Pandas Cheat Sheet: Key Basics and Operations

---

1. Basic Setup

import pandas as pd

---

2. Data Structures

DataFrame: Tabular data structure with rows and columns.

df = pd.DataFrame(data)

Series: One-dimensional labeled array.

series = pd.Series(data)

---

3. Creating a DataFrame

data = {'Column1': [1, 2, 3], 'Column2': [4, 5, 6]}

df = pd.DataFrame(data)

---

4. Basic Inspection

- df.head(n): First n rows (default: 5).
- df.tail(n): Last n rows.
- df.shape: Tuple with (#rows, #columns).
- df.info(): Summary of DataFrame (types, non-null values).
- df.describe(): Statistics summary for numerical columns.

---

5. Selecting Data

Selecting Columns

- df['column_name']: Select a single column.
- df[['col1', 'col2']]: Select multiple columns.

Selecting Rows

- df.loc[row_label]: Select by row label.
- df.iloc[row_index]: Select by row index.
- df.loc[start:end]: Slice rows by labels.
- df.iloc[start:end]: Slice rows by index.

---

6. Filtering Data

- df[df['column'] > value]: Filter rows based on a condition.
- df[(df['col1'] > value1) & (df['col2'] < value2)]: Combine conditions.

---

7. Sorting

- df.sort_values(by='column_name', ascending=True): Sort by a column.
- df.sort_values(by=['col1', 'col2'], ascending=[True, False]): Sort by multiple columns.

---

8. Handling Missing Data

- df.isnull(): Check for missing values.
- df.notnull(): Check for non-missing values.
- df.dropna(): Drop rows with missing values.
- df.fillna(value): Fill missing values with specified value.

---

9. Aggregations and Statistics

- df['col'].mean(): Mean of a column.
- df['col'].sum(): Sum of a column.
- df['col'].min() / df['col'].max(): Minimum/maximum values.
- df['col'].std(): Standard deviation.
- df['col'].count(): Count of non-null values.

---

10. GroupBy Operations

df.groupby('column').agg({'col1': 'mean', 'col2': 'sum'})

- groupby('column').sum()
- groupby('column').mean()
- groupby('column').count()

---

11. Apply Functions

- df['col'].apply(function): Apply a function to a column.
- df.apply(lambda row: row['col1'] + row['col2'], axis=1): Apply across rows.

---

12. Merging and Joining

Merging DataFrames

df1.merge(df2, how='inner', on='key_column')

- how='inner': Inner join (default).
- how='outer': Outer join.
- how='left': Left join.
- how='right': Right join.

Concatenation

pd.concat([df1, df2], axis=0) # Vertical concatenation

pd.concat([df1, df2], axis=1) # Horizontal concatenation

---

13. Pivot Tables

df.pivot_table(values='value_col', index='index_col', columns='col_col', aggfunc='mean')

---

14. Reshaping Data

- df.melt(id_vars='id', value_vars=['col1', 'col2']): Unpivot DataFrame.
- df.pivot(index='row', columns='col', values='val'): Pivot data.

---

15. Exporting Data

- df.to_csv('filename.csv', index=False): Export to CSV.
- df.to_excel('filename.xlsx', index=False): Export to Excel.
- df.to_json('filename.json'): Export to JSON.

---

16. Common Operations

Adding a New Column

df['new_col'] = df['col1'] + df['col2']

Renaming Columns

df.rename(columns={'old_name': 'new_name'}, inplace=True)

Dropping Columns

df.drop(['col1', 'col2'], axis=1, inplace=True)

Resetting the Index

df.reset_index(drop=True, inplace=True)