Columbia University

Driving Scalable Growth: Database Solutions for ABC Foodmart

Adeel Arif, Depali Kulkarni, Suchakrey Nitisanon, Kristen Campbell, Sana Khan, Malaikah Khan

The Client & Challenge

ABC Foodmart Today

A grocery chain operating two locations in Queens, NY, preparing to expand into Brooklyn with three additional stores.

Current State: Relying on large spreadsheets for inventory, staffing, and vendor management across locations.

The Problem

Operational Inefficiencies:

- Error-prone bulk updates
- Weak auditability and tracking
- No cross-location coordination
- Limited analytical capabilities

Multi-borough expansion requires centralized data infrastructure

Our Solution: A Centralized Relational Database

Six Interconnected Tables

Stores, customers, products, aisles, transactions, and transaction items—all normalized to 3NF standards.

Operational Efficiency

Eliminate spreadsheet errors, streamline inventory management, and coordinate across all locations seamlessly.

Customer Loyalty System

Track purchasing behavior, reward points, and transaction history to build lasting customer relationships.

This infrastructure supports ABC Foodmart's expansion while creating a foundation for data-driven decision making and customer engagement.

Sample Data

grocery_chain_data.csv (163.33 kB)

Detail   Compact   Column   10 of 11 columns ⌄

customer_id	store_name	transactio...	aisle	product_n...	quantity	unit_price
2824	GreenGrocer Plaza	2023-08-26	Produce	Pasta	2	7.46
5506	ValuePlus Market	2024-02-13	Dairy	Cheese	1	1.85
4657	ValuePlus Market	2023-11-23	Bakery	Onions	4	7.38
2679	SuperSave Central	2025-01-13	Snacks & Candy	Cereal	3	5.5
9935	GreenGrocer Plaza	2023-10-13	Canned Goods	Orange Juice	5	8.66
7912	Corner Grocery	2023-08-31	Dairy	Onions	3	25.04
1488	GreenGrocer Plaza	2023-10-18	Produce	Ground Beef	3	29.56
4582	City Fresh Store	2024-02-03	Health & Wellness	Orange Juice	2	11.73
9279	GreenGrocer Plaza	2024-11-16	Household Items	Apples	5	22.13
1434	QuickStop Market	2024-10-15	Health & Wellness	Onions	2	20.84
4257	FreshMart Downtown	2025-02-13	Canned Goods	Tomatoes	3	2.91

The dataset contains **1,980 rows across 11 columns**, sourced from Kaggle's grocery store sales records. Key fields include customer_id, store_name, product details, quantity, transaction_date, and pricing information.

This data naturally maps to normalized tables supporting ABC Foodmart's expansion needs while enabling customer behavior analysis and business metrics tracking.

Data Source

Kaggle Grocery Store Sales Dataset

Structured to support multi-store operations and analytical queries

Database Schema Design

Design Principles

Our schema follows **Third Normal Form (3NF)** to minimize redundancy and ensure data integrity at scale.

