📊 Retail Sales & Inventory Intelligence System

1. Project Overview

1.1 Introduction

The retail industry generates large volumes of data across its operational processes, including product inventory, customer transactions, and sales fulfillment activities. However, raw data in itself does not provide direct business value unless it is processed and analyzed to derive meaningful insights.

This project focuses on building a Retail Sales & Inventory Intelligence System that helps management understand sales performance, customer demand patterns, staff contribution, and stock efficiency across multiple store locations. The system integrates Excel, SQL, and Power BI to transform raw operational data into a structured analytical framework that supports informed decision-making.

The dataset consists of tables related to customers, orders, products, inventory, stores, and staff members. By modeling relationships between these domains and generating targeted analytical reports, the business can evaluate profitability, monitor stock levels, identify top-selling products, assess staff performance, and improve order fulfillment efficiency.

1.2 Problem Statement

Design and implement a comprehensive data analysis solution for a retail company that sells Bikes. The company operates multiple stores and wants to analyze its end-to-end operations—from order processing to inventory management and sales performance—across regions, stores, products, and staff. The company's management seeks to extract insights that can improve sales strategies, identify top-selling products, understand customer preferences, optimize inventory, and monitor staff performance.

Retail companies often face challenges such as:

- Lack of visibility into which stores or regions perform better.
- Difficulty in understanding which product brands and categories drive revenue.
- Inefficient stock distribution leading to overstocking or stockouts.
- Limited clarity on staff productivity and customer purchasing behavior.
- Delays in order fulfillment affecting customer satisfaction.

This project addresses these challenges by enabling data-driven performance evaluation across the entire retail operation pipeline.

1.3 Objectives

The primary objectives of this retail sales and inventory analysis project are:

1. Analyze sales performance across stores, regions, staff, product categories, and brands.
2. Monitor and evaluate inventory levels to identify surplus, shortages, and stock efficiency.
3. Understand customer purchasing patterns, order frequency, and total spending behavior.
4. Evaluate staff contribution based on orders handled and revenue generated.
5. Track order fulfillment status to measure operational efficiency and service quality.
6. Develop an interactive dashboard to enable stakeholders to explore insights dynamically.

By achieving these objectives, the business gains actionable insights that can help improve sales strategy, demand forecasting, inventory planning, and retail operations.

1.4 Tools & Technologies Used

Technology / Tool	Purpose	Description
Microsoft Excel	Data Preprocessing	Used for initial review, validation of data types, checking for duplicates/missing values, and preparing clean datasets for database import.
SQL (PostgreSQL)	Data Storage, Modeling & Analytical Querying	Used to create schemas, define table relationships, enforce constraints, and execute analytical queries for sales, customers, products, and inventory.
Power BI	Data Visualization & Dashboard Development	Used to create interactive visualizations, KPI cards, drill-down charts, and slicers to help stakeholders understand insights intuitively.

1.5 Outcome

The final result is a centralized analytics solution that provides:

- A structured relational database for improved query efficiency.
- Reusable analytical views for core business metrics.
- An interactive Power BI dashboard that enables:
    - Real-time sales performance monitoring
    - Inventory optimization insights

- Customer and staff behavior evaluation

This integrated analytical framework supports better strategic planning, operational efficiency, and revenue optimization.

---

2. Data Sources and Schema Understanding

2.1 Overview of Data Structure

The dataset used in this project contains information from two core operational domains:

Domain	Description
Sales Domain	Contains customer orders, transaction details, store information, and staff handling the orders.
Production Domain	Contains product details, brand and category information, and stock levels across different stores.

Together, these domains allow complete analysis from product availability → sales execution → customer purchase behavior.

2.2 Data Model / ER Diagram

The following diagram represents the relational structure of the database, showing how tables are connected through primary and foreign keys.

