Retail Sales Data Analysis using SQL

A comprehensive PostgreSQL project exploring retail sales patterns through data cleaning, exploration, and actionable business insights

Prepared by: Sumit Arora

Project Overview

Data Foundation

PostgreSQL database with comprehensive retail sales records including transactions, customers, and product categories

Data Quality

Rigorous cleaning process to identify and remove NULL values, ensuring data integrity

Business Insights

Strategic SQL queries revealing customer behavior, sales patterns, and profitability metrics

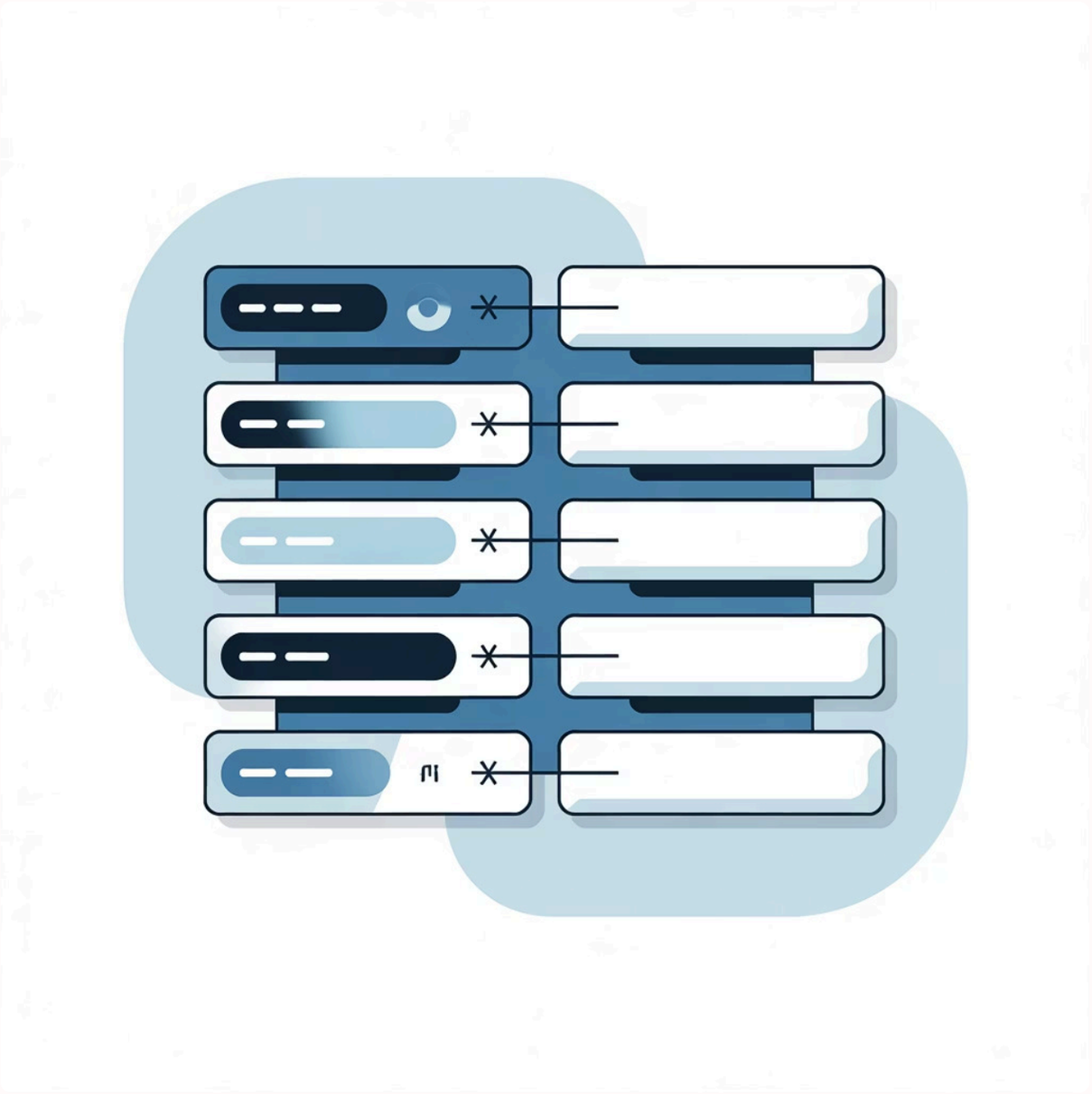
Tools Used: PostgreSQL, SQL, Excel/CSV, pgAdmin 4

Database Structure

Table Schema

The retail_sales table captures complete transaction details:

- Transaction ID (Primary Key)
- Sale date and time
- Customer demographics
- Product category
- Pricing and cost data



