Pandas: Getting Started — Notes

# What is Pandas?

Pandas is a Python library used for working with datasets.  
It helps manage and analyze data using structures like DataFrames (like tables).

# What is a DataFrame?

A DataFrame is a table with rows and columns, like an Excel sheet.  
It's used for handling structured data (organized with labels for rows and columns).

# What is a CSV File?

CSV stands for Comma Separated Values.  
It's a simple text format to store structured data.  
Used to store tables: each row is a record, and each column is a field.  
Structured data means data where each column has meaning (e.g., price, location).  
Panel data is repeated observations over time (e.g., house prices over months).

# CSV Example

property\_type,department,lat,lon,area\_m2,price\_usd  
house,Bogotá D.C,4.69,-74.048,187.0,"$330,899.98"

# Code Examples

## 1. Importing Pandas and Reading a CSV

import pandas as pd  
  
# Load the CSV file into a DataFrame  
df = pd.read\_csv("houses.csv")  
  
# Show the first few rows of the table  
print(df.head())

## 2. Viewing the Data as a Table (DataFrame)

# View the shape of the table (rows, columns)  
print(df.shape)  
  
# View column names  
print(df.columns)

# Key Terms

|  |  |
| --- | --- |
| Term | Meaning |
| Pandas | Python library for data manipulation |
| DataFrame | 2D table (like Excel) with rows and columns |
| CSV | A text file format for storing table data |
| Structured Data | Data arranged in a clear table format |
| Panel Data | Data collected over time (same subjects, different time points) |