Pandas Advanced Project (With Datasets)

Project 1: Data Cleaning Mastery

Goal: Learn how to clean messy real-world data using Pandas.

Dataset (customer_data.csv)

```
CustomerID,Name,Age,Gender,City,Spend
1,Amit,25,M,Delhi,5000
2,Riya,,F,Mumbai,7000
3,Arjun,45,M,Delhi,
4,Neha,29,F,Bangalore,6500
5,Amit,25,M,Delhi,5000
6,Rahul,-5,M,Mumbai,4000
7,Pooja,120,F,Delhi,9000
```

Tasks

1   Detect and handle missing values
2   Remove duplicate rows
3   Fix invalid age values (negative or > 100)
4   Standardize city names
5   Save the cleaned dataset to a new CSV file

Project 2: Multi-File Data Analysis

Goal: Merge multiple CSV files and perform meaningful analysis.

Dataset 1 (customers.csv)

```
CustomerID,Name,City
1,Amit,Delhi
2,Riya,Mumbai
3,Arjun,Delhi
```

Dataset 2 (orders.csv)

```
OrderID,CustomerID,ProductID,Quantity
101,1,1001,2
102,2,1002,1
103,1,1003,5
```

Dataset 3 (products.csv)

```
ProductID,Product,Price
1001,Laptop,55000
1002,Mobile,20000
1003,Chair,3000
```

Tasks

1 Merge all three datasets
2 Calculate total bill amount per order
3 Calculate total spending per customer
4 Identify top customers by spending

Project 3: Time Series Analysis

Goal: Perform advanced analysis using date and time data.

Dataset (sales_timeseries.csv)

```
Date,Sales
2023-01-01,5000
2023-01-02,7000
2023-01-03,6000
2023-02-01,8000
2023-02-02,7500
```

Tasks

1 Convert Date column to datetime format
2 Set Date as index
3 Calculate monthly sales using resampling
4 Calculate rolling average
5 Calculate sales growth rate

Project 4: Pivot Tables and Reshaping

Goal: Summarize and reshape data using pivot tables.

Dataset (regional_sales.csv)

```
Region,Product,Sales
North,Laptop,50000
North,Mobile,30000
South,Laptop,40000
South,Mobile,35000
```

Tasks

1 Create pivot tables
2 Compare product sales by region
3 Calculate percentage contribution

Project 5: Performance Optimization

Goal: Write fast and memory-efficient Pandas code.
Tasks

1   Compare apply() vs vectorized operations
2   Simulate large datasets
3   Optimize memory usage using proper data types

Final Capstone Project: Business Sales Dashboard

This project combines all Pandas concepts into one real-world scenario.

Features

1   Merge multiple CSV files
2   Build a complete data cleaning pipeline
3   Analyze time series sales trends
4   Create pivot-based reports
5   Export final analysis results

Completing all these projects honestly will make you advanced and confident in Pandas.