**🧮 PHẦN 1 — PANDAS (XỬ LÝ DỮ LIỆU / DATA HANDLING)**

*(Cập nhật đầy đủ, song ngữ, có thể dùng như sổ tay tra cứu nhanh)*

**🧰 1. Tạo và đọc dữ liệu / Creating & Reading Data**

| **Lệnh / Hàm** | **Giải thích (Việt)** | **English Explanation** | **Khi nào dùng / When to use** | **Tham số chính (Parameters)** |
| --- | --- | --- | --- | --- |
| import pandas as pd | Gọi thư viện Pandas | Import the Pandas library | Bắt buộc khi làm việc với dữ liệu dạng bảng | — |
| pd.Series(data) | Tạo đối tượng 1D (một cột dữ liệu) | Create a 1D labeled array | Khi cần lưu một cột dữ liệu đơn giản | data: list, dict, array |
| pd.DataFrame(data) | Tạo bảng dữ liệu 2 chiều | Create a 2D labeled table | Khi cần lưu nhiều cột có tên | data: dict, list of dicts, arrays |
| pd.read\_csv('file.csv') | Đọc dữ liệu từ file CSV | Read CSV file into DataFrame | Đọc dữ liệu dạng bảng phổ biến nhất | sep, header, encoding |
| pd.read\_excel('file.xlsx') | Đọc file Excel | Read Excel file | Khi dữ liệu ở định dạng Excel | sheet\_name, usecols |
| pd.read\_json('file.json') | Đọc file JSON | Read JSON data | Khi dữ liệu ở định dạng JSON | — |
| pd.read\_sql(query, conn) | Đọc dữ liệu SQL | Read SQL query result | Khi lấy dữ liệu từ cơ sở dữ liệu | query, connection |

**📄 2. Xem thông tin tổng quan / Inspecting Data**

| **Lệnh** | **Giải thích (Việt)** | **English Explanation** | **Khi nào dùng** | **Tham số chính** |
| --- | --- | --- | --- | --- |
| .head(n) | Xem n dòng đầu tiên | Display first n rows | Kiểm tra dữ liệu đầu bảng | n=5 mặc định |
| .tail(n) | Xem n dòng cuối | Display last n rows | Kiểm tra phần cuối dữ liệu | n=5 |
| .info() | Thông tin cột, kiểu dữ liệu, số null | Show column info & data types | Khi cần hiểu cấu trúc bảng | — |
| .describe() | Tóm tắt thống kê (mean, std, min, max) | Summary statistics | Xem nhanh đặc điểm dữ liệu số | include='all' để lấy cả object |
| .shape | Trả về (số dòng, số cột) | Return (rows, columns) | Xem kích thước bảng | — |
| .columns | Trả về danh sách tên cột | Return column names | Khi cần đổi tên, lọc cột | — |
| .dtypes | Kiểu dữ liệu từng cột | Data types of each column | Khi cần kiểm tra hoặc ép kiểu | — |
| .value\_counts() | Đếm tần suất giá trị | Count occurrences | Khi muốn thống kê phân bố | normalize=True để lấy tỉ lệ |

**🧹 3. Lọc, chọn dữ liệu / Selecting & Filtering**

| **Lệnh** | **Giải thích (Việt)** | **English Explanation** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| df['col'] | Chọn 1 cột | Select one column | Lấy dữ liệu cột | 'col': tên cột |
| df[['col1','col2']] | Chọn nhiều cột | Select multiple columns | Khi cần nhiều cột | list tên cột |
| df.loc[row, col] | Truy cập theo **nhãn (label)** | Access by label | Khi biết tên hàng/cột | row, col có thể là tên hoặc mảng |
| df.iloc[row, col] | Truy cập theo **vị trí (index)** | Access by position | Khi biết vị trí (số thứ tự) | số hoặc slice |
| df[df['col'] > 10] | Lọc theo điều kiện | Filter by condition | Khi muốn lấy dòng thỏa điều kiện | Biểu thức điều kiện |
| df.query('col > 10 and sex == "Male"') | Lọc theo biểu thức query | Filter with SQL-like syntax | Khi viết nhiều điều kiện phức tạp | Chuỗi biểu thức |
| df.sample(n=5) | Lấy mẫu ngẫu nhiên | Random sample | Khi kiểm tra nhanh dữ liệu | n, random\_state |

