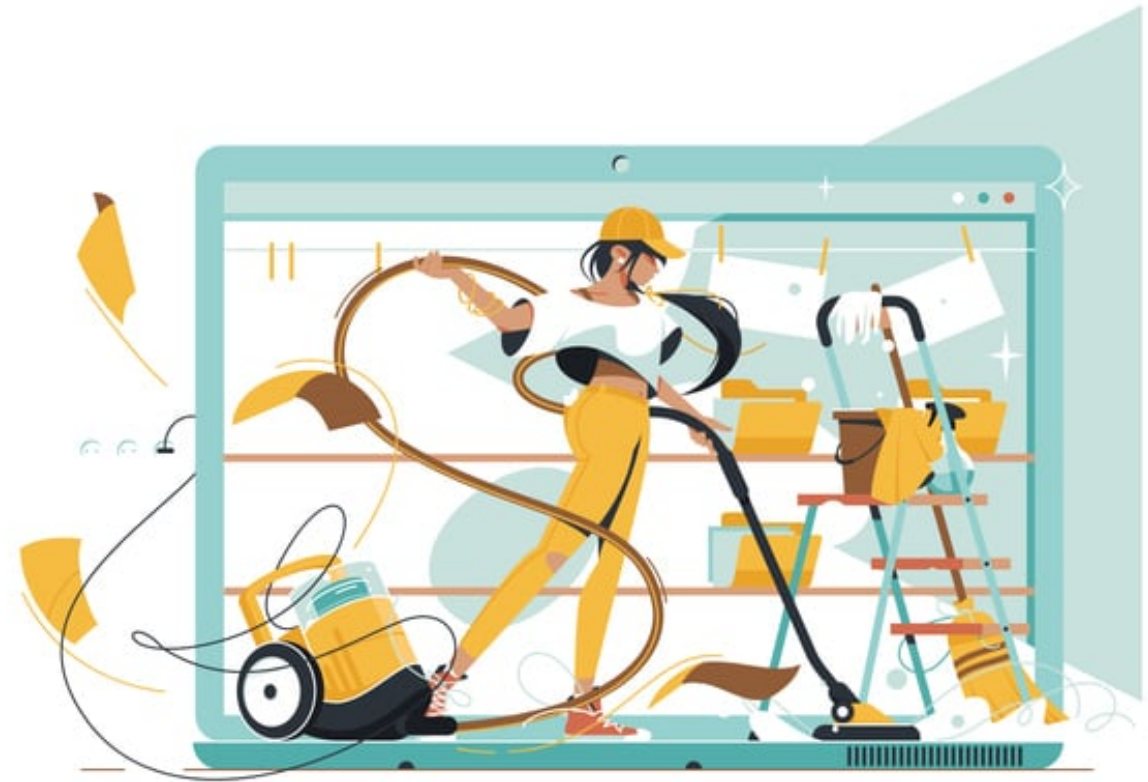
# Excel knowledge

- Build on basic functions and features
- How to use Excel to prepare data
- Prerequisite course:
  - Introduction to Excel



# What is data preparation?

- Data preparation involves:
  - Cleaning data
  - Transforming data
  - Organizing data
- Good quality analysis, needs high quality and "clean" data
- Ensures accurate and efficient data analysis



# Doing laundry

- Collect and sort unclean clothes
- Select settings for the washing machine
- Start the washing machine to clean your clothes
- Organize your clean, dry clothes



# Doing data preparation

- Collect, sort, and filter raw data
- Set the correct formatting and data types
- Clean raw data
  - Remove duplicates
  - Correct errors
  - Fill in missing data
- Check data quality and summarize your data better



