ABC Foodmart Database Implementation

Group 7: Minkyung Kim, Boni Vasius Rosen, Theodore Julien Zaphiris, Peter Zhang

APANS5310

Aug 13, 2025

Problem Statement

ABC Foodmart, a neighborhood grocery chain with existing stores in Queens, is preparing to expand into Brooklyn with three new locations. The company currently relies on paper records and spreadsheets to manage critical business functions such as inventory, vendor relationships, staffing, and sales. This manual, disconnected approach has led to delays, stockouts, and operational errors, limiting the company's ability to scale effectively.

These challenges are not unique to ABC Foodmart, industry research shows that 65% of grocery retailers lack real-time supply chain visibility (PYMNTS, 2025), resulting in inefficiencies that directly impact profitability and customer satisfaction. Without a modern, integrated system, these inefficiencies will persist and potentially worsen as operations grow in size and complexity. The lack of centralized, real-time access to data prevents managers from making timely, informed decisions on staffing, inventory replenishment, and vendor management. These gaps reduce profitability, risk customer dissatisfaction, and limit the company's competitiveness in an increasingly data-driven grocery market.

Proposal

To address ABC Foodmart's operational challenges and prepare for its planned expansion, our team will design and implement a centralized relational database system that integrates all core business data across store locations. The database will be structured in 18 fully normalized tables (3NF) to ensure data integrity and eliminate redundancy.

We will develop automated ETL processes to load, clean, and update data from multiple operational sources, ensuring that information on staffing, inventory, vendors, deliveries, sales, and accounting is always accurate and up to date.

The solution will meet ABC Foodmart's specific business requirements in the following areas:

Core Operations

- Maintain store-level data on staffing, inventory, vendors, deliveries, sales, expenses, and accounting across all locations.
- Provide historical daily and monthly net revenue after discounts and returns for profitability tracking.
- Identify top categories per month by contribution percentage to optimize inventory allocation and marketing focus.

Staffing & Human Resources

- Maintain complete employee records (name, contact, role, assigned store, and department).
- Track scheduled labor hours and headcount per store per day to match staffing with sales demand.

Product & Inventory Management

- Store complete product details (SKU, name, brand, category, shelf location) with linked historical pricing, including promotional prices.
- Generate inventory watchlists for understocked and overstocked items to guide restocking and markdowns.
- Identify frequently returned products and top return reasons by refund value to address quality or fulfillment issues.

Vendors & Deliveries

- Record vendor details, including tier and supplied products.
- Evaluate vendor on-time delivery rates to improve supply chain reliability and prioritize top-performing suppliers.

Sales & Financials

- Capture all sales transactions with product, store, date/time, quantity, payment type, and promotional linkage.
- Compare units/day for promotional vs. non-promotional sales to assess campaign effectiveness.
- Identify top SKUs by incremental units sold during promotions to guide targeted marketing efforts.
- Track refunds during promotional periods by category to detect and resolve promo-related issues.

Manager Reporting & Insights

- Provide strategic management dashboards with business overview, such as daily net revenue trends, monthly top categories, and average basket size. promotion performance, inventory status, vendor on-time delivery rates, and top return reasons.
- Allow analysts to drill into detailed transaction data, while executives and store managers view high-level, actionable summaries.

To deliver on these requirements, we will implement eleven complex SQL queries that produce actionable metrics for decision-making, along with an interactive Metabase dashboard that enables non-technical users to view and explore results in real time.

By adopting this solution, ABC Foodmart will gain centralized operational visibility, improve efficiency in inventory and supply chain management, enhance vendor oversight, and increase sales performance. The system will also provide a scalable foundation for future analytics initiatives such as forecasting, targeted promotions.

Team Structure and Timeline

Team Members:

- Member 1: Minkyung Kim
- Member 2: Boni Vasius Rosen
- Member 3: Theodore Julien Zaphiris
- Member 4: Peter Zhang

Breakdown of Tasks and Responsibilities

Task	Responsible Member(s)	Deadline
Finalize business requirements list	Member 1+ Member 2	July 21
Database schema design (ER diagram + 3NF tables)	Member 1 + Member 2	July 27
Plan / Generate dataset & Python ETL scripts	Member 1 + Member 4	August 1
Develop and test complex SQL queries	Member 1 + Member 2	August 5
Build interactive dashboard (Metabase)	Member 3 + Member 2	August 8
Final report and presentation slide	Member 1 + Member 2	August 12

Timeline of Activities

Phase	Key Deliverables	Target Date
Week 1 (July 18–22)	Finalize business requirements + draft ER diagram	July 21
Week 2 (July 23–29)	Complete relational schema design, begin data sourcing and loading	July 27–29
Week 3 (July 30–Aug 5)	Complete data loading, develop complex SQL queries	August 5
Week 4 (Aug 6–11)	Build and test dashboards, write and finalize report & presentation	August 12
Final Presentation	Submit and present the project	August 13

Most Challenging Tasks

1. **Data Generation with AI & ETL** – Creating realistic datasets to fulfill business requirements and developing Python scripts to load, clean, and integrate them into PostgreSQL while resolving data quality issues.
2. **Relational Schema Design** – Building a fully normalized (3NF) schema that met all requirements and supported complex analytical queries while maintaining good performance for reporting.

