Import & Export Data (CSV, Excel, JSON)

🔹 Reminder

- Yesterday: we created DataFrames manually (`dict → DataFrame`).
- Today: data usually comes from **files** (CSV, Excel, JSON) or **APIs**.

🔹 Data Formats

Notes

- **CSV** → Most common, simple, readable.
- **Excel** → Business world, multiple sheets.
- **JSON** → API/web services, nested structures.

Examples

students.csv

```
Name,Age,City
Alice,25,Nairobi
Bob,30,Mombasa
```

students.xlsx → same as above, saved as Excel.

students.json

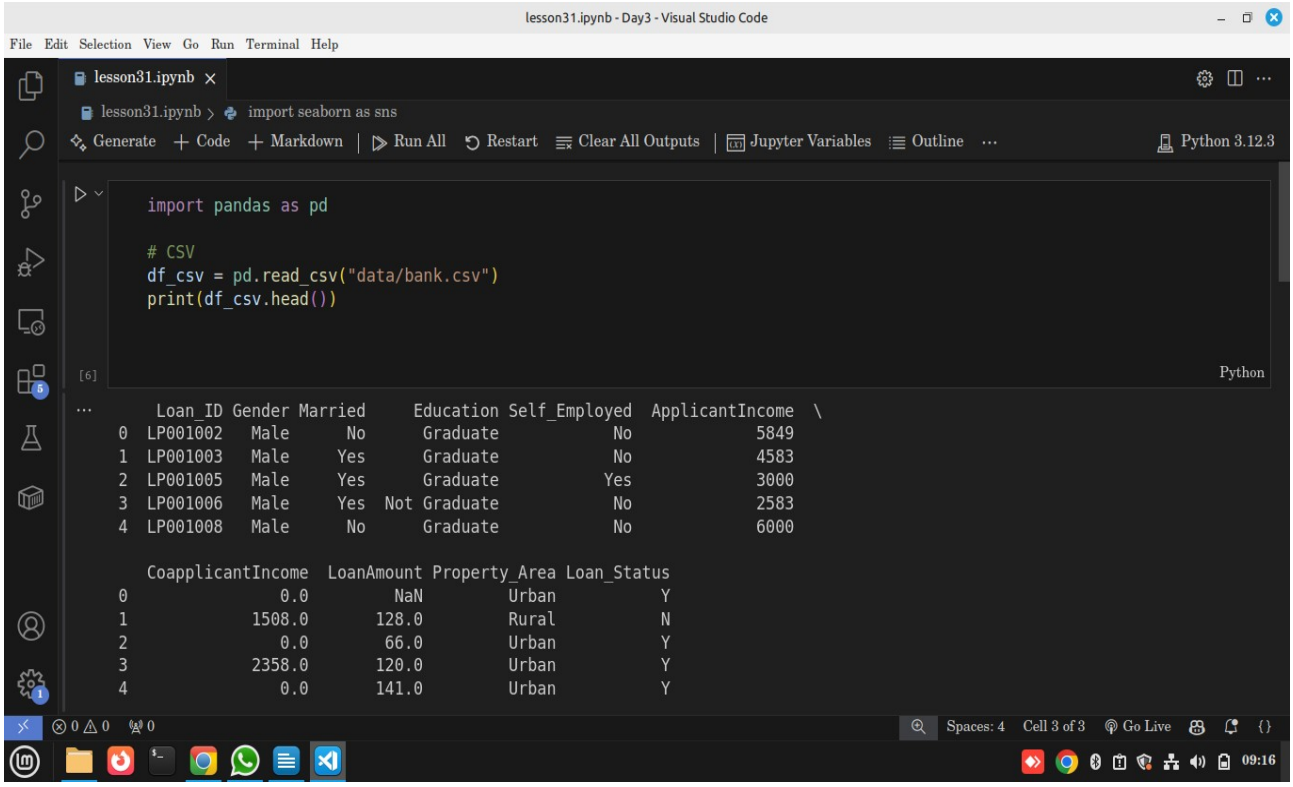
```
[
  {"Name": "Alice", "Age": 25, "City": "Nairobi"},
  {"Name": "Bob", "Age": 30, "City": "Mombasa"}
]
```

Use in Data Science:

- CSV → Kaggle datasets.
- Excel → corporate records.
- JSON → web APIs (Twitter, Weather).