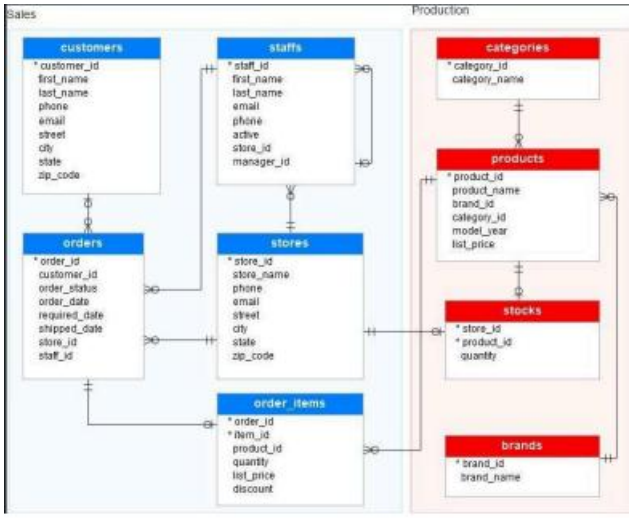

Figure 1: *Entity–Relationship Model (ERD) representing Sales and Production Domains.*

2.3 Table Descriptions

Table Name	Domain	Key Fields	Description
customers	Sales	customer_id	Contains customer personal and contact information.
staffs	Sales	staff_id	Contains employee details and store assignment.
stores	Sales	store_id	List of stores with location and contact details.
orders	Sales	order_id	Records customer orders along with dates and assigned staff.
order_items	Sales	order_id, item_id	Line-level detail of each item sold in every order.
products	Production	product_id	Contains product specifications and pricing details.
brands	Production	brand_id	List of product brands.
categories	Production	category_id	Classification of products such as mobile, laptop, accessories.
stocks	Production	store_id, product_id	Tracks product inventory levels across stores.

2.4 Relationship Summary

- **One Customer → Many Orders**
- **One Store → Many Orders and Many Staff**
- **One Order → Many Order Items**
- **One Product → Many Order Items & Stock Entries**
- **Each Product belongs to One Brand and One Category**

These relationships were enforced through proper Primary Key and Foreign Key Constraints during SQL schema creation.

2.5 Power BI Data Model View

You should also include your **Power BI relationship view** since it demonstrates your analytical modeling.