Database Schema

Full CREATE TABLE scripts are provided in the appendix with a link to the project's GitHub repository and Google Drive.

Design Logic

The design follows a fully normalized Third Normal Form (3NF) structure to:

- Eliminate redundancy and ensure data integrity.
- Maintain clear relationships between entities such as stores, products, promotions, employees, vendors, sales, returns, and deliveries.
- Support complex analytical queries without requiring excessive data cleaning or transformation.

Overall Schema Design

The ABC Foodmart database contains 18 tables grouped into key business domains:

1) Operations & HR

- **Store table** – Stores details such as location, operating hours, and assigned manager.
  Justification: Enables location-specific reporting for sales, expenses, staffing, and operational performance.
- **Department table** – Represents departments within each store (ex. Bakery, Dairy).
  Justification: Allows analysis of performance by department and helps allocate staffing and promotions effectively.
- **Employee table** – Tracks employee records including contact, role, and department assignment. Staff details are linked to the store and department.
  Justification: Supports HR management and employee performance tracking.

- **ShiftSchedule table** – Employee shift records.
  Justification: Enables labor cost analysis, ensures adequate staffing levels.

2) Products, Categories & Pricing

- **Category table** – Categorizes products by type (ex. frozen, fresh, beverage).
  Justification: Facilitates product grouping for reporting.
- **Product table** – Stores product details including SKU, name, brand, and shelf location.
  Justification: Serves as the central reference for linking sales, pricing, promotions, inventory, and vendor relationships.
  ProductPricing table – Maintains a history of regular and promotional pricing by date.
  Justification: Preserves historical pricing for trend analysis, margin calculation, and promotion effectiveness measurement.

3) Inventory & Supply Chain

- **Inventory table** – Store-specific stock levels and restock thresholds.
  Justification: Supports restocking automation, low-stock alerts, and helps prevent stockouts or overstock situations.
- **Vendor table** – Supplier details and tier.
  Justification: Enables vendor performance evaluation and tier-based prioritization for procurement.
- **VendorProduct table** – M:N link between vendors and products.
  Justification: Supports scenarios where multiple vendors can supply the same product and a vendor can supply multiple products.
- **Delivery table** – Delivery events from vendors to stores.
  Justification: Allows tracking of delivery schedules, delays, and supply chain efficiency.
- **DeliveryItem table** – Products and quantities within a delivery.
  Justification: Supports partial deliveries and verification of delivered quantities.

4) Sales, Promotions, Returns, and Expenses

- **Promotion table** – Campaign periods, discounts, linked products.
  Justification: Tracks marketing efforts and allows analysis of sales lift during promotions.
- **Sale table** – Transaction header with store, datetime, and payment type.
  Justification: Forms the basis for daily, weekly, and monthly sales summaries and trend analysis.
- **SaleItem table** – Items within a sale, including promo details.
  Justification: Enables basket analysis, promotional impact tracking, and detailed margin calculations.

- **ProductReturn table** – Returned items linked to sales and reasons.
  Justification: Facilitates return rate analysis, quality control, and refund processing.
- **ReturnReason table** – Codes and descriptions for return category.
  Justification: Ensures consistent categorization of returns for reporting and quality improvement.
- **Expense table** – Store-level operating costs by category.
  Justification: Enables profitability analysis and budget tracking per store.

Triggers and Functions Overview

There are four trigger functions designed to keep inventory levels and status in sync automatically for our client and performance efficiency.

- **update_restock_status()**
  Trigger: trg_update_restock_status on Inventory (BEFORE INSERT/UPDATE)
  Logic: Sets restock_status to 'Restock Needed' if quantity_on_hand <= reorder_threshold, otherwise 'In Stock'.
  Justification: Ensures stock status is always accurate without manual intervention.

- **deduct_inventory_after_sale()**
  Trigger: trg_deduct_inventory on SaleItem (AFTER INSERT)
  Logic: Finds the matching Inventory record based on Sale.store_id and SKU and subtracts quantity_sold.
  Justification: Maintains real-time inventory accuracy after sales transactions.

- **add_inventory_on_return()**
  Trigger: trg_add_inventory_on_return on ProductReturn (AFTER INSERT)
  Logic: Increases the Inventory.quantity_on_hand for the returned item based on store and SKU.
  Justification: Prevents inventory discrepancies when products are returned.

