Step-by-Step Guide to Pandas

# Step 1: Introduction to Pandas

1. Pandas is a Python library for data analysis and manipulation.  
2. It is mainly used for handling tabular data (similar to Excel).  
3. Core Pandas data structures:  
 - Series → One-dimensional (single column).  
 - DataFrame → Two-dimensional (multiple columns, like a table).  
4. Installation command:  
 pip install pandas

# Step 2: Creating DataFrames

1. From CSV File:  
 import pandas as pd  
 df = pd.read\_csv("data.csv")  
  
2. From Excel File:  
 df = pd.read\_excel("data.xlsx")  
  
3. From Python Dictionary:  
 data = {  
 "RollNo": [1, 2, 3],  
 "Name": ["Alice", "Bob", "Charlie"],  
 "Percentage": [85, 72, 91],  
 "Grade": ["A", "B", "A+"]  
 }  
 df = pd.DataFrame(data)  
  
4. From List of Lists:  
 data = [[1, "Alice", 85], [2, "Bob", 72]]  
 df = pd.DataFrame(data, columns=["RollNo", "Name", "Percentage"])

# Step 3: Basic Data Inspection

1. df.head() # first 5 rows  
2. df.head(10) # first 10 rows  
3. df.tail() # last rows  
4. df.shape # (rows, columns)  
5. df.info() # summary info  
6. df.describe() # statistical summary

# Step 4: Accessing Data (Indexing & Slicing)

1. Single Column: df["Name"]  
2. Multiple Columns: df[["Name", "Percentage"]]  
3. Label-based indexing (.loc):  
 df.loc[0] # first row  
 df.loc[0:2] # rows 0 to 2  
 df.loc[0:2, "Name"] # Name column for rows 0 to 2  
  
4. Position-based indexing (.iloc):  
 df.iloc[0] # first row  
 df.iloc[0:3, 0:2] # first 3 rows, first 2 columns

# Step 5: Data Manipulation & Cleaning

1. Sorting:  
 df.sort\_values("Percentage")  
 df.sort\_values("Percentage", ascending=False)  
  
2. Adding a new column:  
 df["Total"] = df["Math"] + df["Science"] + df["English"]  
  
3. Deleting a column:  
 df.drop("Total", axis=1, inplace=True)  
  
4. Handling Missing Values:  
 df.isnull()  
 df.fillna(0)  
 df.dropna()  
  
5. Removing Duplicates:  
 df.duplicated()  
 df.drop\_duplicates(inplace=True)

# Step 6: Conditional Selection & Transformation

Example: Assigning grades based on percentage:  
  
def grade(x):  
 if x < 35: return "Fail"  
 elif x < 50: return "Pass"  
 elif x < 60: return "C"  
 elif x < 75: return "B"  
 else: return "Distinction"  
  
df["Grade"] = df["Percentage"].apply(grade)

# Step 7: Exporting Data

1. Export to CSV:  
 df.to\_csv("output.csv", index=False)  
  
2. Export to Excel:  
 df.to\_excel("output.xlsx", index=False)