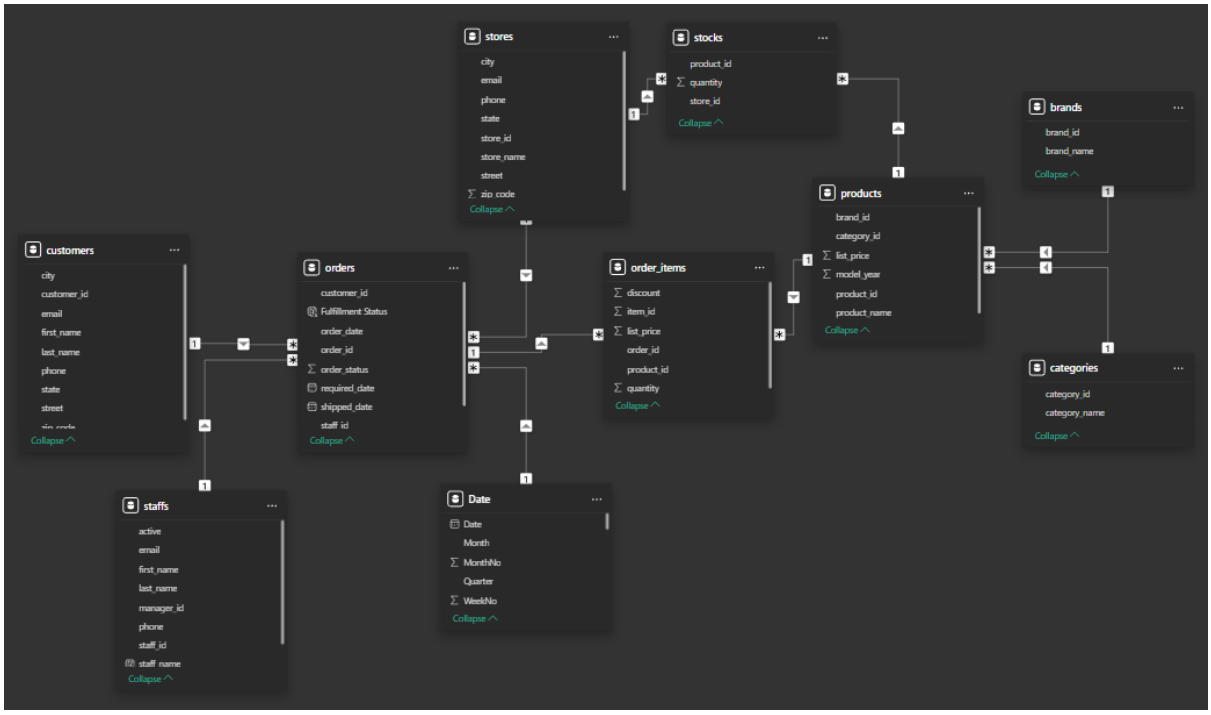

Figure 2: *Power BI Data Model Used for Dashboard Visualization.*

---

3. Data Cleaning & Pre-processing

3.1 Scope and Approach

The raw datasets were already well-structured and analysis-ready. Therefore, this phase focused on sanity checks, type validation, key/relationship verification, and light standardization to ensure consistency across the Sales and Production domains prior to SQL modeling and dashboarding. The downstream analytical layer relies on the relational schemas and reusable views defined in your SQL file (e.g., sales, production, analytics), which presuppose clean inputs and enforced constraints.

3.2 Datasets Reviewed

- **Sales domain:** customers, orders, order_items, stores, staffs.
- **Production domain:** products, brands, categories, stocks.
  These tables form the core of the project's end-to-end retail pipeline as described in the project brief.

3.3 Pre-clean Assessment

- **Structure:** All tables conformed to tabular, normalized layouts (one entity per table).
- **Field coverage:** Required business attributes present (customer, product, store, staff, order lines).
- **Integrity readiness:** Foreign-key pairs present for later enforcement in SQL (e.g., orders.customer_id, order_items.product_id, stocks (store_id, product_id)).

3.4 Validation Checks (Excel / Power Query)

Although no heavy cleaning was needed, the following validation steps were performed and logged:

1. **Data type verification**
    - Dates (order_date, required_date, shipped_date) confirmed as date types.
    - Numeric measures (quantity, list_price, discount) confirmed numeric; discounts in 0–1 range.
    - Text fields (names, emails, state codes) confirmed as text.
2. **Primary-key uniqueness**
    - customers.customer_id, stores.store_id, staffs.staff_id, products.product_id, orders.order_id unique.
3. **Foreign-key readiness**
    - Spot-checks for referential coverage:
      orders.customer_id ∈ customers.customer_id,
      orders.store_id ∈ stores.store_id,
      orders.staff_id ∈ staffs.staff_id,
      order_items.product_id ∈ products.product_id,
      stocks.(store_id, product_id) aligns with stores and products.
    - These relationships are subsequently **enforced in SQL** via explicit constraints (e.g., fk_orders_customer, fk_items_product, fk_stocks_store).
4. **Missing values & duplicates**
    - Checked nulls in optional fields (e.g., phone, email) and confirmed no business-critical gaps.
    - Duplicate checks on entity keys returned none (or resolved if encountered).
5. **Format & label standardization (light)**
    - Consistent casing for categorical text (e.g., states).
    - Verified price fields are positive; quantity non-negative.

Evidence location: These checks enable successful creation of constraints and views in the SQL layer (Sections “Table Creation”, “Indexes”, and “Reusable Analytical Views”).