- **add_inventory_on_delivery()**
  Trigger: trg_add_inventory_on_delivery on DeliveryItem (AFTER INSERT)
  Logic:
    - If the (store_id, sku) already exists in Inventory, increment quantity_on_hand by delivered quantity.
    - If it does not exist, insert a new record with quantity_on_hand set to delivered quantity, a default reorder_threshold of 10, and restock_status set to 'In Stock'.

  Justification: Ensures deliveries are immediately reflected in inventory for accurate stock monitoring.

Key relationships & cardinarlity:

- Store (1) — (M) Employee/ Inventory/ Sale/ Expense/ Delivery
- Store (1) — (1) Employee (Manager)
- Employee (1) — (M) ShiftSchedule
- Department (1) — (M) Employee
- Category (1) — (M) Product
- Product (1) — (M) ProductPricing / Promotion / Inventory / SaleItem ProductReturn / DeliveryItem
- Vendor (1) — (M) Delivery
- Vendor (M) — (M) Product via VendorProduct
- Delivery (1) — (M) DeliveryItem
- Sale (1) — (M) SaleItem / ProductReturn
- ReturnReason (1) — (M) ProductReturn
- SaleItem (M) — (0..1) Promotion

Entity–Relationship Diagram (ERD)

The ER diagram visually represents these relationships, showing primary keys, foreign keys, and cardinalities.

- ER Diagram PDF: Group7_ERD.pdf
- Lucidchart Link: https://lucid.app/lucidchart/8f9f860d-7f98-4a52-b12d-03107da7da06/edit?viewport_loc=1171%2C-234%2C5256%2C2516%2C0_0&invitationId=inv_1ef1578d-791f-43c1-925b-9d110223dd8f

DATASET

To populate the ABC Foodmart relational database with realistic, analysis-ready data, our team generated four primary master tables using a hybrid data creation process that integrated multiple Large Language Models (LLMs), specifically OpenAI GPT-4o, GPT-4o-mini, and GPT-3o. Each master table is designed to align with the finalized ERD and meet the referential integrity constraints defined in our schema. These datasets serve as the foundation for the ETL process, analytical queries, and dashboard visualizations. Raw master datasets and a record of data generation with LLMs are shared in the Google Drive link and GitHub repository in the appendix.

1. Sales_Master.csv

- **Purpose:** Captures detailed sales transactions at the item level.
- **Design Notes:**
    - Dates span a realistic operational range to support time-series analysis.
    - SKUs match the Product and delivery table to maintain referential integrity.
    - Promotional pricing is linked with the promotion ID.
    - Includes mixed payment methods for realistic transaction diversity.

2. Expense_Master.csv

- **Purpose:** Records store-level operational expenses by category.
- **Design Notes:**
    - Categories cover rent, utilities, salaries, spoilage, and miscellaneous costs.
    - Expense amounts are proportionate to store size and sales volume for profitability analysis.
    - Ensures no negative values and matches store IDs defined in the Store table.

3. Delivery_Master.csv

- **Purpose:** Tracks deliveries from vendors to stores, including product quantities.
- **Design Notes:**
    - Includes varied delivery statuses such as *Completed*, *Delayed*.
    - SKUs and vendor IDs align with Vendor and Product tables.
    - Delivery quantities are designed to affect inventory thresholds for trigger testing.

4. Shift_Master.csv

- **Purpose:** Manages employee scheduling across departments and stores.
- **Design Notes:**
    - Covers all active employees, ensuring each store and department has complete daily coverage.
    - Department assignments align with the Department table for staffing analysis.

ETL Process

We built a Python-based ETL pipeline using pandas and psycopg2 to load four master datasets: Sales_Master.csv, Expense_Master.csv, Delivery_Master.csv, and Shift_Master.csv into the normalized ABC Foodmart PostgreSQL schema. Full ETL code is provided in the Appendix and available via the project's GitHub repository and Google Drive folder for reference and reproducibility.

Steps:

1. Create tables & triggers (18 tables, 4 inventory-related triggers)
2. Extract required columns for each target table
3. Clean & type data (remove duplicates, cast types, handle nulls).

4. Load in FK-safe order:
    - **Core:** Store → Department → Employee → ShiftSchedule
    - **Products**: Category → Product → ProductPricing
    - **Supply**: Inventory → Vendor → VendorProduct → Delivery → DeliveryItem
    - **Sales:** Promotion → Sale → SaleItem
    - **Finance:** Expense → ReturnReason → ProductReturn
5. Loop-based Inserts:
    - For each master dataset, rows were iterated using Python for loops, and insert statements were executed for each record.
6. Triggers automatically update inventory after sales, returns, or deliveries and adjust restock_status.
7. Upserts (ON CONFLICT DO NOTHING) allow safe re-runs without duplicate inserts.

Outcome:

Clean, validated data with enforced PK/FK constraints, automated stock updates, and consistent inventory levels ready for complex querying by analysts or dashboard implementation.

