Legend Design:

Note: It's a big data, 541910 columns/transactions); **need to make it smaller!**

The dataset includes 8 variables:

- InvoiceNo: transaction identifier;
- StockCode: product identifier;
- Description: Product name;
- Quantity: Number of each product purchased per transaction;
- InvoiceDate: Date and time of transaction;
- UnitPrice: Product price per unit in pounds sterling (£);
- CustomerID: customer identifier;
- Country: Customer's country of residence;

Important Libraries:

Preprocess with Polars

- Remove null values (step 1).
- Filter Quantity > 0 (step 1).
- Convert Invoice No

to grouped transactions (step 1).

Step 1 Dataset (D) Summary:

- Import CSV file with Polars for optimized performance.
- Filter, summarize, group, and list all from the dataset.

Step 2 Frequent Itemsets (FK) Summary:

- Count each individual item (from grouped transactions).
- Apply Minimum Support Filters.
- Frequent itemsets filtering.

Step 3 Create Candidates (CK) Summary:

- Iteratively generate and prune itemsets until no more frequent itemsets can be found. Generate Ck from Lk+1.
- Output final frequent itemsets (LK).

Step 4 Generate Rules (FI) Summary:

- Generate rules from frequent itemsets (X → Y) with different parameters and show the results at final.

Step 5 Filter and Rank Rules (RG) Summary:

- Generate the lift-defined ratio of the conditional x -> Y and generate the certainty factor

Step 6 Visualize and Evaluate (RF) Summary:

- Final results tests, comparisons, knowledge and feedback about all steps before.