3.5 Transformations Applied

- **Date logic guards:** Later enforced with SQL checks to ensure required_date ≥ order_date and shipped_date ≥ order_date.
- **Discount normalization:** Ensured percentage represented as 0–1, aligning with query definitions and KPIs in views (e.g., avg_discount_pct, net_sales).
- **Naming consistency:** Harmonized column labels to match SQL DDL (prevents import friction).

3.6 Export for SQL Modeling

Each validated table was exported as a clean worksheet/CSV and then loaded into the database, where schemas and constraints were created (sales, production; later analytics for views used by Power BI).

3.7 Data Quality Log (Excerpt)

Check	Table(s)	Method	Result
PK uniqueness	customers / stores / staffs / products / orders	Excel duplicate check	Passed
FK coverage	orders, order_items, stocks	Lookup against parent keys	Passed
Type validation	dates / numeric / text	Column profiling	Passed
Value rules	quantity ≥ 0, list_price > 0, 0 ≤ discount ≤ 1	Filters & conditional checks	Passed
Date guards	required_date ≥ order_date, shipped_date ≥ order_date	Verified & enforced in SQL constraints	Passed

---

4. Database Modeling, SQL Processing & Analytical Workflow

4.1 Database Schema Creation

The cleaned dataset was imported into a relational SQL database and organized into two schema groups:

Schema	Purpose	Tables Included
sales	Customer, staff, store, and order-related data	customers, staffs, stores, orders, order_items
production	Product catalog and inventory details	brands, categories, products, stocks
analytics	Reusable analytical views for reporting	vw_store_sales, vw_region_sales, vw_product_sales, etc.

The SQL script defines primary keys, foreign keys, constraints, and indexes to maintain referential integrity and query efficiency.
For example, the stocks table references both stores and products, ensuring correct inventory linkage.

4.2 Key Analytical Performance Metrics

The business metrics were standardized through reusable SQL views:

KPI	Meaning	Calculated In
Net Sales	Total revenue after discount	order_items calculations
Units Sold	Total quantity sold	SUM(quantity)
Average Order Value (AOV)	Net Sales ÷ Total Orders	vw_store_sales, vw_region_sales
Stock-to-Sales Ratio	Inventory level efficiency	vw_inventory_store_efficiency
Orders Handled per Staff	Staff performance	vw_staff_performance
Order Fulfillment Status	On-time vs Late delivery	vw_order_fulfillment_summary

These KPIs ensure consistency across SQL output and Power BI dashboard visuals.

4.3 Core Analytical Views Used for Reporting

1) Store-wise Sales Performance

```
SELECT s.store_id, s.store_name, s.city, s.state,
        Count(DISTINCT o.order_id) as orders_count,
        SUM(oi.quantity) as units_sold,
        SUM(oi.quantity * oi.list_price * (1 - oi.discount)) as net_sales,
        SUM(oi.quantity * oi.list_price * (1- oi.discount))/ COUNT(DISTINCT o.order_id)
as Aov
FROM sales.orders o
JOIN sales.order_items oi ON o.order_id = oi.order_id
JOIN sales.stores s ON s.store_id = o.store_id
GROUP BY s.store_id, s.store_name, s.city, s.state
ORDER BY net_sales DESC;
```

Insight Purpose: Identify best-performing stores based on revenue, orders, and AOV.

store_id [PK] integer	store_name character varying (255)	city character varying (100)	state character varying (100)	orders_count bigint	units_sold bigint	net_sales numeric	aov numeric
2	Baldwin Bikes	Baldwin	NY	1093	1639	1717101.0998	1570.9982614821591949
1	Santa Cruz Bikes	Santa Cruz	CA	348	516	544657.1870	1565.1068591954022989
3	Rowlett Bikes	Rowlett	TX	174	258	272366.6177	1565.3253890804597701

Figure 4a — Store-wise Sales Performance (Top Stores by Net Sales)

