45. Joining DataFrames

Task:

• a) Load the orders\_data.csv file into a Pandas DataFrame. This file contains:

Order ID, Customer ID, and Order Date.

• b) Load the customer\_info.csv file into another Pandas DataFrame. This file contains:

Customer ID, Name, Email, and Phone Number.

• c) Merge the two DataFrames on the Customer ID column to create a single DataFrame that includes both order and customer details.

• d) Calculate the average time it takes for a customer to place another order after their first one (i.e., time between consecutive orders).

CODE:

import pandas as pd

orders\_df = pd.read\_csv("order\_data.csv")

customers\_df = pd.read\_csv("customer\_info.csv")

merged\_df = pd.merge(orders\_df, customers\_df, on="Customer ID", how="inner")

print(merged\_df)

merged\_df['Order Date'] = pd.to\_datetime(merged\_df['Order Date'])

merged\_df = merged\_df.sort\_values(by=['Customer ID', 'Order Date'])

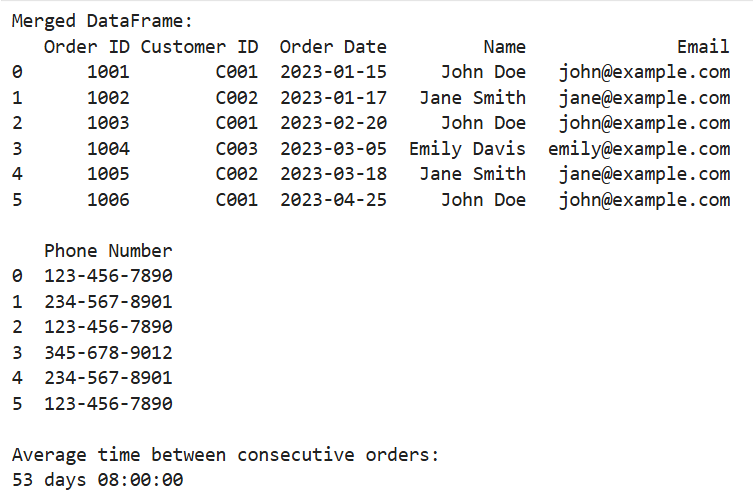
merged\_df['Time Diff'] = merged\_df.groupby('Customer ID')['Order Date'].diff()

avg\_time\_diff = merged\_df['Time Diff'].dropna().mean()

print("\nAverage time between consecutive orders:")

print(avg\_time\_diff)

OUTPUT:



Dataset:

Customer data:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Customer ID | Name | Email | Phone Number | |  |
| C001 | John Doe | john@example.com | 123-456-7890 | |  |
| C002 | Jane Smith | jane@example.com | 234-567-8901 | |  |
| C003 | Emily Davis | emily@example.com | 345-678-9012 | |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Order details;

|  |  |  |  |
| --- | --- | --- | --- |
| Order ID | Customer ID | Order Date | |
| 1001 | C001 | 1/15/2023 |  |
| 1002 | C002 | 1/17/2023 |  |
| 1003 | C001 | 2/20/2023 |  |
| 1004 | C003 | 3/5/2023 |  |
| 1005 | C002 | 3/18/2023 |  |
| 1006 | C001 | 4/25/2023 |  |
|  |  |  |  |