Pandas Documentation — Reference Guide

Reading & Writing Data

```python
import pandas as pd

# Read
df = pd.read_csv("data.csv")
df = pd.read_excel("data.xlsx")
df = pd.read_json("data.json")
df = pd.read_sql("SELECT * FROM users", connection)

# Write
df.to_csv("output.csv", index=False)
df.to_excel("output.xlsx", index=False)
df.to_json("output.json", orient="records")
```

Data Cleaning Cheat Sheet

Task	Code
Drop duplicates	`df.drop_duplicates()`
Rename columns	`df.rename(columns={"old": "new"})`
Change dtype	`df["col"].astype(int)`
Replace values	`df["col"].replace("old", "new")`
Drop columns	`df.drop(columns=["col1", "col2"])`
Reset index	`df.reset_index(drop=True)`

Aggregation

```python
# Single aggregation
df.groupby("category")["sales"].sum()

# Multiple aggregations
df.groupby("category").agg(
    total_sales=("sales", "sum"),
    avg_price=("price", "mean"),
```

```
    count=("id", "count")
)
```

Pivot Tables

```
pivot = df.pivot_table(
    values="sales",
    index="region",
    columns="product",
    aggfunc="sum",
    fill_value=0
)
```

Key Takeaways

1. Pandas can read from CSV, Excel, JSON, SQL, and more.
2. Chain methods for clean, readable data transformations.
3. Use `.agg()` with named aggregations for clear output.
4. Pivot tables are powerful for cross-tabulation summaries.

LearnQuest — Python for Data Science