



LUMINA MART Co.

— Egyptian Dairy Market —
Your Premium Dairy Partner



Filters



Navigation



Executive Overview



Sales Performance



Product Analysis



Promotions Analysis



Returns & Quality



Survey Insights



26M

Total Profit



22%

Profit Margin



50.15%

Net Sales YoY %



36%

Return Rate %



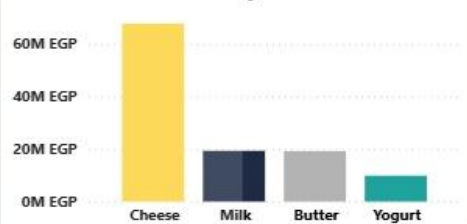
108M

Net Sales After Returns

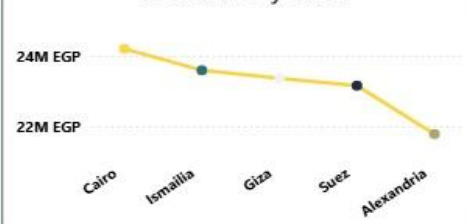
Customer Revenue Ranking



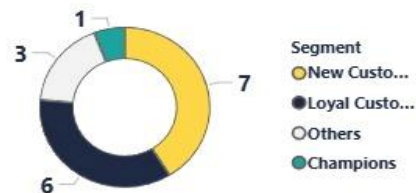
Net Sales by Product



Net Sales by Area



Segment Distribution



0.58

Aver...

This metric reflects overall revenue performance based on sales, profit, and return behavior.

0.00 0.50 0.75 1.00



0.00 – 0.50 ● Poor
 0.50 – 0.75 ● Average
 0.75 – 1.00 ● Good

2 Datasets.

111 + 999 rows.

Zero errors.

How I cleaned, transformed, and modeled two real-world datasets in Power Query & Power BI

Survey Dataset

Sales Dataset

Power Query

Feature Engineering

Power BI Model



Two datasets, one pipeline, one goal

DATASET 1 — SURVEY

- 111 respondents — Arabic survey data
- 21 columns raw → 32 cleaned
- 20+ Replace Value steps
- 4 binary dummy variables
- 7 engineered feature columns
- 2 queries merged (Left Outer Join)
- Final model: 33 columns, 0 nulls

DATASET 2 — SALES

- 999+ transactions — commercial sales
- 52 columns, pre-validated clean data
- Focus: business logic, not remediation
- Order Size & Profitability classification
- Offer Period & Credit Check logic
- Return Rate validation
- Full 52-col model loaded to Power BI

The raw data was a mess.

BEFORE

✗ Arabic text inconsistencies

القاهرة ≠ القاهرة

الجيزة ≠ الجيزة

✗ Multi-value columns

'البن، زبادي، جبن، زبدة'
(impossible to analyze)

✗ Columns with no meaning

Timestamp, Column 3 (100% empty)

✗ No analytical codes

Text only – unusable in DAX



AFTER

- ✓ Canonical Arabic forms
- 20+ Replace Value steps
- 13 governorates deduplicated

- ✓ 4 binary dummy variables
Milk_Flag | Yogurt_Flag
Cheese_Flag | Butter_Flag

- ✓ Removed + reordered columns
- 19 clean, named columns

- ✓ 7 engineered feature columns
- Ordinal codes + categories



5 phases. 35+ documented steps.

01

Data Import & Profiling

Load Excel → Promote Headers → Set Types → Column Profile → 100% Valid confirmed

02

Text Cleaning & Standardization

20+ Replace Values on Arabic text · Governorate deduplication · Trim + Remove null row

03

Column Splitting & Binary Encoding

Multi-value column split by delimiter → 4 positional cols → 4 binary dummy flags + count

04

Feature Engineering

7 new columns: ordinal codes, impact scores, categories, Sort By Column assignments

05

Validation & Model Load

All 19 columns: 100% Valid · 0 errors · 0 nulls · 35+ named steps in Query Settings

Not just cleaning. Analytical thinking.



Ordinal Encoding — not obvious

Converting Arabic text responses to ordered numeric codes (1–4) enables Power BI to sort visuals correctly. Without this, charts show alphabetical disorder.

Fair_Price_Code · Promo_Impact_Score ·
Seasonal_Consumption_Code



Sort By Column Assignment

Assigning a numeric code as the Sort By Column for a text field is a Power BI modeling decision, not a cleaning step — it prevents visual sorting bugs in every chart downstream.










Binary Flags — the right way

Multi-value column 'البن، جين، زيادي' split → 4 dummy variables. Each flag = 1 if product present, 0 otherwise. This is the standard approach for ML-ready categorical encoding.



Clean data ≠ no work to do.

 *The sales dataset arrived pre-validated with 0 structural errors. The professional move was recognizing this — and shifting focus entirely to business rule implementation and analytical enrichment, not remediation.*

 Order Size:	Small / Medium / Large — classified from order data
 Profitability Level:	Low / Medium / High — derived from margin calculations
 Stock Check Flag:	'Sufficient' vs 'Reorder Required' — inventory threshold logic
 Offer Period Check:	3 outcomes: Within Period / Review / No Offer — conditional column
 Credit Validation:	'Valid Cash' / 'Valid Credit' / 'Review' — payment integrity check
 Return Rate:	Calculated_Return_Rate validated against source — exact match confirmed

Two queries. One unified model.

Sheet 1

Survey Responses

Response_ID ← Join Key

Governorate

Fair_Price + Code

Promo_Impact + Score

Seasonal flags (4)

Economic_Impact_Category

7 engineered columns

**LEFT
OUTER
JOIN**



Join key: ID

Sheet 1 (2)

Consumer Behaviour

ID ← Join Key

Age + Family_Size

Purchase_Frequency

Milk/Yogurt/Cheese flags

Butter/Cream flags

Preferred_Brand

Brand_Loyalty_Type

Result: 33 columns × 111 rows — fully merged, 0 errors, loaded to Power BI



What the pipeline produced.

111

Survey Respondents

999+

Sales Transactions

20+

Replace Value Steps

7

Engineered Features

35+

Applied Steps Logged

5

Binary Product Flags

33

Merged Survey Cols

0

Errors / Nulls

Both datasets fully loaded into Power BI with correct types, English snake_case naming, and Sort By Column assignments.

The full stack behind this project.



Power Query (M)

Primary transformation engine — all applied steps, custom columns, and merges



Power BI Desktop

Data model, Sort By Column, field panel configuration and DAX-ready structure



Arabic Text Processing

Manual normalization of Arabic categorical inconsistencies across 13 governorates



Ordinal Encoding

Numeric code columns for unordered Arabic text → enables sorting & DAX measures



Binary Flag Engineering

Dummy variable creation from multi-value delimited columns via Split + Conditional logic



Query Merging

Left Outer Join on Response ID — preserves all III records with full column expansion

This is just the data foundation.

The LUMINA MART Dairy Intelligence Dashboard — built on this exact pipeline — is live in Power BI.

- > Full 68-slide Data Prep report available on request
- > Dashboard covers 6 analytical pages
- > Open to feedback, collaboration, and data conversations