# Joining DataFrames Report

## a) Load orders\_data.csv

Sample data:

Order ID Customer ID Order Date  
 1 101 2025-01-01  
 2 102 2025-01-03  
 3 101 2025-01-10  
 4 103 2025-01-15  
 5 102 2025-01-20  
 6 101 2025-01-25

## b) Load customer\_info.csv

Sample data:

Customer ID Name Email Phone Number  
 101 Alice Smith alice@example.com 123-456-7890  
 102 Bob Johnson bob@example.com 234-567-8901  
 103 Charlie Lee charlie@example.com 345-678-9012

## c) Merge on Customer ID

merged\_df = pd.merge(orders\_df, customers\_df, on='Customer ID')

Merged data:

Order ID Customer ID Order Date Name Email Phone Number  
 1 101 2025-01-01 Alice Smith alice@example.com 123-456-7890  
 3 101 2025-01-10 Alice Smith alice@example.com 123-456-7890  
 6 101 2025-01-25 Alice Smith alice@example.com 123-456-7890  
 2 102 2025-01-03 Bob Johnson bob@example.com 234-567-8901  
 5 102 2025-01-20 Bob Johnson bob@example.com 234-567-8901  
 4 103 2025-01-15 Charlie Lee charlie@example.com 345-678-9012

## d) Average Time Between Orders

merged\_df\_sorted = merged\_df.sort\_values(by=['Customer ID', 'Order Date'])  
merged\_df\_sorted['Time Between Orders'] = merged\_df\_sorted.groupby('Customer ID')['Order Date'].diff()  
avg\_time\_between\_orders = merged\_df\_sorted['Time Between Orders'].dropna().mean()

Data with time differences:

Order ID Customer ID Order Date Name Email Phone Number Time Between Orders  
 1 101 2025-01-01 Alice Smith alice@example.com 123-456-7890 NaT  
 3 101 2025-01-10 Alice Smith alice@example.com 123-456-7890 9 days  
 6 101 2025-01-25 Alice Smith alice@example.com 123-456-7890 15 days  
 2 102 2025-01-03 Bob Johnson bob@example.com 234-567-8901 NaT  
 5 102 2025-01-20 Bob Johnson bob@example.com 234-567-8901 17 days  
 4 103 2025-01-15 Charlie Lee charlie@example.com 345-678-9012 NaT

Average time between orders: 13 days 16:00:00