01

Entity Identification

Customers, stores, products, aisles, transactions, transaction items

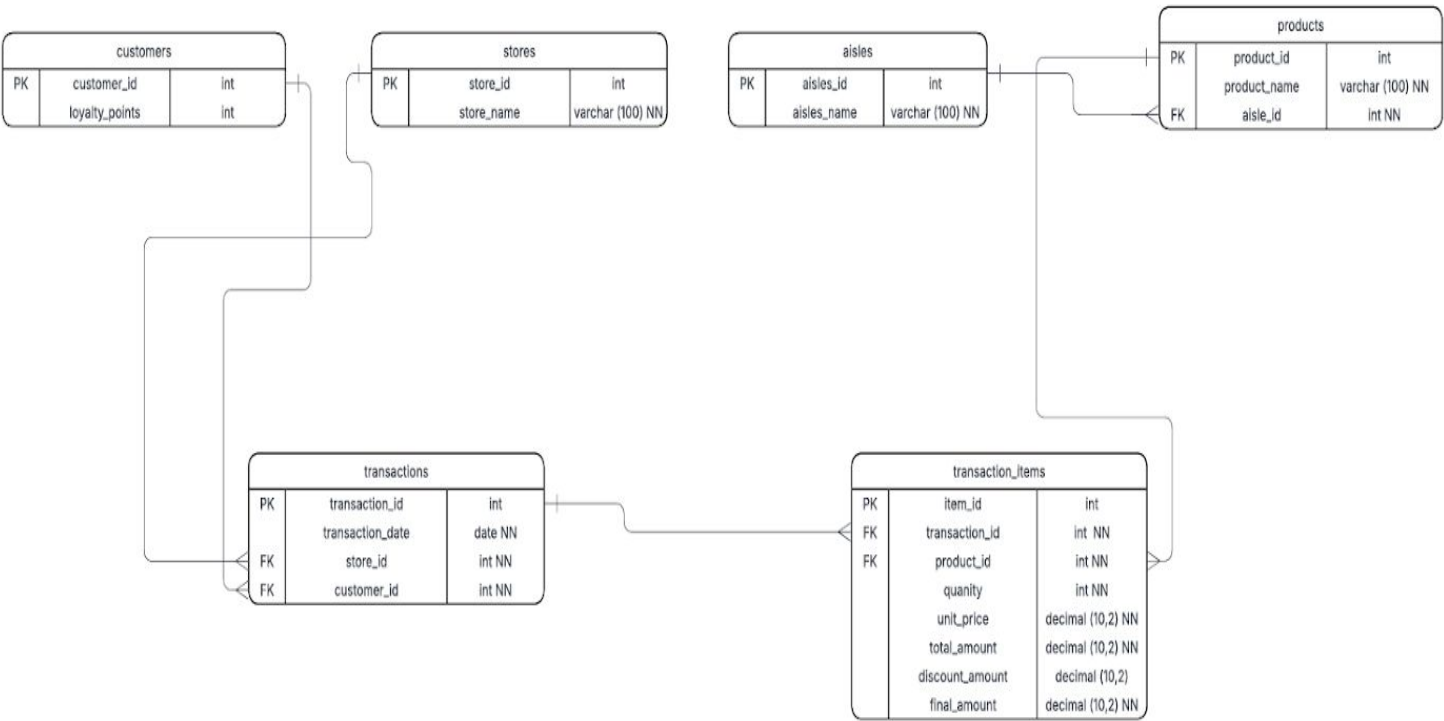02

Relationship Mapping

One-to-many relationships between core entities via foreign keys

03

Extensibility

Room to add staffing, vendor, and supply chain data without breaking existing structure

Entity Relationship Diagram showing table connections and foreign key relationships

The ETL Pipeline

1 **Extract**

Load raw Kaggle dataset into Pandas DataFrame for quality

assessment

2 **Transform**

Correct product-to-aisle mappings, normalize data into six tables,

preserve original customer IDs

3 **Load**

Use SQLAlchemy to insert data into PostgreSQL following foreign

key dependency order

Critical Fix: Manual correction of incorrect product-aisle

assignments (e.g., pasta was incorrectly categorized under produce)

ensured data integrity before database loading.

Customer Needs: What & Why

For Analysts

- **Sales Performance Analysis**
  Understand which stores, products, and aisles drive revenue

- **Customer Behavior Insights**
  Track purchasing patterns, repeat visits, and loyalty metrics

- **Discount Effectiveness**
  Measure promotion usage on sales volume

- **Time-Based Trends**
  Identify daily, weekly, and seasonal sales patterns for inventory planning

For Executives

- **Real-Time Performance Monitoring**
  Track KPIs across all locations without technical expertise

- **Strategic Decision Support**
  Data-driven insights for staffing, inventory, and expansion planning

- **Operational Efficiency**
  Identify low-performing areas and optimize resource allocation

- **Accessible Analytics**
  Visual dashboards replace complex SQL queries

Analytical Tools: The How

1

PostgreSQL + pgAdmin4

For Analysts: Direct database

access with 16 pre-built SQL

scripts answering key business

questions

- Store revenue analysis

- Product revenue rankings

- Customer spending patterns

- Discount impact assessment

2

Metabase Dashboards

For Executives: Interactive

visualizations with no SQL

knowledge required

- Real-time sales monitoring

- Aisle traffic visualization

- Store comparison metrics

- Automated email reports

3

Python ETL Scripts

For Data Engineers: Automated

data transformation and loading

using Pandas and SQLAlchemy

- Data validation and cleaning

- Dependency-aware loading

- Error handling and logging

Analytical Procedures: Business Questions Answered Through SQL

Store Performance & Revenue Optimization

Understand how each store performs financially and operationally using key metrics such as:

Total revenue and revenue after discounts

Average order value (AOV) and transaction volume

Store-level discount usage and its impact on earnings

Product & Aisle Sales Intelligence

Understand what drives sales performance across product lines and categories using insights such as:

Top-revenue products and top-selling SKUs
 Aisle-level revenue and units sold
 Daily sales trends for inventory and demand planning

Customer Behavior & Loyalty Insights

Understand customer value and engagement patterns using insights such as:

Highest-spending and highest-value customers
 Repeat visit frequency across locations
 Lifetime spend patterns for loyalty and retention

SQL Insight Deep Dive: Post-Discount Revenue Analysis

Query Purpose

Calculate each store's total revenue after subtracting all discounts.

Key Fields Used

Store-Level Fields

store_id, store_name

Transaction Fields

final_amount, discount_amount

Logic Summary

Revenue After Discount =
SUM(final_amount - discount_amount)

store_name character varying (100) 🔒	revenue_after_discount numeric 🔒
ValuePlus Market	15896.97
FreshMart Downtown	15131.58
GreenGrocer Plaza	14773.07
Corner Grocery	14198.28
MegaMart Westside	13187.62

What the Business Can Learn

Identify the **true** top-performing stores after removing discount inflation.

Understand whether current discount strategies are boosting sales or silently eroding margins.

Impact on Decision-Making

Leadership can redesign or reduce discount programs that aren't generating real revenue uplift.

