

■ Dunkin' Donuts Sales Summary & Analytics System

Complete Project Documentation

Generated: December 30, 2025

1. Project Overview

The Dunkin Sales Summary & Analytics System is a comprehensive data pipeline and analytics dashboard designed to track, analyze, and visualize sales performance across multiple Dunkin' Donuts franchise locations. The system automates data collection, transformation, storage, and visualization to provide actionable insights for business decision-making.

Key Features:

- Automated data pipeline for processing daily sales reports
- Multi-store performance tracking and comparison
- Interactive web-based dashboard with multiple analytics views
- Real-time data synchronization with Supabase cloud database
- Comprehensive reporting covering sales, labor, tender types, and product mix
- Duplicate detection and data validation
- PDF export capabilities for all reports

2. System Architecture

The system follows a three-tier architecture consisting of:

Component	Technology	Purpose
Data Collection	Python Scripts	Download reports from Gmail/Manual
Data Processing	pandas, openpyxl	Parse & transform Excel files
Data Storage	Supabase (PostgreSQL)	Cloud database for all metrics
Dashboard	Streamlit	Interactive web-based analytics
Visualization	Plotly	Charts and graphs

3. Database Schema & Tables

The system uses 6 primary tables in Supabase (PostgreSQL) to store different aspects of store performance. All tables share common identifying columns: store, pc_number, and date.

3.1 *sales_summary*

Primary sales metrics and financial performance

Column	Type	Description
id	INTEGER	Primary key
store	TEXT	Store name/location
pc_number	TEXT	Profit center number
date	DATE	Transaction date
gross_sales	REAL	Total sales before deductions
net_sales	REAL	Sales after discounts/refunds
dd_adjusted_no_markup	REAL	DD adjusted sales (no markup)
pa_sales_tax	REAL	Pennsylvania sales tax
dd_discount	REAL	Total discounts applied
guest_count	INTEGER	Number of customers
avg_check	REAL	Average transaction amount
gift_card_sales	REAL	Gift card purchases
void_amount	REAL	Total voided transactions
refund	REAL	Total refunds issued
void_qty	INTEGER	Number of void transactions
cash_in	REAL	Cash received

3.2 *sales_by_daypart*

Sales performance by time of day

Column	Type	Description
id	INTEGER	Primary key
store	TEXT	Store name/location
pc_number	TEXT	Profit center number
date	DATE	Transaction date
daypart	TEXT	Time period (Morning, Afternoon, Evening)
net_sales	REAL	Sales for this daypart
percent_sales	REAL	% of total daily sales
check_count	INTEGER	Number of transactions
avg_check	REAL	Average transaction amount

3.3 tender_type_metrics

Payment method breakdown

Column	Type	Description
id	INTEGER	Primary key
store	TEXT	Store name/location
pc_number	TEXT	Profit center number
date	DATE	Transaction date
tender_type	TEXT	Payment method (Visa, Cash, etc.)
detail_amount	REAL	Amount paid via this method

3.4 labor_metrics

Employee hours and payroll tracking

Column	Type	Description
id	INTEGER	Primary key
store	TEXT	Store name/location
pc_number	TEXT	Profit center number
date	DATE	Transaction date
labor_position	TEXT	Job role/position
reg_hours	REAL	Regular hours worked
ot_hours	REAL	Overtime hours worked
total_hours	REAL	Total hours (reg + OT)
reg_pay	REAL	Regular pay amount
ot_pay	REAL	Overtime pay amount
total_pay	REAL	Total pay (reg + OT)
percent_labor	REAL	Labor cost as % of sales

3.5 sales_by_order_type

Sales breakdown by service channel

Column	Type	Description
id	INTEGER	Primary key
store	TEXT	Store name/location
pc_number	TEXT	Profit center number
date	DATE	Transaction date
order_type	TEXT	Service type (Dine-in, Drive-thru, etc.)
net_sales	REAL	Sales for this order type
percent_sales	REAL	% of total daily sales

guests	INTEGER	Number of customers
percent_guest	REAL	% of total daily guests
avg_check	REAL	Average transaction amount

3.6 sales_by_subcategory

Product category performance tracking

Column	Type	Description
id	INTEGER	Primary key
store	TEXT	Store name/location
pc_number	TEXT	Profit center number
date	DATE	Transaction date
subcategory	TEXT	Product category (Coffee, Donuts, etc.)
qty_sold	INTEGER	Quantity of items sold
net_sales	REAL	Sales for this subcategory
percent_sales	REAL	% of total daily sales

4. Data Pipeline Process

The data pipeline consists of multiple stages that transform raw Excel reports into structured, analyzable data in the cloud database.

4.1 Data Collection

Reports are downloaded from Gmail or manually placed in the /data/raw_emails directory. Six report types are required for each day:

1. Labor Hours Report - Employee schedules and payroll
2. Sales by Daypart - Time-based sales breakdown
3. Sales by Subcategory - Product mix performance
4. Tender Type Report - Payment method details
5. Sales Mix Detail (Sales Summary) - Overall financial metrics
6. Menu Mix Metrics (Order Type) - Service channel breakdown

4.2 Data Processing

The batch_processor.py script orchestrates the entire pipeline: • Scans for new report files • Validates that all 6 report types are present • Parses Excel files using specialized flattening functions • Normalizes store names and formats • Validates data integrity • Checks for duplicates before upload

4.3 Data Storage

Processed data is uploaded to Supabase (PostgreSQL) with: • Automatic duplicate detection • Transaction-based uploads for data integrity • Error logging and retry mechanisms • Connection pooling for performance

4.4 Key Scripts & Files

Script	Purpose
batch_processor.py	Main pipeline orchestrator - processes multiple dates
compile_store_reports.py	Parse and flatten Excel reports
load_to_sqlite.py	Upload data to Supabase database
download_from_gmail.py	Auto-download reports from Gmail
cleanup_duplicates.py	Remove duplicate entries
check_supabase_data.py	Validate data in database

5. Dashboard & Analytics

The Streamlit-based dashboard provides 9 comprehensive analytical views, each focusing on specific aspects of business performance.

