Training Day – 40

November 12, Tuesday\*

* \*Topic:\* Advanced Groupby Operations
* Applied multiple aggregation functions to grouped data.
* Example: Calculated mean and max for grouped columns.

# Advanced Groupby Operations

Import Libraries

python Copy code import pandas as pd import numpy as np

Create the Dataset

python Copy code data = {

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

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

"Salary": [50000, 60000, 80000, 90000, 70000, 75000, 62000, 88000],

"Bonus": [5000, 7000, 10000, 12000, 8000, 8500, 6000, 11000], "Years": [2, 3, 5, 6, 4, 4, 3, 7]

}

df = pd.DataFrame(data) df

## Output of Dataset

### Department Employee Salary Bonus Years

HR Alice 50000 5000 2

HR Bob 60000 7000 3

IT Charlie 80000 10000 5

IT David 90000 12000 6

Finance Eve 70000 8000 4

Finance Frank 75000 8500 4

HR Grace 62000 6000 3

IT Hank 88000 11000 7

# Applying Advanced Groupby Operations

## 1. Multiple Aggregations

python Copy code

grouped = df.groupby("Department").agg({

"Salary": ["mean", "sum", "max"],

"Bonus": ["sum", "max"],

"Years": ["mean"]

})

print(grouped)

## Output

### Salary Bonus Years

Department mean sum max sum max mean

HR 57333.33 172000 62000 18000 7000 2.67 IT 86000.00 258000 90000 33000 12000 6.00

Finance 72500.00 145000 75000 16500 8500 4.00

#### 2. Custom Aggregation Function

python Copy code def custom\_salary\_range(series): return series.max() - series.min()

grouped\_custom = df.groupby("Department").agg({

"Salary": ["mean", custom\_salary\_range],

"Bonus": "sum"

})

print(grouped\_custom)

## Output

### Salary Bonus

Department mean custom\_salary\_range sum

|  |  |  |
| --- | --- | --- |
| HR | 57333.33 12000 | 18000 |
| IT | 86000.00 10000 | 33000 |
| Finance | 72500.00 5000 | 16500 |

3. Broadcasting Aggregation Results python Copy code df["Total Salary by Dept"] = df.groupby("Department")["Salary"].transform("sum") df["Max Bonus by Dept"] = df.groupby("Department")["Bonus"].transform("max") print(df)

## Output

### Department Employee Salary Bonus Years Total Salary by Dept Max Bonus by Dept

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HR | Alice | 50000 5000 2 | 172000 | 7000 |
| HR | Bob | 60000 7000 3 | 172000 | 7000 |
| IT | Charlie | 80000 10000 5 | 258000 | 12000 |
| IT | David | 90000 12000 6 | 258000 | 12000 |
| Finance | Eve | 70000 8000 4 | 145000 | 8500 |
| Finance | Frank | 75000 8500 4 | 145000 | 8500 |
| HR | Grace | 62000 6000 3 | 172000 | 7000 |
| IT | Hank | 88000 11000 7 | 258000 | 12000 |

# Discussion

Advanced groupby operations are crucial for deriving insights from grouped data. These techniques include:

* Applying multiple aggregation functions.
* Using custom functions to extract specific insights.
* Broadcasting results back to the original dataset for further analysis.