Analytical Procedures

To provide valuable insight to the client, the analytical procedures below have been developed based on business requirements. Full complex query code is provided in the Appendix and is available via the project's GitHub repository and Google Drive folder.

1) **Daily Net Revenue (after discounts & returns)**
   Client question: "What did we actually make each day (by store), net of promos and refunds?"
   Valuable insight: Track true daily performance (not just gross sales), separate promo effects from refunds, and spot days/stores dragging net revenue so you can adjust pricing, promos, and staffing.

2) **Top Categories per Month (ranked, contribution %)**
   Client question: "Which categories actually carry the month (per store) after returns?"
   Valuable insight: Observe the few categories contributing most of the net revenue (Pareto), quantify contribution %, and prioritize shelf space, promo spend, and inventory depth where it matters.

3) **Average Basket Size (items per order)**
   Client question: "Are customers buying more items per trip (per store × day)?"

Valuable insight: track attachment trends, measure impacts of cross-sell and merchandising, and correlate basket size lifts with promotions or layout changes.

4) **Inventory Watchlist (Understock / Overstock)**
Client question: "Where are we below threshold or tying up cash in overstock?"
Valuable insight: provide a daily watchlist to prevent stockouts on fast movers and reduce capital locked in slow movers; guides reorders and markdowns.

5) **Scheduled Labor Hours (headcount & hours)**
Client question: "How many hours/people did we schedule per store × day?"
Valuable insight: adjust staff to demand. Spot under/over staffing versus revenue or footfall for better service and labor cost control

6) **Units/Day: Promo vs. Non-Promo (uplift%)**
Client question: "Which SKUs move faster when discounted (and by how much)?"
Valuable insight: quantify promo effectiveness per SKU; stop discounting items that don't lift, double-down on winners.

7) **Top 5 SKUs by Incremental Units (promo vs. not)**
Client question: "Which SKUs gained the most additional unit volume during promos?"
Valuable insight: Focus displays, inventory, and ad spend on SKUs where promos add volume (not just margin erosion).

8) **Refunds During Promo Window (by category)**
Client question: "Are our promos spiking returns in any category?"
Valuable insight: Catch promo/quality issues early (e.g., size, packaging, misleading offers) and refine promo mechanics or vendor QA

9) **On-Time Delivery Rate**
Client question: "Which vendors are reliably on time?"
Valuable insight: Steer purchase orders to vendors with high on-time rates; escalate or renegotiate with chronic delayers to protect availability.

10) **Top Return Reasons by $ value**
Client question: "What's costing us the most in returns (defective/damaged/wrong item/etc)?"
Valuable insight: Attack the biggest dollar drains first—tighten vendor SLAs, packaging, or fulfillment accuracy where it hurts most.

11) Market Basket: Products Bought Together

Client question: “Which pairs are frequently purchased together so we can co-merchandise/promote?”

Valuable insight: Build end-caps and bundles with high-lift pairs to grow basket size and

Database Workflow

ABC Foodmart’s database system is designed for three user groups: analysts, executives, and non-technical staff.

Analysts will connect directly to the PostgreSQL database via PgAdmin, Metabase’s SQL editor, or Python notebooks. They will have access to 11 parameterized SQL procedures stored in both PgAdmin and Metabase, enabling them to run ad-hoc queries, schedule automated pipelines, and leverage materialized views for faster performance.

Executives will access a Metabase “Performance & Operation” dashboard showing all key metrics such as daily net revenue, category Pareto, basket size, inventory alerts, promo uplift, vendor performance—filterable by date, store, or category. They will also receive a scheduled PDF or email with KPI summaries and exception alerts.

Non-technical users, such as store managers and procurement teams, will use simplified dashboards with pre-filtered views, tooltips, and CSV export for quick operational actions, no SQL coding skill required.

The platform combines PostgreSQL for secure, centralized data, Metabase for interactive dashboards and reporting, and Python for automation, reproducibility, and advanced analysis. This ensures all teams, from data specialists to executives, interact with a single source of truth, enabling consistent, timely, and actionable insights.

