Data Analyst Training Programme

Sorting & Grouping Data

Introduction to Data Organisation

Why Organisation Matters

- Raw data tells no story
- Sorted data reveals patterns
- Grouped data creates insights

Two Key Concepts

- **ORDER BY**: arranges individual rows
- **GROUP BY**: combines rows into summaries

Business Applications

- Top performers analysis
- Sales summaries by region
- Trend identification over time

Mastering ORDER BY

Basic Sorting Syntax

- ORDER BY column_name [ASC|DESC]

Sorting Options

- **ASC (ascending)**: A-Z, 1-10, oldest-newest
- **DESC (descending)**: Z-A, 10-1, newest-oldest
- Default is ascending

Multiple Column Sorting

- Primary sort, then secondary sort
- Each column can have different direction
- Order matters: first column dominates

Sorting Different Data Types

- **Text**: alphabetical order
- **Numbers**: numerical order
- **Dates**: chronological order

Introduction to GROUP BY

GROUP BY Purpose

- Combines rows with identical values
- Creates summary statistics
- Foundation for analytical reporting

Basic Grouping Concept

- Similar records become one row
- Individual details disappear
- Aggregate functions provide summaries

What You Can SELECT

- Grouped columns only
- Aggregate functions (**COUNT**, **SUM**, **AVG**, etc.)
- Cannot mix individual rows with groups

Common Business Uses

- Sales by region
- Orders per customer
- Products by category

Aggregate Functions

Essential Aggregate Functions

- **COUNT()**: number of records
- **SUM()**: total of numeric values
- **AVG()**: average of numeric values
- **MIN()**: smallest values
- **MAX()**: largest values

COUNT Variations

- **COUNT(*)**: counts all rows including NULL
- **COUNT(column)**: counts non-NULL values only
- **COUNT(DISTINCT column)**: counts unique values

Handling NULL Values

- Most aggregates ignore NULL
- Can affect results unexpectedly
- Always consider data quality

Advanced Grouping and Filtering

HAVING Clause

- Filters groups after aggregation
- WHERE filters before grouping
- Uses aggregates functions in conditions

WHERE vs HAVING

- **WHERE**: filters individual rows
- **HAVING**: filters grouped results
- Often used together

Combining ORDER BY with GROUP BY

- Sort grouped results
- Can sort by aggregate values
- Reveals top and bottom performers

Complex Grouping Patterns

- Multiple grouping patterns
- Nested analysis levels

Assignment

Complete these sorting and grouping challenges using the Northwind database:

Basic Sorting

- List all products ordered by price (highest first)
- Show customers alphabetically by country, then by name
- Display recent orders first (most recent OrderDate first)

Simple Grouping

- Count how many products exist in each category
- Find the average price of products by category
- Count customers by country

Advanced Analysis

- Show categories with more than 10 products
- Find countries with exactly 5 customers
- List the most expensive product in each category

Complex Challenges

- Find the top 3 countries by customer count
- Show monthly order counts for 1996 (grouped by year and month)
- Identify categories where average product price exceeds £30
- Calculate total sales revenue by category (requires joining with OrderDetails)

Until Next Week Sunday...

See you next week on Sunday, **[student name]**.

Your foundation is now complete. You can find any data by filtering, organize it by sorting, and summarize it by grouping. Next week, we will start looking into **Data Visualization**.

Thank you, [student name].

Any Questions?