Department of Engineering Sciences and Technology,

Second Year Btech in Computer Science

Project Based Learning-Python

Assignment - 15

Name - Paritosh kolwadkar

SRN – 31231313

Roll no – 39

Batch – D2

Problem statement : **Write a program to group data in a DataFrame by a specific column using group by() and compute aggregate statistics (e.g., sum, mean) on grouped data.**

Pre-requisites: Install the Pandas library:

```
pip install pandas
```

Knowledge of DataFrame grouping and aggregation in Pandas.

Code:

```
# Import Pandas

import pandas as pd



# Sample data

data = {

   "Department": ["HR", "IT", "HR", "Finance", "IT", "Finance", "HR"],

   "Employee": ["Alice", "Bob", "Charlie", "David", "Eve", "Frank", "Grace"],

   "Salary": [50000, 60000, 55000, 70000, 65000, 80000, 52000],

   "Bonus": [5000, 6000, 5500, 7000, 6500, 8000, 5200]

}
```

```
# Create a DataFrame
df = pd.DataFrame(data)

# Group data by the 'Department' column
grouped = df.groupby("Department")

# Compute aggregate statistics
salary_stats = grouped["Salary"].agg(["sum", "mean"])
bonus_stats = grouped["Bonus"].agg(["sum", "mean"])

# Display the original DataFrame
print("Original DataFrame:")
print(df)

# Display salary statistics
print("\nSalary Statistics by Department:")
print(salary_stats)

# Display bonus statistics
print("\nBonus Statistics by Department:")
print(bonus_stats)
```

Explanation :

- **Create a DataFrame:**
  - **The `data` dictionary contains columns for `Department`, `Employee`, `Salary`, and `Bonus`.**
  - **The dictionary is converted into a Pandas DataFrame using `pd.DataFrame()`.**
- **Group Data by a Specific Column:**

    - **The `groupby()` method is used to group rows in the DataFrame by the `Department` column.**
- **Compute Aggregate Statistics:**
    - **Aggregation methods such as `sum` and `mean` are applied to the grouped data for the `Salary` and `Bonus` columns.**
    - **The `agg()` method allows multiple aggregations to be performed at once.**
- **Display Results:**
    - **The original DataFrame, salary statistics, and bonus statistics are printed.**

Output:

Original DataFrame:

	Department	Employee	Salary	Bonus
0	HR	Alice	50000	5000
1	IT	Bob	60000	6000
2	HR	Charlie	55000	5500
3	Finance	David	70000	7000
4	IT	Eve	65000	6500
5	Finance	Frank	80000	8000
6	HR	Grace	52000	5200

Salary Statistics by Department:

Department	sum	mean
Finance	150000	75000.000000
HR	157000	52333.333333
IT	125000	62500.000000

Bonus Statistics by Department:

Department	sum	mean
Finance	15000	7500.000000
HR	15700	5233.333333
IT	12500	6250.000000

Output Explained:

- **`groupby()` groups rows based on the unique values in a specified column (e.g., `Department`).**
- **Aggregation functions such as `sum` and `mean` compute statistics for each group.**
- **The program demonstrates how to analyze grouped data for meaningful insights.**

This method is commonly used in data analysis tasks, such as summarizing information by categories or departments.