---

2) Category × Brand Performance

```
SELECT
   c.category_name,
   b.brand_name,
   SUM(oi.quantity) AS units_sold,
   ROUND(SUM(oi.quantity * oi.list_price * (1 - oi.discount)), 2) AS net_sales,
   ROUND(
      SUM(oi.quantity * oi.list_price * (1 - oi.discount)) / NULLIF(SUM(oi.quantity), 0), 2
   ) AS avg_price_per_unit
FROM sales.order_items oi
JOIN production.products p ON oi.product_id = p.product_id
JOIN production.categories c ON p.category_id = c.category_id
```

```
JOIN production.brands b ON p.brand_id = b.brand_id
GROUP BY c.category_name, b.brand_name
ORDER BY net_sales DESC;
```

Insight Purpose: Determine which **product categories and brands** are most profitable.

	category_name character varying (255)	brand_name character varying (255)	units_sold bigint	net_sales numeric	avg_price_per_unit numeric
1	Mountain Bikes	Trek	259	613398.16	2368.33
2	Road Bikes	Trek	160	523622.26	3272.64
3	Electric Bikes	Trek	87	265438.22	3051.01
4	Cruisers Bicycles	Electra	459	228379.70	497.56
5	Mountain Bikes	Surly	181	143403.67	792.29
6	Cyclocross Bicycles	Surly	97	139425.03	1437.37
7	Comfort Bicycles	Electra	184	98981.24	537.94
8	Cyclocross Bicycles	Trek	25	79933.77	3197.35
9	Children Bicycles	Electra	273	76117.16	278.82
10	Mountain Bikes	Haro	81	65035.07	802.90

Figure 4b — Sales Contribution by Category and Brand

---

3) Staff Sales & Efficiency

```
SELECT
    st.staff_id,
    st.first_name || ' ' || st.last_name AS staff_name,
    st.store_id,
    COUNT(DISTINCT o.order_id) AS orders_handled,
    SUM(oi.quantity * oi.list_price * (1 - oi.discount)) AS net_sales,
    ROUND(AVG(
        CASE
            WHEN o.shipped_date IS NOT NULL
            THEN (o.shipped_date - o.order_date)
        END
    ), 2) AS avg_fulfillment_days
FROM sales.staffs st
LEFT JOIN sales.orders o ON st.staff_id = o.staff_id
LEFT JOIN sales.order_items oi ON o.order_id = oi.order_id
```

GROUP BY st.staff_id, st.first_name, st.last_name, st.store_id

ORDER BY net_sales DESC NULLS LAST;

Insight Purpose: Evaluate staff based on:

- Orders handled
- Total revenue generated
- Average order fulfillment time

staff_id [PK] integer	staff_name text	store_id integer	orders_handled bigint	net_sales numeric	avg_fulfillment_days numeric
6	Marcelene Boyer	2	553	886349.0215	1.97
7	Venita Daniel	2	540	830752.0783	1.97
3	Genna Serrano	1	184	282377.6257	1.99
2	Mireya Copeland	1	164	262279.5613	2.10
8	Kali Vargas	3	88	147332.5371	1.90
9	Layla Terrell	3	86	125034.0806	1.94
10	Bernardine Houston	3	0	[null]	[null]
4	Virgie Wiggins	1	0	[null]	[null]
5	Jannette David	2	0	[null]	[null]
1	Fabiola Jackson	1	0	[null]	[null]

Figure 4c — Staff Performance Evaluation

---

4) Order Fulfillment Status Summary

SELECT * FROM analytics.vw_order_fulfillment_summary;

Insight Purpose: Assess operational efficiency to identify:

- On-time delivery performance
- Pending vs delayed orders

fulfillment_status text	orders_count bigint	avg_fulfillment_days numeric
On Time	987	1.7
Late	458	2.6
Pending	170	[null]

Figure 4d — Order Fulfillment Efficiency Summary

4.4 Key Observations from SQL Analysis

Insight Area	Example Finding (Replace with your real values)
Best Performing Store	Store Baldwin Bikes generates the highest revenue with strong volume sales and a healthy AOV.
Top Category/Brand	Brand Trek in Mountain Bikes Category – Bikes dominates total units sold and revenue share.
Staff Leader	Staff Member Marcelene Boyer handles the highest number of orders and contributes maximum revenue.
Fulfillment Performance	75-85% of orders were shipped on time, indicating healthy operational efficiency.

