

Module 2: Loading Data

Mastering Ingestion and Inspection in Jupyter



Introducing Pandas

The Engine of Data

Analysis

What is Pandas?

The Data Analyst's Swiss Army Knife

Pandas is the primary library for data manipulation and analysis in Python. It provides high-performance, easy-to-use data structures.

- Handles structured data (tables)
- Built on top of NumPy
- Essential for cleaning and transforming



The DataFrame Concept

Excel Comparison

Think of a DataFrame as a digital spreadsheet. It has rows and columns, with labels for both. Every operation you do in Excel can be automated here.

Key Features

DataFrames are heterogeneous, meaning columns can hold different types (integers, strings, dates) while maintaining strict structure.



File Ingestion

Bringing Raw Files to
Life

Loading CSV Files

Comma-Separated Values (CSV) are the universal language of raw data. In Jupyter, ingestion is a single line of code away.

Syntax:

```
df = pd.read_csv('filename.csv')
```

Note: Ensure the file is in the same folder as your notebook.

```
import os
from flask import Flask, render-template
import,
import sys
    app = Flask b(((<maN1>)))
}
def index::;
    return tenderalater 'index.html')
```

Loading Excel Sheets



Simple Import: Use `pd.read_excel('file.xlsx')` to ingest standard workbook files directly into a variable.



Sheet Selection: Target specific tabs using the `sheet_name` parameter (e.g., `sheet_name='Q3_Sales'`).



Back-end Power: Pandas uses engines like `openpyxl` under the hood to preserve complex spreadsheet formatting.



Jupyter Advantage: Unlike Excel, Jupyter won't "freeze" when loading datasets with hundreds of thousands of rows.



Data Inspection

Auditing Your Spreadsheet
Structure

Quick Visual Audit



Using `.head()` and `.tail()`

Never print 1,000,000 rows. Use `.head()` to see the first 5 rows and verify your headers imported correctly.

Use `.tail()` to check the end of the file for potential corruption or summary rows.

The Technical Deep Dive



`.info()` Command

Provides a concise summary of the DataFrame including column names and data types (Dtypes).



Spot Nulls

Quickly identify missing data by looking at "Non-Null Count" for every column in the output.



Memory Usage

See exactly how much RAM your dataset is consuming within the Jupyter environment.

Command Cheat Sheet

Command	Action	Outcome
<code>read_csv()</code>	Import text data	New DataFrame created
<code>read_excel()</code>	Import .xlsx files	Spreadsheet loaded
<code>df.head()</code>	Preview top rows	Visual data check
<code>df.info()</code>	Technical audit	Schema & Dtypes check

Questions?

Thank you for your
attention.

www.anaconda.com | support@training.com

Image Sources



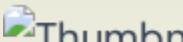
https://seobutler.com/wp-content/uploads/2025/09/Best-laptop-for-data-scientist-and-analytics%E2%80%8B-seobutler.com_-1024x576.jpg

Source: seobutler.com



<https://media2.dev.to/dynamic/image/width=1000,height=420,fit=cover,gravity=auto,format=auto/https%3A%2F%2Fdev-to-uploads.s3.amazonaws.com%2Fuploads%2Farticles%2F4lav9t6ha164k605cm8x.png>

Source: dev.to



https://media.istockphoto.com/id/2249792062/photo/data-analytics-and-visualization-with-magnifying-glass-over-digital-charts-and-dashboards.jpg?s=612x612&w=0&k=20&c=YU2Nf5CvO2TNVuxUSvG7J_sBIthM4YhojaHohUoMFmU=

Source: www.istockphoto.com