Pandas Cheat Sheet (Quick & Practical)

1. Importing & Basics

```
import pandas as pd
```

---

2. Creating DataFrames

From dict

```
df = pd.DataFrame({"name": ["A", "B"], "age": [20, 25]})
```

From List of Lists

```
df = pd.DataFrame([[1,2],[3,4]], columns=["x", "y"])
```

From CSV

```
df = pd.read_csv("file.csv")
```

---

3. Viewing Data

```
df.head()        # first 5 rows
df.tail()        # last 5 rows
df.info()        # summary
df.describe()    # stats summary
```

---

4. Selecting Data

Columns

```
df["col"]
df[["col1", "col2"]]
```

Rows

```
df.loc[5]          # by label
df.iloc[5]         # by index
```

Row + Column

```
df.loc[0, "age"]
df.iloc[0, 1]
```

---

5. Filtering Rows

```
df[df["age"] > 20]
df[(df.age > 20) & (df.city == "Hyd")]
```

---

6. Adding / Updating Columns

```
df["new"] = df["age"] * 2
```

---

7. Dropping

```
df.drop("col", axis=1, inplace=True)      # drop column
df.drop(3, axis=0, inplace=True)           # drop row
```

---

8. Missing Values

```
df.isna()
df.dropna()
df.fillna(0)
```

---

9. Sorting

```
df.sort_values("age")
df.sort_values(["age", "name"], ascending=[True, False])
```

---

10. Group By

```
df.groupby("city")["age"].mean()
df.groupby("city").agg({"age": ["mean", "max"]})
```

---

11. Merging / Joining

```
pd.merge(df1, df2, on="id", how="inner")
pd.concat([df1, df2], axis=0)   # rows
```

---

12. Apply Functions

```
df["age2"] = df["age"].apply(lambda x: x + 1)
```

---

13. Column Renaming

```
df.rename(columns={"age": "Age"}, inplace=True)
```

---

14. Index Reset / Set

```
df.reset_index(drop=True)
df.set_index("id")
```

---

15. Exporting

```
df.to_csv("out.csv", index=False)
df.to_excel("out.xlsx", index=False)
```

---

16. Common Useful Tricks

```
df["age"].unique()
df["age"].value_counts()
df.sample(5)
df.nunique()
```

---

17. Date Handling

```
df["date"] = pd.to_datetime(df["date"])
df["date"].dt.year
df["date"].dt.month
```

---

18. Copy vs View Warning

```
# Safe assignment
df = df.copy()
```

---

19. Pivot Tables

```
pd.pivot_table(df, values="age", index="city", aggfunc="mean")
```

---

20. MultiIndex

```
df.set_index(["state", "city"]).loc["AP"]
```