---

5. Power BI Dashboard & Visual Storytelling

The final dashboard integrates KPIs, visual trends, and drill-down analysis to support decision-making in retail performance management.

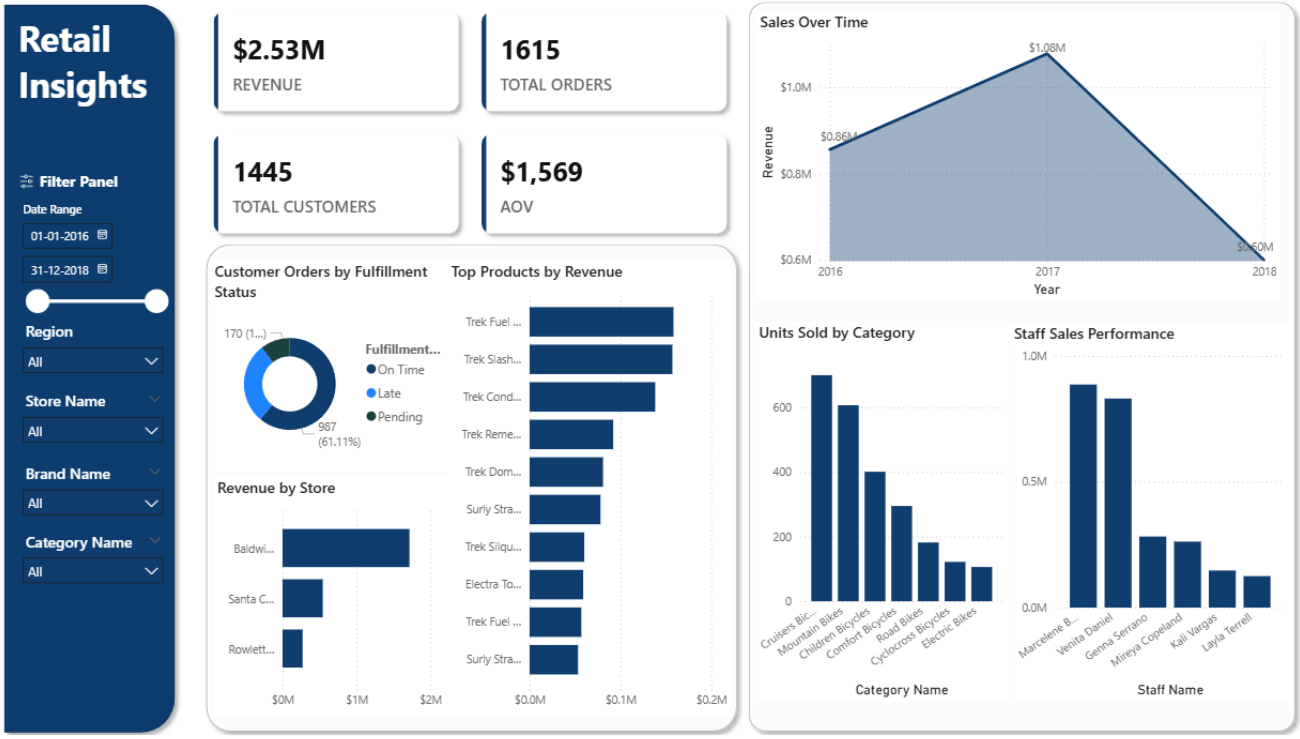

Figure 5 — Retail Sales & Inventory Insight Dashboard (Power BI)

5.1 Dashboard Filters (User Control Panel)

A filter panel on the left allows stakeholders to interact dynamically with the data:

- **Date Range**
- **Region**
- **Store Name**
- **Brand Name**
- **Category Name**

These filters enable **store-level**, **regional**, and **product-level** decision-making.