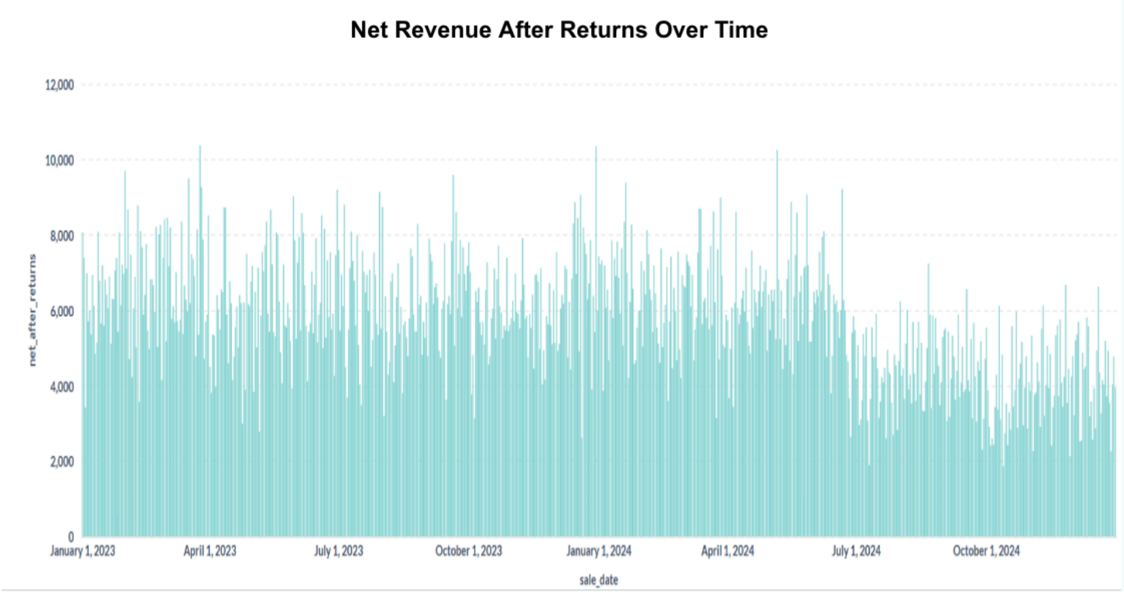
Database Design Redundancies

The database design includes redundancy through automated daily backups, PK/FK constraints, and trigger-based inventory synchronization to prevent data loss or inconsistency. For performance, indexes were created on high-traffic columns and queries optimized for dashboard rendering. Given scalability needs, real-time access, and reduced maintenance overhead, **cloud hosting (e.g., AWS RDS or Azure Database for PostgreSQL) is recommended over on-premises deployment.**

Dashboard Walkthrough

We developed an integrated Metabase dashboard suite to provide ABC Foodmart with real-time operational insights across sales, promotions, inventory, and returns.

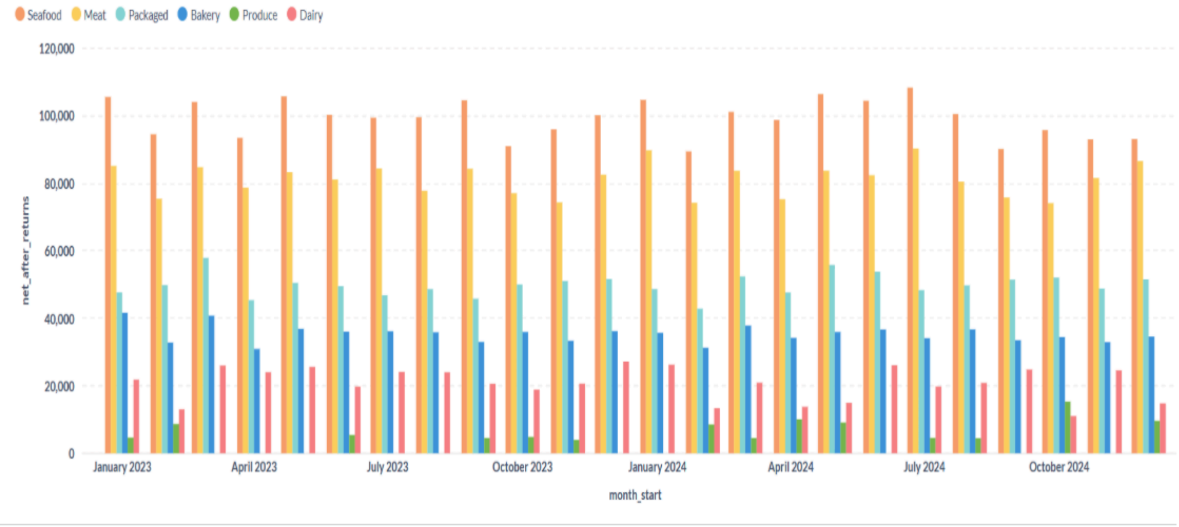
- **Net Revenue After Returns Over Time** – Tracks actual daily revenue per store after accounting for promotional discounts and refunds, enabling managers to monitor profitability trends.