Data Cleaning Process

01

Identify NULL Values

Comprehensive scan across all columns to detect missing or incomplete data

02

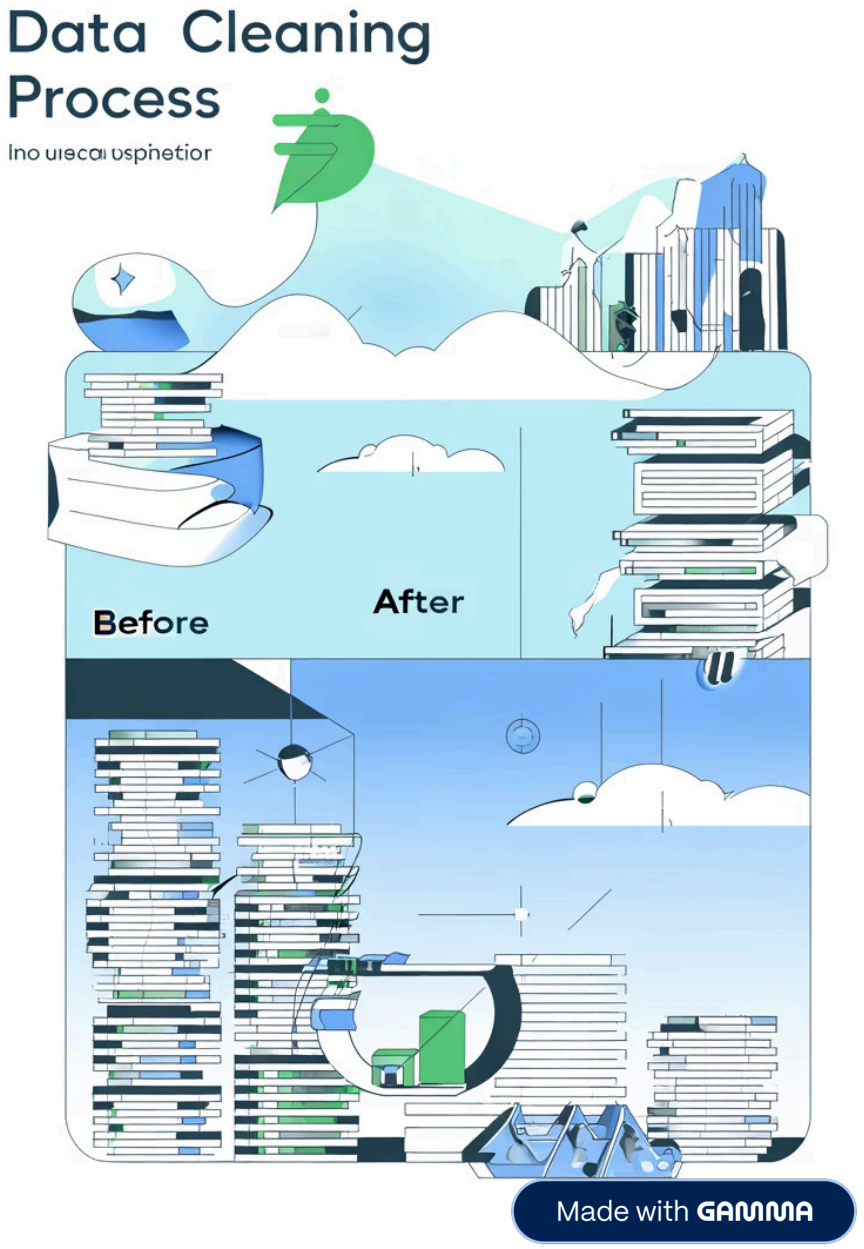
Validate Data Quality

Check for inconsistencies in quantity, pricing, and cost fields

03

Remove Invalid Records

Delete rows with NULL values in critical fields to ensure analysis accuracy



Data Exploration Findings

1,997

Total Sales

Complete transactions analyzed

155

Unique Customers

Individual shoppers tracked

3

Product Categories

Distinct merchandise types

Customer Insights

Top 5 High-Spending Customers

Identified through aggregated total_sale analysis, revealing key revenue drivers and loyalty patterns

Age Group Spending Analysis

- Teenagers (under 18)
- Young Adults (18-25)
- Early Professionals (26-35)
- Middle Age (36-45)
- Mature Adults (46-60)
- Senior Citizens (60+)

Sales Pattern Analysis

1

Category Performance

Total sales aggregated by product category, revealing top-performing merchandise segments

2

Peak Shopping Hours

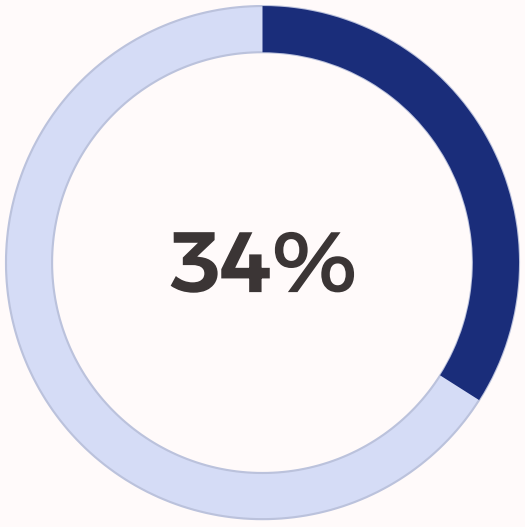
Busiest hour identified through transaction time analysis, optimizing staffing decisions

3

Weekend vs Weekday

Sales comparison showing customer shopping behavior patterns across the week

Profitability Metrics



Profit Margin

Overall profitability calculated from total sales minus cost of goods sold

Category Contribution

Percentage breakdown of each category's contribution to total revenue

Maximum Sales Day

Identified the date and category combination generating highest revenue through aggregated analysis

Key SQL Techniques Applied

Aggregation Functions

GROUP BY, HAVING, SUM, AVG, and COUNT for data summarization

Time-Based Analysis

EXTRACT and TO_CHAR functions for temporal pattern discovery

CASE Statements

Customer segmentation and conditional logic for age groups and day categories

Project Outcomes & Next Steps

Skills Strengthened

SQL proficiency, data cleaning, aggregation, and business analysis capabilities enhanced

Business Insights

Actionable findings on customer behavior, sales patterns, and profitability drivers

Future Plans

Visualizing insights using Power BI or Tableau for enhanced stakeholder communication