# Gathering raw data

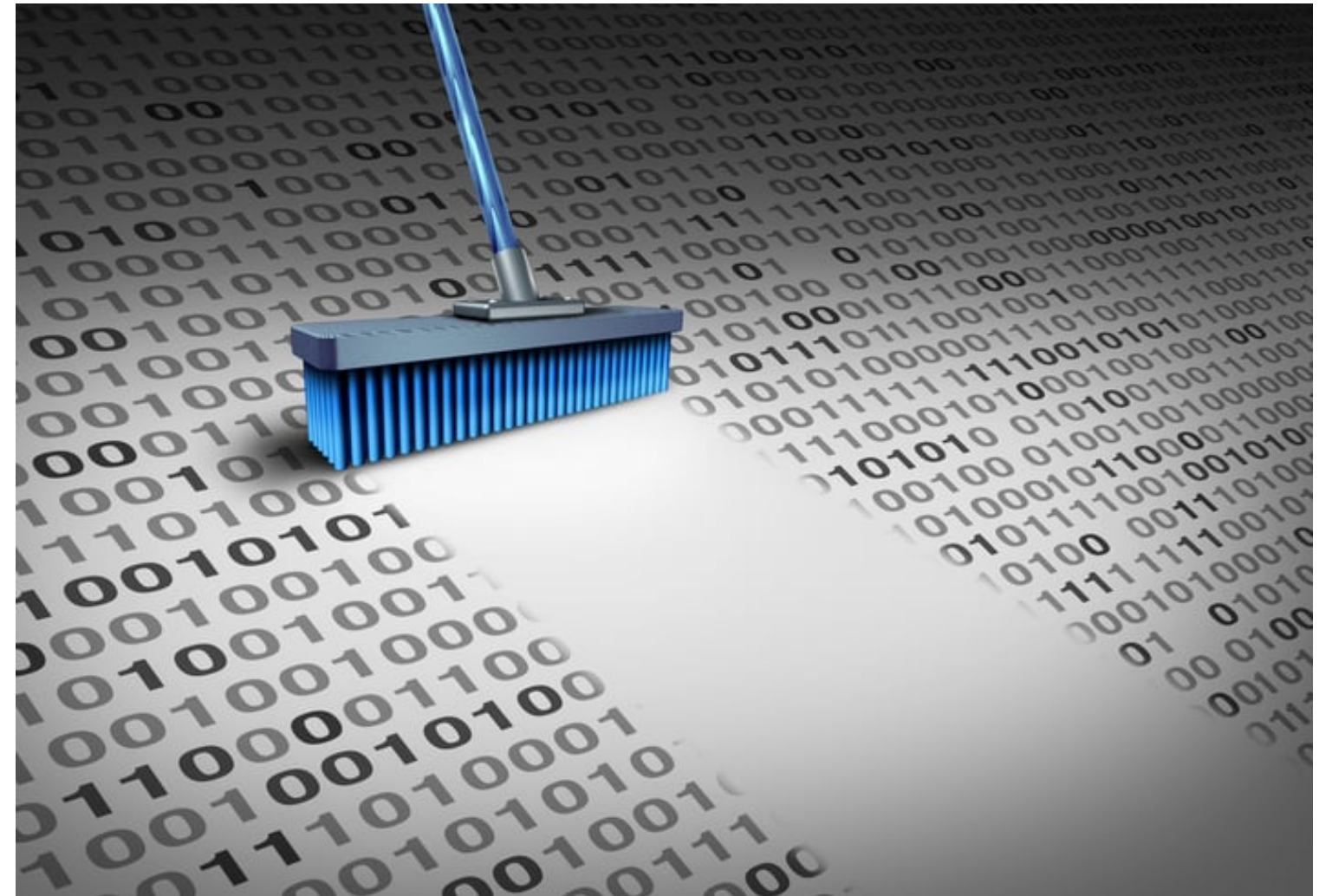
- Collect the required raw data in Excel
- Databases like SQL Server can export data to Excel
- Enter data manually into Excel
  - Can be very time-consuming
- Import data from other files:
  - CSV files
  - Text files
  - Web (HTML) files





# Removing duplicates

- Key step in data preparation process
- Excel has a "Remove Duplicates" feature
- Important to remove incorrectly recorded duplications
  - Keep values that are correctly repeated



# Fill options

- Excel fill features include:
  - Flash-fill
  - Advanced Fill Series
- **Flash-fill** fills columns based on patterns from existing data
  - Ex: Full name column
- **Advanced Fill Series** allows you to customize your column fill process
  - Ex: Date values based on start and end date





# Dataset



- Raw dataset from a fictitious Global supply chain firm, DataCo
- Recently hired as an analyst
- Data contained in three different files

# Let's practice!

DATA PREPARATION IN EXCEL

# Load and clean data

DATA PREPARATION IN EXCEL



**Iason Prassides**

Content Developer, DataCamp

# Let's practice!

DATA PREPARATION IN EXCEL