Net Revenue After Returns Over Time

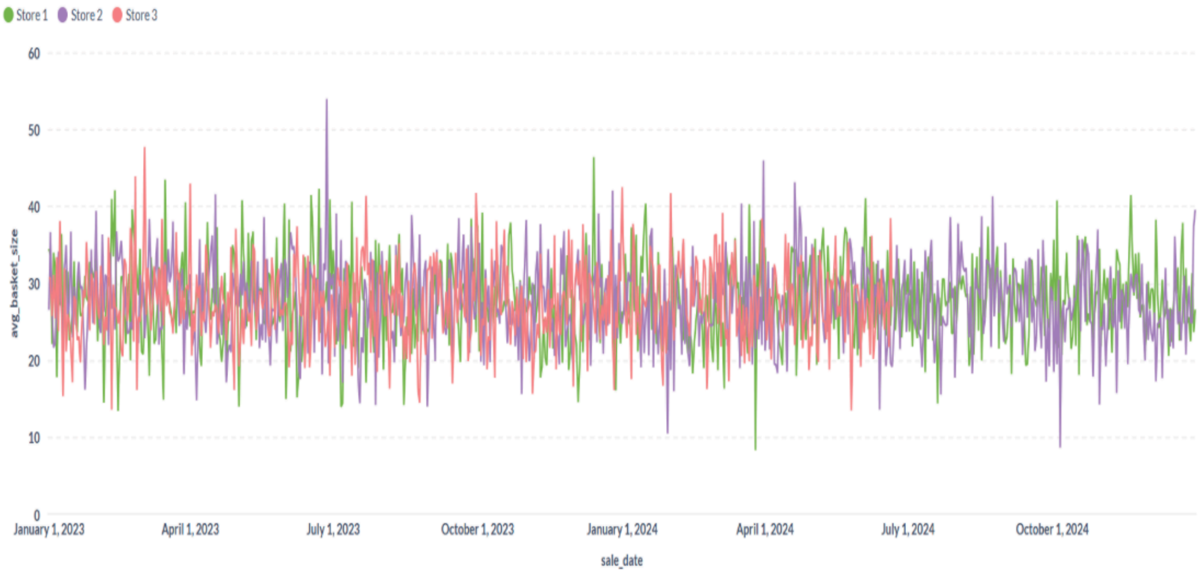
- **Top 5 Categories by Net Revenue After Returns (Monthly, per Store)** – Highlights the highest-performing product categories, supporting targeted marketing and stock allocation decisions.

Top 5 Categories by Net Revenue After Returns (Monthly, per Store)



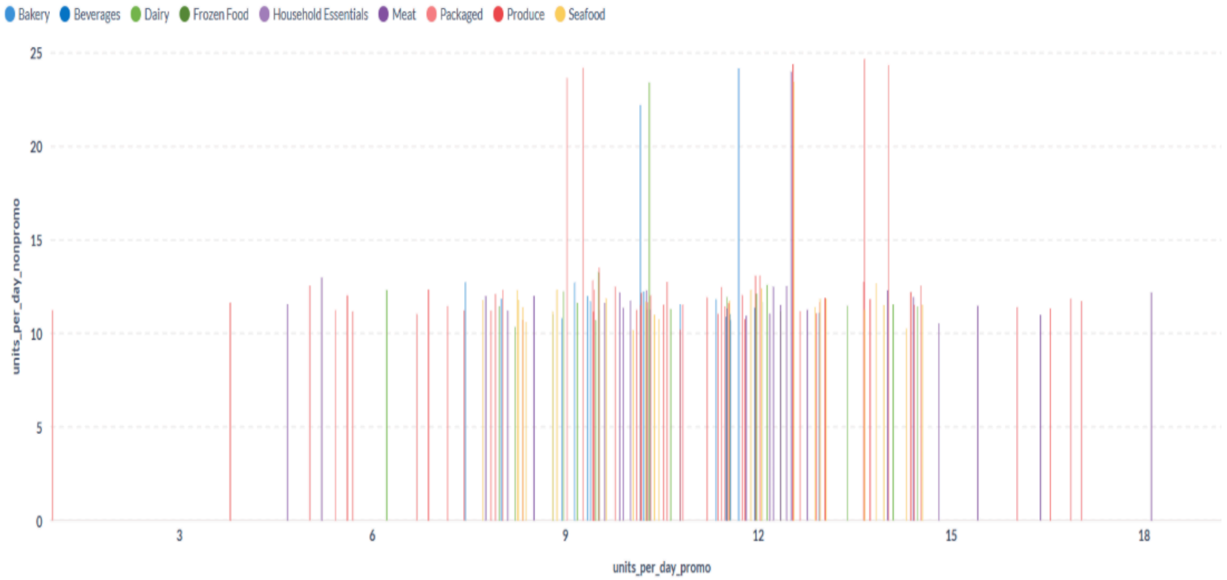
- **Average Basket Size by Store and Day** – Measures average transaction value, helping identify shifts in customer purchasing patterns.