🔹 Importing Data

Practical – Reading Different Formats



- `pd.read_csv()` → loads CSV into DataFrame.
- `pd.read_excel()` → needs `openpyxl` installed.
- `pd.read_json()` → parses JSON into DataFrame.

Use in Data Science

These are **entry points** for every project — from raw file → DataFrame → analysis.

🔹 Exporting Data

Practical – Saving Data

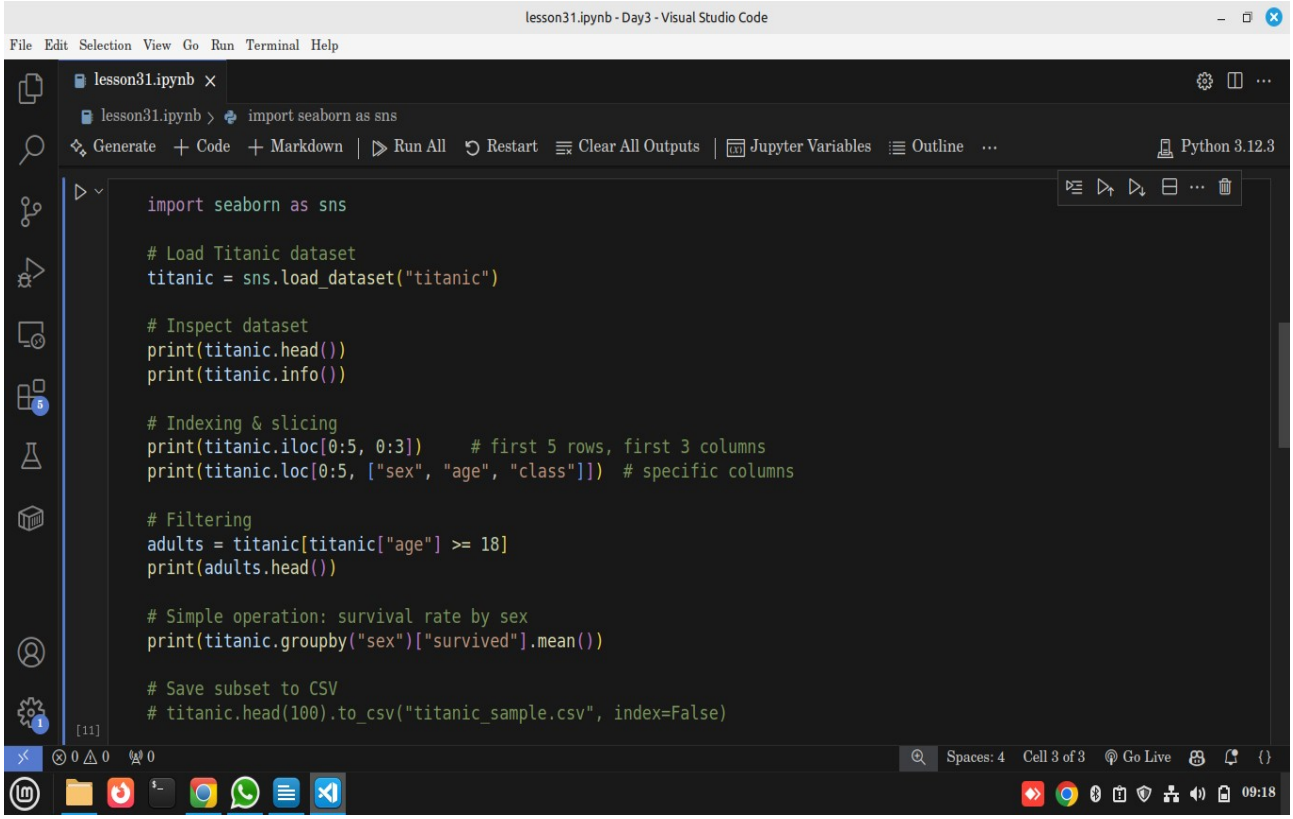
```
# Save to CSV
df_csv.to_csv("output.csv", index=False)

# Save to Excel
df_csv.to_excel("output.xlsx", index=False)

# Save to JSON
df_csv.to_json("output.json", orient="records")
```

