

*Worksheet 4 - Sep 26, 2025*

**Build and Clean Your Own Dataset**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

✨**Goal:** Learn how real-world data becomes messy, and how to clean it — by **making your own messy dataset!**

**✏️ Quick Review: Python Lists and Dictionaries**

### 🔹 What is a List?

A **list** is a collection of items. It's like a row of backpacks 🧳 — each one holds something.

A close-up of a name

Description automatically generated

✅ Lists are written inside square brackets []  
✅ Items are separated by commas  
✅ You can store **strings**, **numbers**, or **mix** them

**🔹 What is a Dictionary?**

A **dictionary** connects **keys** to **values**, like a label on a drawer.

A close-up of a text

Description automatically generated

✅ Dictionaries use **curly braces {}**  
✅ Each entry is a **key → value** pair  
✅ The keys are like **column names**  
✅ The values are **lists** of data for that column

**🔄 DataFrame = a Table Made from a Dictionary**

In pandas, we use pd.DataFrame() to turn a **dictionary** into a **table** (called a DataFrame).

🔹 Exercise 1 – Create Your First DataFrame

A computer screen shot of a code

Description automatically generated

🧠 **Question:**

1. How many rows and columns are in this table?
2. What are the column names?
3. What type of data is in each column?

**🔹 Exercise 2 – Look at Individual Columns**



🧠 **Question:**

* What happens when you look at just one column?
* Try printing another column like data["Name"]. What do you notice?

**🔹 Exercise 3 – Add a New Column**

A number and equation on a white background

Description automatically generated

🧠 **Question:**

* What happens to the table when you add a new column?
* Can you guess what this does?

**🔹 Exercise 4 – Edit a Value**

A close-up of a text

Description automatically generated

🧠 **Question:**

* What changed?
* What does loc[1, "Score"] mean?

**🔹 Exercise 5 – Your Turn 🎯**

Now YOU try it:

* Create your own table with 3 rows and these columns:
  + "City" (e.g. "Milwaukee", "Chicago", "New York")
  + "Temperature" (e.g. 75, 62, 89)
  + "Rainy" (Yes/No)

Then answer:

* How many rows and columns?
* What happens if you print just ["City"]?

### 🧠 Reflection

What’s the difference between a list and a DataFrame?  
When would you use one vs the other?