Average Basket Size by Store and Day



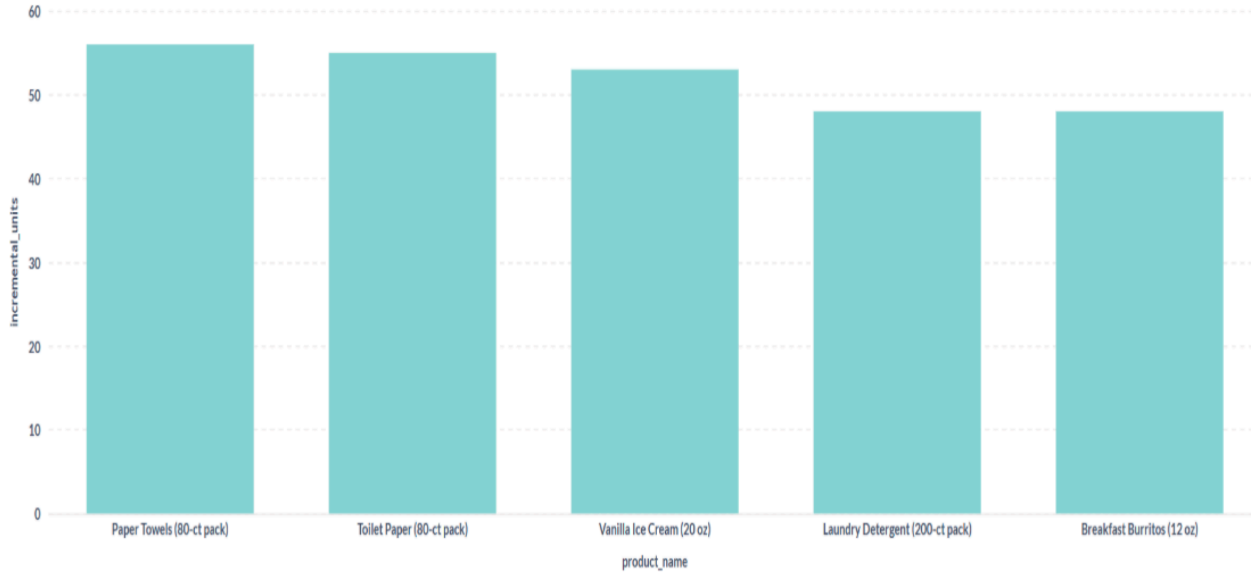
- **Average Daily Units Sold: Promotional vs. Non-promotiona**l – Compares sales velocity of promotional items versus regular-priced items, quantifying promotion effectiveness.

Average Daily Units Sold: Promotional vs. Non-promotional



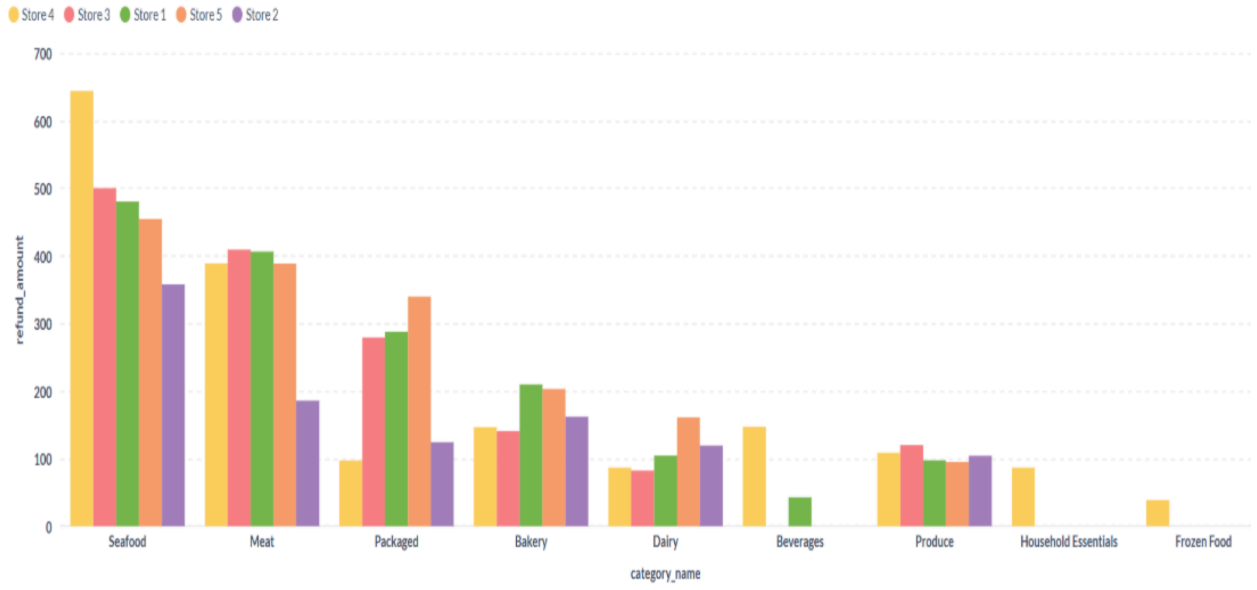
- **Top 5 Products Driving Incremental Sales from Promotions** – Pinpoints SKUs with the largest lift from promotions, guiding future campaign planning.

