Pandas Step-by-Step Guide with Examples

# Step 1: Introduction to Pandas

```python  
# Install pandas (run in terminal)  
# pip install pandas  
  
import pandas as pd  
print(pd.\_\_version\_\_) # Check installed version  
```  
✅ Output:  
```  
2.2.2 # (version may vary)  
```

# Step 2: Creating DataFrames

1. From CSV File (First create CSV):  
  
```python  
import pandas as pd  
  
data = {  
 "RollNo": [1, 2, 3],  
 "Name": ["Alice", "Bob", "Charlie"],  
 "Percentage": [85, 72, 91],  
 "Grade": ["A", "B", "A+"]  
}  
df = pd.DataFrame(data)  
df.to\_csv("sample.csv", index=False)  
print("sample.csv created successfully!")  
```  
  
Now read it:  
```python  
df = pd.read\_csv("sample.csv")  
print(df)  
```  
  
✅ Output:  
```  
 RollNo Name Percentage Grade  
0 1 Alice 85 A  
1 2 Bob 72 B  
2 3 Charlie 91 A+  
```

2. From Excel File:  
  
```python  
df = pd.read\_excel("sample.xlsx")  
print(df.head())  
```

3. From Dictionary:  
  
```python  
data = {  
 "RollNo": [1, 2, 3],  
 "Name": ["Alice", "Bob", "Charlie"],  
 "Percentage": [85, 72, 91],  
 "Grade": ["A", "B", "A+"]  
}  
df = pd.DataFrame(data)  
print(df)  
```

4. From List of Lists:  
  
```python  
data = [[1, "Alice", 85], [2, "Bob", 72], [3, "Charlie", 91]]  
df = pd.DataFrame(data, columns=["RollNo", "Name", "Percentage"])  
print(df)  
```

# Step 3: Basic Data Inspection

```python  
print(df.head(2)) # first 2 rows  
print(df.tail(2)) # last 2 rows  
print(df.shape) # rows, columns  
print(df.info()) # summary info  
print(df.describe()) # statistics  
```

# Step 4: Accessing Data

```python  
print(df["Name"]) # single column  
print(df[["Name", "Percentage"]]) # multiple columns  
  
print(df.loc[0]) # row by label  
print(df.loc[0:1, "Name"]) # rows 0-1, Name column  
  
print(df.iloc[0]) # row by index  
print(df.iloc[0:2, 0:2]) # first 2 rows, 2 columns  
```

# Step 5: Data Manipulation & Cleaning

Sorting:  
  
```python  
print(df.sort\_values("Percentage")) # ascending  
print(df.sort\_values("Percentage", ascending=False)) # descending  
```  
  
Adding Column:  
  
```python  
df["Math"] = [90, 80, 95]  
df["Science"] = [85, 70, 88]  
df["Total"] = df["Math"] + df["Science"]  
print(df)  
```  
  
Deleting Column:  
  
```python  
df.drop("Total", axis=1, inplace=True)  
```  
  
Handling Missing Values:  
  
```python  
df.loc[1, "Science"] = None  
print(df.isnull())  
df["Science"].fillna(0, inplace=True)  
```  
  
Removing Duplicates:  
  
```python  
df.loc[3] = [3, "Charlie", 91, 95, 88] # duplicate row  
print(df.duplicated())  
df.drop\_duplicates(inplace=True)  
```

# Step 6: Conditional Selection

```python  
df["Result"] = df["Percentage"].apply(  
 lambda x: "Fail" if x < 35 else "Pass"  
)  
print(df[["Name", "Percentage", "Result"]])  
```  
✅ Output:  
```  
 Name Percentage Result  
0 Alice 85 Pass  
1 Bob 72 Pass  
2 Charlie 91 Pass  
```

# Step 7: Exporting Data

```python  
df.to\_csv("output.csv", index=False)  
df.to\_excel("output.xlsx", index=False)  
```  
  
✅ Output: Creates output.csv and output.xlsx files.