High-performing stores *(post adjustment)* become models for operational best practices and resource allocation.

Dashboard Insights in
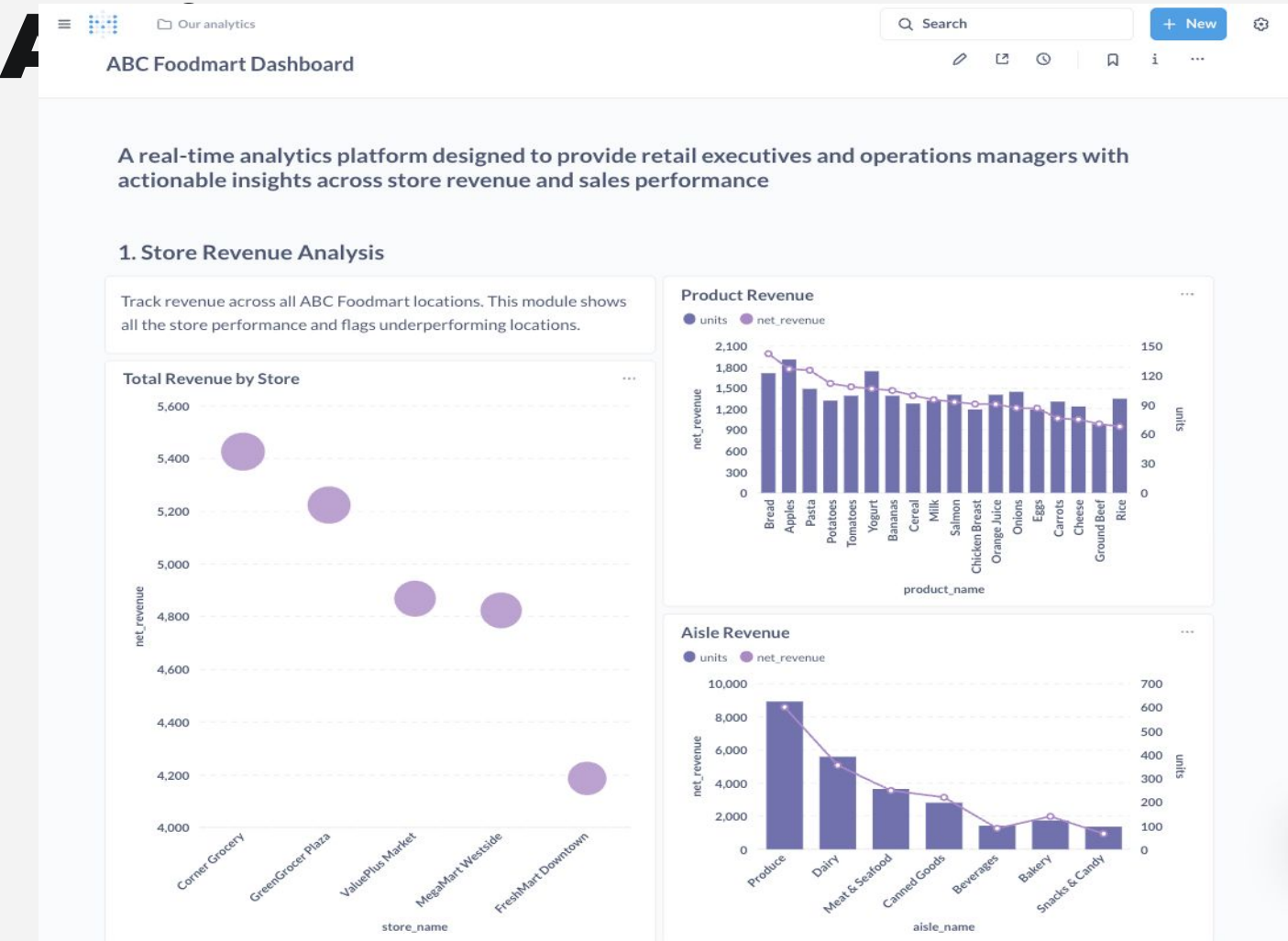
Daily Revenue Tracking
Monitor store performance in real-time with customizable date rage, location, product, and aisle breakdowns

Aisle Traffic Analysis
Identify high and low-traffic areas to optimize product placement and promotional strategies

Promotion Effectiveness
Track promotion usage by store with customized filters such as product and aisle to refine pricing strategies

Dashboard Demo

Impact & Next Steps

What We Delivered

☐ **Replaced error-prone spreadsheets** with a centralized PostgreSQL database following 3NF standards

☐ **Built a Python ETL pipeline** with data validation, transformation, and dependency-aware loading

☐ **Created 16 SQL analytical procedures** for deep-dive analysis of sales, customers, and products

☐ **Designed executive dashboards** in Metabase for real-time monitoring without technical skills

Business Impact

Scalability: Infrastructure ready to support Brooklyn expansion and beyond

Data Integrity: Elimination of spreadsheet errors and improved auditability

Faster Insights: Real-time analytics enable agile decision-making

Customer Loyalty: Transaction tracking powers rewards programs