Top 5 Products Driving Incremental Sales from Promotions



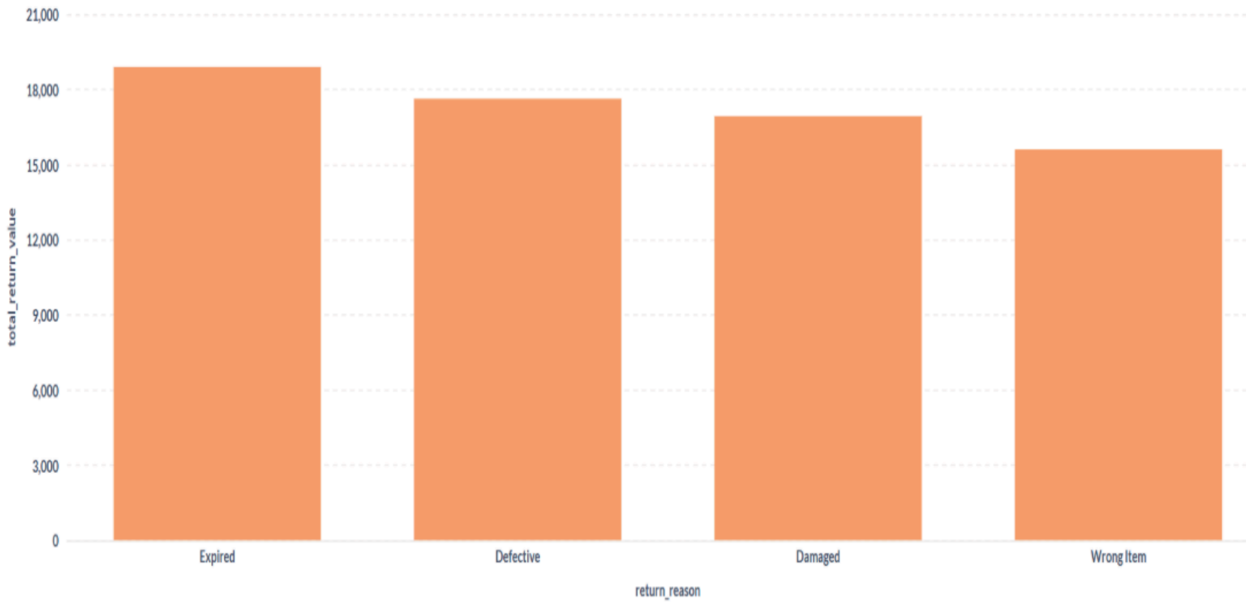
- **Top Categories by Refunds During Promotional Sales** – Identifies categories with high refund rates during promos, enabling quality control and offer adjustments.

Top Categories by Refunds During Promotional Sales



- **Top Return Reasons by Dollar Value** – Summarizes primary drivers of lost revenue due to returns, supporting vendor negotiations and product quality improvements.

Top Return Reasons by Dollar Value



Overall Benefits:

These dashboards transform complex SQL outputs into intuitive visualizations accessible to both analysts and executives. Analysts can drill down into specific transactions for root cause analysis, while executives receive clear, high-level metrics to guide strategic decisions. The real-time nature of the dashboards ensures timely intervention to optimize revenue, manage inventory, and reduce losses.

Conclusion

Our goal was to equip ABC Foodmart with a centralized, scalable, and analytics-ready relational database to replace fragmented manual systems, support expansion, and enable data-driven decision-making. We achieved this through a fully normalized 3NF schema, automated Python-based ETL pipelines, 11 advanced analytical SQL procedures, and interactive Metabase dashboards.

The RDMS ensures data integrity, eliminates redundancy, and provides a single source of truth across operations, inventory, sales, staffing, and vendor management. The ETL process delivers clean, timely data, while the analytical queries and dashboards translate raw information into actionable insights—such as true net revenue, promo effectiveness, inventory alerts, and vendor reliability.

As a result, ABC Foodmart can make faster, more informed decisions, optimize stock levels, improve labor allocation, enhance vendor performance, and focus marketing spend where it delivers the highest ROI. This implementation not only resolves current inefficiencies but also establishes a strong foundation for predictive analytics, scalability, and sustained competitive advantage.

Appendix:

Google Drive Link: SQL_PROJECT_GROUP7
Github Repository Public Link: https://github.com/ginny-1334/APAN5310_SQL_Group7

Note: *Database Schema SQL, ERD, ETL Python, Dashboard Screenshot, AI Procedure for data generation, and 11 Complex Analytical Query files are both uploaded to GDrive and GitHub. However, the generated master datasets are only uploaded to GDrive, as GitHub does not allow files over 30MB.*

Reference:

Pymnts, & Pymnts. (2025, March 26). 65% of grocery retailers lack Real-Time supply chain data. PYMNTS.com. https://www.pymnts.com/news/retail/2025/65percent-of-grocery-retailers-lack-real-time-supply-chain-data/