**✏️ 4. Chỉnh sửa dữ liệu / Modifying Data**

| **Lệnh** | **Giải thích (Việt)** | **English Explanation** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| .rename(columns={'old':'new'}) | Đổi tên cột | Rename columns | Khi cần đổi tên cột dễ hiểu hơn | dict ánh xạ cũ→mới |
| .drop(columns=['col']) | Xóa cột | Drop columns | Khi cột không cần thiết | axis=1 cho cột |
| .drop(index=[0,1]) | Xóa dòng | Drop rows | Khi cần loại dòng theo index | axis=0 cho hàng |
| .insert(loc, column, value) | Thêm cột mới tại vị trí | Insert new column | Khi muốn thêm cột vào vị trí cụ thể | loc, column, value |
| .assign(newcol=df['a']+df['b']) | Thêm cột mới | Add column via assign | Khi thêm cột tính toán | Biểu thức |
| .astype('float') | Ép kiểu dữ liệu | Convert data type | Khi cần chuẩn hóa kiểu dữ liệu | 'int', 'float', 'str' |

**📈 5. Thống kê, nhóm dữ liệu / Aggregation & Grouping**

| **Lệnh** | **Giải thích (Việt)** | **English Explanation** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| .groupby('col') | Gom nhóm theo cột | Group by column | Khi cần tính toán theo nhóm | Cột nhóm |
| .agg({'col':'mean'}) | Hàm tổng hợp | Aggregate with function | Tính trung bình, tổng,... | dict: {col: func} |
| .mean(), .sum(), .count() | Tính trung bình, tổng, đếm | Mean, sum, count per group | Thống kê cơ bản | — |
| .pivot\_table(values, index, columns, aggfunc) | Bảng tổng hợp (giống Excel Pivot) | Pivot table summary | Khi cần thống kê đa chiều | aggfunc='mean' mặc định |
| .crosstab(index, columns) | Bảng chéo tần suất | Cross-tabulation | Khi cần đếm tần suất theo 2 chiều | — |

**🔗 6. Gộp & nối dữ liệu / Merge, Join, Concatenate**

| **Lệnh** | **Giải thích (Việt)** | **English Explanation** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| pd.concat([df1, df2]) | Nối bảng (theo hàng/cột) | Concatenate DataFrames | Khi cần nối 2 bảng | axis=0/1 |
| pd.merge(df1, df2, on='key') | Gộp 2 bảng theo cột chung | Merge two tables by key | Khi kết hợp dữ liệu có khóa chung | how='inner'/'left' |
| df.join(df2) | Gộp theo index | Join on index | Khi muốn nối theo index | — |

**🧮 7. Hàm thống kê cơ bản / Descriptive Stats**

| **Lệnh** | **Giải thích** | **English** | **Khi nào dùng** | **Ghi chú** |
| --- | --- | --- | --- | --- |
| .mean(), .median(), .mode() | Trung bình, trung vị, mốt | Mean, Median, Mode | Khi phân tích mô tả | — |
| .min(), .max() | Giá trị nhỏ/lớn nhất | Minimum/Maximum | — | — |
| .var(), .std() | Phương sai, độ lệch chuẩn | Variance, Std. Dev | — | — |
| .corr() | Hệ số tương quan | Correlation matrix | Khi phân tích mối quan hệ | — |

**🔍 8. Xử lý dữ liệu khuyết / Missing Data Handling**

| **Lệnh** | **Giải thích** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| .isna() / .notna() | Kiểm tra giá trị thiếu | Check missing values | Khi cần xác định NaN | — |
| .fillna(value) | Thay giá trị NaN | Fill missing values | Dọn dữ liệu | method='ffill' để lặp trước |
| .dropna() | Xóa hàng chứa NaN | Drop rows with missing | Khi NaN không thể thay | axis=0/1 |

**🕹 9. Sắp xếp & chỉ số / Sorting & Indexing**

| **Lệnh** | **Giải thích** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| .sort\_values(by='col') | Sắp xếp theo cột | Sort by column | Khi cần thứ tự tăng/giảm | ascending=True/False |
| .set\_index('col') | Đặt cột làm chỉ số | Set index column | Khi cần index cụ thể | — |
| .reset\_index() | Trả index về mặc định | Reset index | Sau khi groupby hoặc set\_index | drop=True |

**🧾 10. Xuất dữ liệu / Exporting**

| **Lệnh** | **Giải thích** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| .to\_csv('file.csv', index=False) | Xuất dữ liệu ra file CSV | Export to CSV | Lưu kết quả xử lý | index=False bỏ cột index |
| .to\_excel('file.xlsx') | Xuất file Excel | Export to Excel | Khi cần nộp file Excel | sheet\_name, index |
| .to\_json() | Xuất dữ liệu JSON | Export JSON | Dùng cho API | — |