**Combining DataFrames**

Topics covered:

* concat()
* merge()
* join()

**1. concat() – Concatenating DataFrames (Stacking)**

**➤ Vertical Stacking (Row-wise)**

import pandas as pd

df1 = pd.DataFrame({'ID': [1, 2], 'Name': ['Arihant', 'Ravi']})

df2 = pd.DataFrame({'ID': [3, 4], 'Name': ['Meena', 'Preeti']})

result = pd.concat([df1, df2], ignore\_index=True)

**➤ Horizontal Stacking (Column-wise)**

df3 = pd.DataFrame({'Math': [90, 80]})

df4 = pd.DataFrame({'Science': [85, 95]})

result = pd.concat([df3, df4], axis=1)

**➤ Add hierarchical keys (multi-index)**

result = pd.concat([df1, df2], keys=['Batch1', 'Batch2'])

**2. merge() – SQL-style Joins**

**➤ Inner Join (only matching rows)**

df1 = pd.DataFrame({'ID': [1, 2], 'Name': ['A', 'B']})

df2 = pd.DataFrame({'ID': [1, 3], 'Score': [90, 80]})

result = pd.merge(df1, df2, on='ID', how='inner')

**➤ Left Join**

result = pd.merge(df1, df2, on='ID', how='left')

**➤ Right Join**

result = pd.merge(df1, df2, on='ID', how='right')

**➤ Outer Join (all rows)**

result = pd.merge(df1, df2, on='ID', how='outer')

**➤ Merge on multiple columns**

pd.merge(df1, df2, on=['ID', 'Name'], how='inner')

**3. join() – Join by index**

df1 = pd.DataFrame({'Name': ['A', 'B']}, index=[1, 2])

df2 = pd.DataFrame({'Score': [90, 95]}, index=[1, 2])

result = df1.join(df2)

**Reshaping & Pivoting Data**

Topics covered:

* pivot()
* pivot\_table()
* melt()
* stack() and unstack()
* transpose() or .T

**1. pivot() – Reshape by setting index/columns/values**

df = pd.DataFrame({

'ID': [1, 2, 1, 2],

'Subject': ['Math', 'Math', 'Science', 'Science'],

'Score': [90, 85, 95, 80]

})

pivoted = df.pivot(index='ID', columns='Subject', values='Score')

**2. pivot\_table() – Similar to pivot but with aggregation**

df = pd.DataFrame({

'Department': ['HR', 'IT', 'IT', 'HR'],

'Gender': ['F', 'M', 'F', 'F'],

'Salary': [40000, 50000, 45000, 42000]

})

pivot\_table = df.pivot\_table(index='Department', columns='Gender', values='Salary', aggfunc='mean')

**3. melt() – Unpivot your data (columns to rows)**

df = pd.DataFrame({

'ID': [1, 2],

'Math': [85, 90],

'Science': [80, 95]

})

melted = pd.melt(df, id\_vars='ID', value\_vars=['Math', 'Science'], var\_name='Subject', value\_name='Marks')

**4. stack() – Columns to index (longer format)**

df = pd.DataFrame({

'Name': ['Arihant', 'Ravi'],

'Math': [90, 80],

'Science': [85, 95]

}, index=['Student1', 'Student2'])

stacked = df.stack()

**5. unstack() – Index to columns (wider format)**

unstacked = stacked.unstack()

**6. transpose() or .T – Flip rows and columns**

df = pd.DataFrame({

'Name': ['A', 'B'],

'Score': [88, 92]

})

transposed = df.T

-------------------------------------------------------------------------------------------------------------

**Practice Questions**

**Q1. Vertical Concatenation**

Create two DataFrames:

* df1 with columns Name, Age
* df2 with columns Name, Age  
  Now concatenate them vertically using concat().

**Q2. Horizontal Concatenation**

Use two DataFrames:

* df1: Student\_ID, Math
* df2: Science, English  
  Stack them horizontally using concat().

**Q3. Inner Merge**

Merge two DataFrames orders and customers on customer\_id using an inner join.

**Q4. Left Join**

Given DataFrames employees and departments, merge them using a left join on dept\_id.

**Q5. Join using Index**

Use join() to combine:

* df1: Employee Name (indexed by ID)
* df2: Salary (indexed by ID)

**Q6. pivot()**

Given:

df = pd.DataFrame({

'Name': ['A', 'B', 'A', 'B'],

'Subject': ['Math', 'Math', 'Science', 'Science'],

'Score': [90, 85, 95, 80]

})

Use pivot() to convert it into wide format.

**Q7. pivot\_table() with aggregation**

Using the same DataFrame, create a pivot table that calculates average scores for each Name.

**Q8. melt()**

Given a DataFrame:

df = pd.DataFrame({

'ID': [1, 2],

'Math': [90, 85],

'Science': [88, 92]

})

Use melt() to convert subject columns into rows.

**Q9. stack() and unstack()**

Create a DataFrame with multi-level index and apply stack() and unstack().

**Q10. transpose()**

Transpose a DataFrame having columns Product, Price, Quantity.

**3 Mini Projects with Code**

**Project 1: Student Score Analyzer**

**Goal**: Combine marks from different subjects and reshape to get subject-wise scores per student.

import pandas as pd

math = pd.DataFrame({'ID': [1, 2], 'Math': [85, 90]})

science = pd.DataFrame({'ID': [1, 2], 'Science': [78, 88]})

english = pd.DataFrame({'ID': [1, 2], 'English': [92, 81]})

# Merge all subject scores

merged = math.merge(science, on='ID').merge(english, on='ID')

# Melt to long format

long\_df = pd.melt(merged, id\_vars='ID', var\_name='Subject', value\_name='Marks')

# Pivot to see per-student subject scores

pivot\_df = long\_df.pivot(index='ID', columns='Subject', values='Marks')

print(pivot\_df)

**Project 2: Sales Report Dashboard**

**Goal**: Merge order and customer data, summarize monthly sales using pivot table.

orders = pd.DataFrame({

'OrderID': [1, 2, 3],

'CustomerID': [101, 102, 101],

'Amount': [250, 150, 300],

'Month': ['Jan', 'Jan', 'Feb']

})

customers = pd.DataFrame({

'CustomerID': [101, 102],

'Name': ['Arihant', 'Ravi']

})

# Merge customer info

data = pd.merge(orders, customers, on='CustomerID')

# Pivot: total sales per customer per month

sales\_summary = pd.pivot\_table(data, index='Name', columns='Month', values='Amount', aggfunc='sum')

print(sales\_summary)

**Project 3: Employee Department Join**

**Goal**: Join employee data with department data using index.

employees = pd.DataFrame({

'Name': ['Anu', 'Rahul'],

'Dept\_ID': [101, 102]

}).set\_index('Dept\_ID')

departments = pd.DataFrame({

'Dept\_Name': ['HR', 'IT']

}, index=[101, 102])

# Join based on index

result = employees.join(departments)

print(result)