

KNIME Exercise: Database Data Manipulation (SQLite)

Problem Statement:

You are given a SQLite database file named `WebActivity.sqlite` inside the data folder. This database contains a table called `web_activity` that stores customers and their web activity category.

Your task is to build a KNIME workflow that will identify the most popular web activity category (the category with the maximum number of customers) and then extract all records from the original table for only that category.

Goal

Create a KNIME workflow that:

1. Connects to the database
 2. Reads the table in database mode (DB nodes)
 3. Finds the most popular web activity category
 4. Filters the original table for that category
 5. Loads the final filtered result into a KNIME data table
-

Input

- Database file: `WebActivity.sqlite`
- Table name: `web_activity`

The table contains at least the following columns:

- `CustomerKey`
 - `WebActivity`
 - (may include additional columns like count)
-

Tasks to Perform

Build the workflow to complete the following steps:

1) Connect to SQLite

Connect KNIME to the database `WebActivity.sqlite` located in the data folder.

2) Select Table

Select the table `web_activity` from the database.

3) Customer Count Per Category

Calculate the number of customers in each `WebActivity` category.

Expected result:

A table showing each `WebActivity` category and how many customers belong to it.

Example structure: |WebActivity|Customer_Count|:--:|--:|0|...|1|...|2|...|

4) Maximum Customers Across Categories

From the aggregated table, compute the maximum value of `Customer_Count`.

Expected result:

one row:

`Max_Customer_Count`

...

5) Identify Most Popular Category

Filter the aggregated data so that only the category/categories having:

- `Customer_Count = Max_Customer_Count` remain.

Expected result:

one row (or more if tie).

6) Filter Original Data

Using the most popular category value, `filter` the original `web_activity` table so you only keep rows belonging to that category.

>Expected result: filtered original records (not aggregated).

7) Read into KNIME Table

Finally, read the filtered DB output into KNIME using a `DB Reader` so it becomes a standard KNIME data table.

Constraints / Rules

- You must use DB nodes only (no CSV reader / no normal groupby).
- You should perform aggregation and filtering while data is still inside the database.
- Only in the last step you should bring data to KNIME using DB Reader.

Final Expected Output

At the end, you should have a KNIME output table that contains:

- Only customers belonging to the most popular WebActivity category
- All original columns (CustomerKey, WebActivity`, etc.)

Bonus Challenge (Optional)

1. What if two categories have the same maximum count? Modify your workflow to return both.
2. Add a node to display the most popular WebActivity category clearly.
3. Rename WebActivity values (0,1,2,3...) into readable labels using Rule Engine after DB Reader.

□ End of Project

Exercise Database Data Manipulation

Concise Problem Statement

- 1) Connect to the WebActivity.sqlite database available in the "data" folder
- 2) Select the web_activity table
- 3) Calculate the number of customers in each web activity category
- 4) Calculate the maximum number of customers out of all web activity categories
- 5) Filter the aggregated table to the most popular web activity category
- 6) Filter the original table according to the most popular web activity category
- 7) Read the filtered data into a KNIME data table

