🐼 Pandas Cheat Sheet

📌 What is Pandas?

Pandas is a Python library for working with tabular data (like spreadsheets, SQL tables, CSV files).

* **Series** → 1D labeled array (like a column)
* **DataFrame** → 2D labeled table (like a spreadsheet)

📌 How To Import Pandas?

python

import pandas as pd

*The convention is always pd.*

📌 How to Create Data?

You can create data from Python lists, dictionaries, or external files.

python

# Series

s = pd.Series([10, 20, 30], index=['a', 'b', 'c'])

# DataFrame from dict

df = pd.DataFrame({

'Name': ['Alice', 'Bob', 'Charlie'],

'Age': [25, 30, 35],

'City': ['NY', 'LA', 'Chicago']

})

📌 How to Read and Write Data?

python

# Read from CSV

df = pd.read\_csv('data.csv')

# Save to CSV

df.to\_csv('output.csv', index=False)

# Other formats

df.to\_excel('data.xlsx', index=False) # Excel

df.to\_json('data.json') # JSON

📌 How to Explore the Data?

python

df.head() # First 5 rows

df.tail() # Last 5 rows

df.shape # (rows, columns)

df.columns # Column names

df.info() # Data types & non-null counts

df.describe() # Summary statistics

📌 How to Select the Data?

python

# Column

df['Name'] # Series

df[['Name', 'Age']] # Multiple columns

# Row by index

df.loc[0] # By label/index

df.iloc[0] # By position

# Row + Column

df.loc[0, 'Name'] # First row, "Name" column

📌 How to Filter the Data?

python

# Filter by condition

df[df['Age'] > 30]

# Multiple conditions

df[(df['Age'] > 25) & (df['City'] == 'NY')]

📌 How to Modify the Data?

python

# Add new column

df['Age in 5 Years'] = df['Age'] + 5

# Update values

df.loc[df['Name'] == 'Alice', 'Age'] = 26

# Drop column/row

df.drop(columns=['City'], inplace=True)

df.drop(index=0, inplace=True)

📌 How to Aggregate and Group Data?

python

df['Age'].mean() # Average age

df.groupby('City')['Age'].mean() # Avg age per city

📌 How to Sort the Data?

python

df.sort\_values(by='Age', ascending=False)

df.sort\_index()

📌 How to Handle Missing Data?

python

df.isna() # Check missing

df.fillna(0) # Replace missing with 0

df.dropna() # Drop missing rows

📌 How to Merge Data?

python

pd.merge(df1, df2, on='ID') # SQL-style join

pd.concat([df1, df2], axis=0) # Stack rows

📌 What Kind of Jobs Can Pandas Fulfill?

Pandas is designed for data manipulation, cleaning, and analysis.  
It supports many job roles, such as:

* **Data Analyst** – analyzing trends, creating reports, making dashboards.
* **Data Scientist** – preparing datasets for machine learning, statistical analysis.
* **Business Analyst** – turning raw business data into insights.
* **Financial Analyst** – processing stock data, sales reports, or budgets.
* **Researcher** – cleaning survey data, preparing datasets for experiments.
* **Data/ETL Engineer** – moving and transforming data between systems.

*Basically: if your work involves structured data (tables, CSVs, Excel, SQL), pandas can help.*

📌 Who Uses Pandas?

* Tech companies – log analysis, A/B testing results, user analytics.
* Banks/Finance – stock data, risk modeling, fraud detection.
* Healthcare – patient records, medical research datasets.
* Government & NGOs – census data, economic reports, social research.
* Academia – cleaning survey data, statistical prep for studies.
* E-commerce – product sales analytics, customer behavior analysis.

💡 *Pandas is like Excel for programmers – anywhere Excel is used, pandas can replace or enhance it.*

📌 What Can I Do With Pandas?

🔹 Data Cleaning

* Remove duplicates
* Fill in missing values
* Rename or reorder columns
* Fix wrong data types

🔹 Data Transformation

* Merge/join multiple datasets
* Group data and summarize (totals, averages, counts)
* Pivot and reshape tables
* Apply custom functions

🔹 Data Analysis

* Calculate statistics (mean, median, std, correlations)
* Time-series analysis
* Identify outliers
* Filter by conditions

🔹 Data Input/Output

* Read/write CSV, Excel, JSON, SQL, Parquet, etc.
* Handle millions of rows faster than Excel
* Automate repetitive reporting

🔹 Visualization (with matplotlib/seaborn)

* Quick charts (bar, line, scatter)
* Time-series plots
* Histograms