Page	Primary Table(s)	Key Metrics
Executive Summary	sales_summary	Net Sales, Guest Count, Avg Check, Discounts
Sales Mix	sales_by_order_type, sales_by_subcategory	Order Type %, Product Mix, Guest Distribution
Daypart Analysis	sales_by_daypart	Morning/Afternoon/Evening Sales, Transaction Patterns
Labor Efficiency	labor_metrics	Hours Worked, Labor Cost, Labor % by Position
Tender Type	tender_type_metrics	Payment Method Breakdown, Card vs Cash
Store Comparison	All tables	Multi-store Performance Comparison
Cash Reconciliation	sales_summary, tender_type_metrics	Cash In, Expected vs Actual
Location Metrics	sales_summary	Store-specific KPIs and Trends
Payroll Metrics	labor_metrics	Detailed Payroll Analysis

5.1 Common Dashboard Features

- Multi-store filtering with checkbox selections
- Flexible date range or single-day analysis
- Interactive Plotly charts (bar, pie, line, stacked)
- Real-time data from Supabase
- PDF export for each page
- Responsive layout for different screen sizes
- Color-coded metrics and visualizations

6. Stores & Locations

The system currently tracks 7 Dunkin' Donuts franchise locations across Pennsylvania:

Store Name	PC Number	Location
Paxton	301290	Paxton, PA
Mount Joy	343939	Mount Joy, PA
Enola	357993	Enola, PA
Columbia	358529	Columbia, PA
Lititz	359042	Lititz, PA
Marietta	363271	Marietta, PA
E-Town	364322	Elizabethtown, PA

7. Technical Stack

Category	Technology	Version/Details
Language	Python	3.13+
Database	Supabase	PostgreSQL cloud
Dashboard	Streamlit	Web framework
Data Processing	pandas	Data manipulation
Excel Parsing	openpyxl	Excel file handling
Visualization	Plotly	Interactive charts
PDF Generation	ReportLab, WeasyPrint	Report exports
Version Control	Git	Source control

8. Daily Usage Workflow

The typical daily workflow for adding new data:

1. Download reports - Obtain all 6 report types from data source
2. Place in directory - Save files to /data/raw_emails
3. Run batch processor - Execute: python scripts/batch_processor.py
4. Select date - Choose interactive mode or provide date parameter
5. Validation - Script validates all files are present
6. Processing - Automatic parsing and data transformation
7. Upload - Data uploaded to Supabase with duplicate checks
8. Verification - Check dashboard for new data
9. Analysis - Use dashboard pages to analyze performance

9. Data Quality & Validation

The system includes multiple layers of data validation to ensure accuracy:

- File presence validation - Ensures all 6 report types exist
- Date parsing verification - Validates date formats in filenames
- Duplicate detection - Prevents re-uploading existing data
- Schema validation - Ensures columns match expected structure
- Data type checking - Validates numeric and date fields
- Null handling - Manages missing values appropriately
- Store name normalization - Standardizes location names
- Transaction integrity - Uses database transactions for uploads

10. Project Directory Structure

The project follows a well-organized directory structure:

```
/dashboard/ - Streamlit web application
/pages/ - Individual dashboard pages
/components/ - Reusable UI components
/utils/ - Database and utility functions
app.py - Main dashboard entry point
/scripts/ - Data processing scripts
batch_processor.py - Main pipeline orchestrator
compile_store_reports.py - Excel parsing logic
load_to_sqlite.py - Database upload functions
check_*.py - Various validation scripts
/data/ - Data storage
/raw_emails/ - Downloaded Excel reports
/compiled/ - Processed CSV files
/db/ - Database schemas and SQL
sales_schema.sql - Table definitions
init_db.py - Database initialization
/logs/ - Processing logs and error tracking
/exports/ - Generated reports and exports
```

11. Potential Enhancements

- Automated email report fetching with scheduled jobs
- Real-time alerts for anomalies (low sales, high discounts, etc.)
- Predictive analytics using machine learning
- Mobile-responsive dashboard improvements
- Advanced forecasting models for inventory and staffing
- Integration with POS systems for real-time data
- Automated report generation and distribution
- Historical trend analysis with year-over-year comparisons
- Customer segmentation analysis
- Advanced labor scheduling optimization

12. Summary

The Dunkin Sales Summary & Analytics System provides a comprehensive solution for managing and analyzing sales data across multiple franchise locations. By automating data collection, processing, and visualization, the system enables data-driven decision making and provides valuable insights into business performance.

The modular architecture ensures maintainability and scalability, while the use of modern cloud technologies (Supabase) guarantees data accessibility and reliability. The interactive dashboard transforms raw sales data into actionable insights, helping franchise owners optimize operations, manage labor costs, understand customer behavior, and maximize profitability.

With 6 comprehensive database tables tracking everything from sales and labor to payment methods and product mix, the system provides a 360-degree view of business operations. The combination of automated processing, robust validation, and intuitive visualization makes this a powerful tool for franchise management.