

# Unit 4: Datastores, Databases, and Processing CSV Data

---

Prerequisites: [Processing and Validating User Inputs](#)

## Learning Objectives

In this unit, we'll illustrate the concept of program-data independence as we explore integrations between software and datastores. We'll focus on the Comma-separated Values (CSV) format and learn how to use Python to process CSV data.

## Discussions and Notes

- Program-Data Independence
- [Machine-readable Data Formats](#)
- [Managing Files](#):
  - [The os Module](#) (focus on filepath construction)
- Processing JSON Data:
  - [The json Module](#)
- Processing CSV Files:
  - [The csv Module](#)
  - [The pandas Package](#)

Optional Further Exploration (for students seeking greater difficulty level now or when revisiting from Unit 9):

- Processing Google Sheets:
  - [The gspread Package](#)
- Processing Relational (SQL) Data:
  - [The psycopg Package](#) (for PostgreSQL databases)
  - [The pymysql Package](#) (for MySQL databases)
- Processing XML and HTML Data:
  - [The BeautifulSoup Package](#)
- Parsing PDF Files:
  - [The pdftotext Utility](#) (for parsing PDFs)
  - [The fpdf Package](#) (for generating PDFs)

## Activities and Exercises

- ["Omniparser" Exercise](#)
- ["Monthly Sales Reporting" Exercise](#)