Pandas Exercises

Question 1:

Given the following DataFrame representing student data:

import pandas as pd

data = {

'Name': ['John', 'Mary', 'Alex', 'Sarah', 'Jack'],

'Math\_Score': [85, 78, 92, 95, 80],

'Science\_Score': [90, 92, 88, 89, 85],

'History\_Score': [88, 85, 91, 92, 86]

}

df = pd.DataFrame(data)

a) Calculate the total number of students in the DataFrame.

b) Calculate the average score for each subject (Math, Science, and History).

c) Find the student with the highest total score (sum of all three subjects).

Question 2:

Given the following DataFrame representing sales data:

data = {

'Date': ['2023-07-01', '2023-07-02', '2023-07-03', '2023-07-04', '2023-07-05'],

'Product': ['Laptop', 'Phone', 'Tablet', 'Laptop', 'Phone'],

'Sales': [1000, 500, 800, 1200, 600],

'Profit': [200, 100, 150, 250, 120]

}

df = pd.DataFrame(data)

a) Calculate the total sales and total profit for each product.

b) Find the product with the highest total sales and highest total profit.

c) Add a new column "Profit\_Percentage" that contains the percentage of profit for each sale.

Question 3:

Given the following DataFrame representing information about employees in a company:

data = {

'Employee\_ID': [1001, 1002, 1003, 1004, 1005],

'Name': ['John', 'Sarah', 'Alex', 'Emily', 'Jack'],

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

'Salary': [50000, 60000, 55000, 58000, 52000]

}

df = pd.DataFrame(data)

a) Calculate the average salary for each department and display the results as a new DataFrame.

b) Find the department with the highest average salary.

c) Add a new column "Salary\_Category" based on the following conditions:

High: Salary > average salary for the department.

Low: Salary <= average salary for the department.

Question 4:

Given the following DataFrame representing sales data for different regions:

data = {

'Region': ['North', 'South', 'East', 'West', 'North', 'West'],

'Sales': [1000, 800, 1200, 900, 1100, 750],

'Profit': [200, 150, 250, 180, 220, 130]

}

df = pd.DataFrame(data)

a) Calculate the total sales and total profit for each region.

b) Find the region with the highest total sales and highest total profit.

c) Display the number of occurrences of each region in the DataFrame.

Question 5:

Given the following DataFrame representing student scores in different subjects:

data = {

'Name': ['John', 'Mary', 'Alex', 'Sarah', 'Jack'],

'Math\_Score': [85, 78, 92, 95, 80],

'Science\_Score': [90, 92, 88, 89, 85],

'History\_Score': [88, 85, 91, 92, 86]

}

df = pd.DataFrame(data)

a) Calculate the average score for each student.

b) Find the student with the highest average score.

c) Add a new row for a student "Ella" with scores [88, 84, 90] in Math, Science, and History, respectively.

Question 6:

Given the following DataFrame representing sales data for different products:

data = {

'Product': ['Laptop', 'Phone', 'Tablet', 'Laptop', 'Phone'],

'Sales': [1000, 500, 800, 1200, 600],

'Profit': [200, 100, 150, 250, 120]

}

df = pd.DataFrame(data)

a) Calculate the total sales and total profit for each product.

b) Find the product with the highest total sales and highest total profit.

c) Calculate the profit percentage for each product (Profit / Sales \* 100) and add it as a new column.

Question 7:

Given the following DataFrame representing employee data:

data = {

'Name': ['John', 'Sarah', 'Alex', 'Emily', 'Jack'],

'Age': [25, 30, 22, 27, 29],

'Gender': ['Male', 'Female', 'Male', 'Female', 'Male'],

'Salary': [50000, 60000, 55000, 58000, 52000]

}

df = pd.DataFrame(data)

a) Calculate the average age and average salary for all employees.

b) Find the employee with the highest salary and display their details.

c) Add a new employee with Name: "Ella", Age: 24, Gender: "Female", Salary: 56000.

Question 8:

Given the following DataFrame representing sales data for different products:

data = {

'Product': ['Laptop', 'Phone', 'Tablet', 'Laptop', 'Phone'],

'Sales': [1000, 500, 800, 1200, 600],

'Profit': [200, 100, 150, 250, 120]

}

df = pd.DataFrame(data)

a) Calculate the total sales and total profit for each product.

b) Find the product with the highest total sales and highest total profit.

c) Add a new column "Profit\_Percentage" that contains the percentage of profit for each sale.

Question 9:

Given the following DataFrame representing sales data for different regions:

data = {

'Region': ['North', 'South', 'East', 'West', 'North', 'West'],

'Sales': [1000, 800, 1200, 900, 1100, 750],

'Profit': [200, 150, 250, 180, 220, 130]

}

df = pd.DataFrame(data)

a) Calculate the total sales and total profit for each region.

b) Find the region with the highest total sales and highest total profit.

c) Display the number of occurrences of each region in the DataFrame.

Question 10:

Given the following DataFrame representing student data:

data = {

'Name': ['John', 'Mary', 'Alex', 'Sarah', 'Jack'],

'Math\_Score': [85, 78, 92, 95, 80],

'Science\_Score': [90, 92, 88, 89, 85],

'History\_Score': [88, 85, 91, 92, 86]

}

df = pd.DataFrame(data)

a) Calculate the overall average score for each student and add it as a new column "Average\_Score."

b) Find the student with the highest average score and display their details.

c) Add a new row for a student "Ella" with scores [88, 84, 90] in Math, Science, and History, respectively.