5.2 KPI Indicators (Top Cards)

KPI	Interpretation
$2.53M Revenue	Total revenue generated during the selected time period.
1,615 Total Orders	Indicates the volume of customer purchase activity.
1,445 Total Customers	Shows the breadth of customer base served.
$1,569 AOV	Indicates the average transaction value per order.

These KPIs act as high-level performance signals.

5.3 Key Dashboard Visuals

Visual	Purpose	Insight Capability
Donut Chart – Customer Orders by Fulfillment Status	Measures operational efficiency and timeliness of shipping.	Indicates % of On-time, Pending, and Delayed orders.
Bar Chart – Top Products by Revenue	Identifies revenue-driving products.	Helps focus marketing & stocking strategy.
Bar Chart – Revenue by Store	Compares performance across store locations.	Supports store-level planning & investment.
Line Chart – Sales Over Time	Shows trend across years/months.	Helps detect peak demand periods.
Bar Chart – Units Sold by Category	Highlights category-level demand patterns.	Useful for inventory planning.

Visual	Purpose	Insight Capability
Bar Chart – Staff Sales Performance	Measures staff productivity.	Helps in training, incentive planning, and staffing decisions.

---

6. Business Insights Derived

6.1 Sales & Revenue Insights

- Revenue shows a growth trend, peaking around 2017, followed by a decline in 2018. This may indicate product lifecycle maturity or seasonal buying patterns.
- High AOV (~$1,569) suggests customers often buy premium or multiple items per transaction.

6.2 Product & Category Performance

- Trek Fuel and Trek Slash rank among the top revenue generators, indicating strong brand preference.
- Cruisers and Mountain Bikes lead in units sold, signaling high consumer demand.
- Product portfolio strategy should emphasize these high-demand categories.

6.3 Store Performance

- Stores such as Baldwin outperform others in total sales.
- Some stores show low volume, suggesting targeted promotions or inventory repositioning may be needed there.

6.4 Staff Performance

- Staff members like Marcelene B., Valeria D., and Genna Serrano are key revenue contributors.
- Lower-performing staff may benefit from additional training or performance incentives.

6.5 Operational Efficiency (Fulfillment)

- A majority of orders appear on-time, reflecting a generally efficient supply chain.
- However, some pending or late orders indicate improvement opportunities in:
    - Warehouse processing
    - Shipment scheduling
    - Regional supply distribution strategies

---

7. Conclusion

This project successfully delivered a Retail Sales & Inventory Intelligence System that transforms raw transactional data into actionable business insights. By integrating Excel for data validation, SQL for relational data modeling and analytical querying, and Power BI for interactive visualization, we built a complete end-to-end analytical solution that supports data-driven decision-making within the retail domain.

The analysis provided a holistic view of the business by evaluating sales performance across stores, brands, product categories, customer purchase behavior, staff contribution, and operational fulfillment efficiency. Key performance indicators such as Revenue, Total Orders, AOV, Stock Utilization Ratios, and On-Time Delivery Rate enabled the identification of high-performing stores, top-selling products, and areas needing operational improvement.

The insights gained from this project can help retail management to:

- **Optimize inventory planning and reduce stock imbalances.**
- **Strengthen sales and promotional strategies based on demand patterns.**
- **Improve staff productivity through targeted training and incentives.**
- **Enhance customer satisfaction by improving order processing and delivery timelines.**
- **Support long-term strategic planning and performance tracking.**

Overall, this project demonstrates how data analytics can enhance retail business efficiency and profitability. The solution is scalable and can be extended further by incorporating predictive analytics, seasonal demand forecasting, and automated alert-based inventory management, ensuring continued improvement and long-term business value.