

# EDA-Groupby Aggregation

---

## GroupBy Aggregation

GroupBy aggregation is a powerful tool in data analysis that allows you to perform various operations on grouped data. It's a crucial step in the Exploratory Data Analysis (EDA) process.

### What does it do?

GroupBy aggregation groups data by one or more columns, performs an aggregation operation (e.g., sum, mean, count), and returns the result for each group. This helps you to:

- 1. Identify patterns:** Grouping similar data points together reveals trends, correlations, and relationships within your dataset.
- 2. Aggregate metrics:** Calculate summary statistics like means, sums, counts, or standard deviations for each group.

### Types of aggregations

Common aggregation operations include:

- `mean()`: calculate the mean value
- `sum()`: sum up values
- `count()`: count the number of rows in each group
- `std()`: calculate the standard deviation

### Example: GroupBy Aggregation with Pandas (Python)

Suppose we have a dataset `orders` with columns `order_id`, `customer_id`, `product_name`, and `order_total`.

```

import pandas as pd

# sample data
data = {
    'order_id': [1, 2, 3, 4, 5],
    'customer_id': [101, 102, 103, 104, 105],
    'product_name': ['Product A', 'Product B', 'Product C', 'Product A', 'Product B']
    'order_total': [100.0, 200.0, 300.0, 150.0, 250.0]
}

df = pd.DataFrame(data)

# GroupBy aggregation
grouped_df = df.groupby('product_name')['order_total'].mean()

print(grouped_df)

```

Output:

product_name	order_total
Product A	125.0
Product B	225.0
Product C	300.0

In this example, we grouped the data by `product_name` and calculated the mean of `order_total` for each group.

### EDA Use Cases

GroupBy aggregation is commonly used in EDA to:

- Analyze customer behavior:** Group orders by customer ID, product name, or order date to understand their purchasing habits.
- Identify top-selling products:** Group sales data by product category to determine which products are generating the most revenue.
- Investigate geographic trends:** Group data by region or country to explore patterns in demographics, sales, or other metrics.

By applying GroupBy aggregation, you can gain valuable insights into your data and inform business decisions.