Explanation

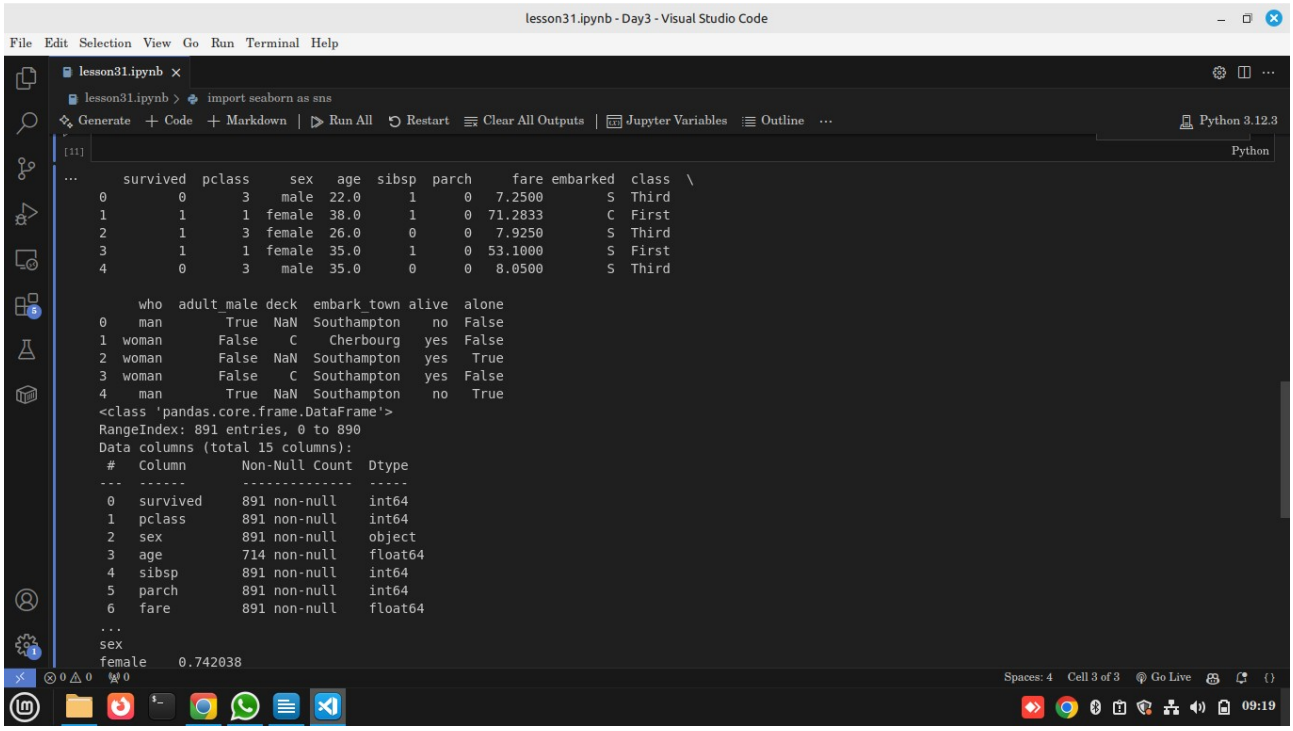
- `index=False` → avoids extra column.
- `orient="records"` → JSON-friendly format.

Use in Data Science
Exporting lets you share cleaned/processed data with **colleagues, dashboards, or ML pipelines**.

🔹 Hands-On: Titanic Dataset



Sample Output



Explanation

- `iloc` → integer-based indexing.
- `loc` → label-based indexing.
- Filtering (e.g., age ≥ 18) → data wrangling basics.
- `.groupby("sex")` → real analysis: shows women had higher survival rates.
- `.to_csv()` → saves results for later.

Use in Data Science
This shows how **imported data** can immediately be **inspected, filtered, and analyzed** — bridging file formats with real analysis.

🔍 Reflection

- Which format is most common? (**CSV**).
- Why use JSON? (**APIs / real-time data**).
- Why is Excel risky? (**Formatting errors, multiple sheets**).
- How does indexing/slicing help? (**Lets us zoom into relevant data quickly**).