📆 DAY 10 — Pandas Basics

Goal: Read, inspect, and manipulate tabular data

1️⃣ Why Pandas?

Real ML data comes as:

- CSV files
- Excel sheets
- Tables (rows & columns)

Pandas gives you:

✔️ DataFrames

✔️ Easy filtering

✔️ Fast analysis

2️⃣ Importing Pandas

```
import pandas as pd
```

Standard convention:

- `pd` = pandas

3️⃣ Creating a DataFrame

Form Dictionary

```
data = {
    "name": ["Alice", "Bob", "Charlie"],
    "age": [25, 30, 35],
    "score": [85, 90, 88]
}

df = pd.DataFrame(data)
print(df)
```

```
# output
      name  age  score
0    Alice   25     85
1      Bob   30     90
2  Charlie   35     88
```

4️⃣ Reading a CSV File

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

This is how **90% of ML datasets start**.

5️⃣ Inspecting Data (Very Important)

```
df.head()      # first 5 rows
df.tail()      # last 5 rows
df.shape       # (rows, columns)
df.columns     # column names
df.info()      # data types & nulls
df.describe()  # statistics
```

6️⃣ Selecting Data

Column

```
df["age"]
```

Multiple columns

```
df[["age", "score"]]
```

Row selection

```
df.iloc[0]    # by index
df.loc[0]     # by label
```

7️⃣ Filtering Rows

```
df[df["age"] > 25]
```

Multiple conditions:

```
df[(df["age"] > 25 & (df["score"] > 85)]
```

8️⃣ Handling Missing Values (Basic)

```
df.isnull()
df.isnull().sum()
```

Fill missing values:

```
df["age"].fillna(df["age"].mean(), inplace=True)
```

9️⃣ Add / Remove Columns

Add Column

```
df["passed"] = df["score"] > 85
```

Drop Column

```
df.drop("passed", axis=1, inplace=True)
```

1️⃣0️⃣ Real ML Example

```
X = df[["age","score"]]
y = df["passed"]
```

This is **exactly how ML training data is prepared**.
