رواد مصر الرقمية

ANALYTICS REPORT

2025

Super Store Sales

Table Of Contents

Introduction

About the dataset:

In today's competitive retail landscape, data-driven decision-making is essential for businesses aiming to optimize operations, improve customer satisfaction, and maximize profitability. This report presents an in-depth analysis of the **Super Store** dataset, leveraging **Power BI** to uncover actionable insights into sales trends, customer behavior, product performance, and return rates. The goal of this analysis is to assist stakeholders in identifying key areas for improvement and implementing data-backed strategies to drive growth and efficiency.

The **Super Store Data Analysis** Project focuses on evaluating business performance across different dimensions, including:

- **Sales Performance**: Understanding seasonal trends, high-performing regions, and revenue distribution.
- **Customer Segmentation**: Analyzing consumer purchasing behavior to tailor marketing efforts.
- **Product Performance**: Identifying top-selling products and those with high return rates.
- **Profitability Insights**: Assessing the impact of discounts and operational costs on revenue.

Through a structured approach involving **data cleaning**, **modeling**, and **DAX calculations**, this report provides a comprehensive overview of the business, highlighting strengths, weaknesses, and areas for potential optimization.

Objectives of This Analysis

The primary objective of this analysis is to **extract actionable insights** from the sales data that can help in decision-making and operational improvements. By leveraging data analytics, we aim to:

- **Understand Sales Trends:** Identify best-selling products, high-revenue regions, and seasonal sales patterns.
- **Assess Customer Segmentation:** Analyze the behavior of different customer segments to improve marketing strategies.
- **Evaluate Shipping Performance:** Determine the impact of various shipping modes on customer satisfaction and delivery efficiency.
- **Analyze Profitability:** Examine factors influencing revenue, profit margins, and discount strategies.

Business Questions

This analysis aims to answer key business questions that drive decision-making and strategic planning:

1- Sales Performance:

- What are the best-selling products?
- How do sales trends change over time?
- What is the revenue distribution across different product categories?

2- Customer Behavior:

- Who are the top-spending customers?
- What are the common purchasing patterns?
- What is the customer retention rate?

3- Inventory Management:

- Which products need restocking?
- Which products have slow turnover rates?

4- Financial Performance:

- What is the net profit over different periods?
- How do operational costs fluctuate over time?

5- Shipping & Logistics Analysis:

- Which shipping mode is most cost-effective and delivers the highest customer satisfaction?
- Are there any regions where delayed shipments impact sales and customer retention?

6- Regional & Market Expansion Analysis:

- Which states and regions have the highest and lowest sales performance?
- Are there regional preferences for specific product categories?
- What factors contribute to regional variations in profit margins?
- Which geographic areas should the company prioritize for expansion?
- How can advertising and marketing efforts be optimized for underperforming regions?

Dataset Overview

The dataset consists of multiple fields that provide insights into different aspects of sales performance and business operations.

1. Order Information

- **Order ID, Order Date, and Ship Date:** Track individual transactions and fulfillment timelines.
- **Ship Mode:** Categorizes different shipping methods such as **Standard Class, First Class, Second Class, and Same Day,** which impact delivery speed and cost.

2. Customer Data

- **Customer ID and Customer Name:** Identifies unique customers to track repeat purchases**.**
- **Segment:** Classifies customers into three main groups**—Consumer, Corporate, and Home Office—**helping in targeted marketing strategies**.**

3. Product Details

- **Category & Sub-Category:** Provides insights into the different product types, such as **Furniture, Office Supplies, and Technology.**
- **Product ID & Product Name:** Helps in tracking individual items and their performance in sales.

4. Sales & Financial Metrics

- **Sales Revenue:** Total transaction value for each order.
- **Discount & Profit:** Evaluates the impact of discounting strategies on overall profit margins.

5. Geographical Insights

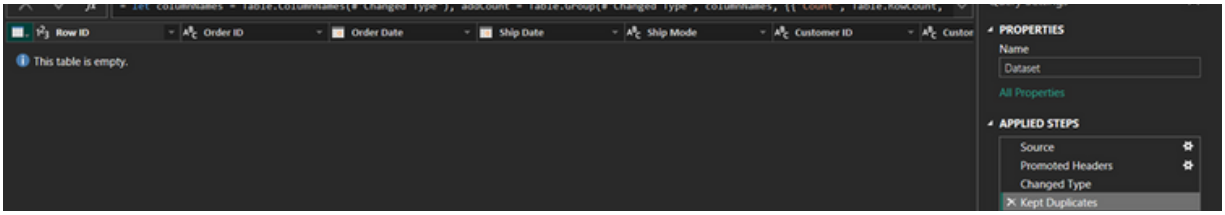
- **City, State, and Region:** Provides location-based insights into sales distribution and market demand.
- **Postal Code:** Useful for geo-based analysis and customer distribution mapping.

Data Exploration, Cleaning, and Modeling Documentation

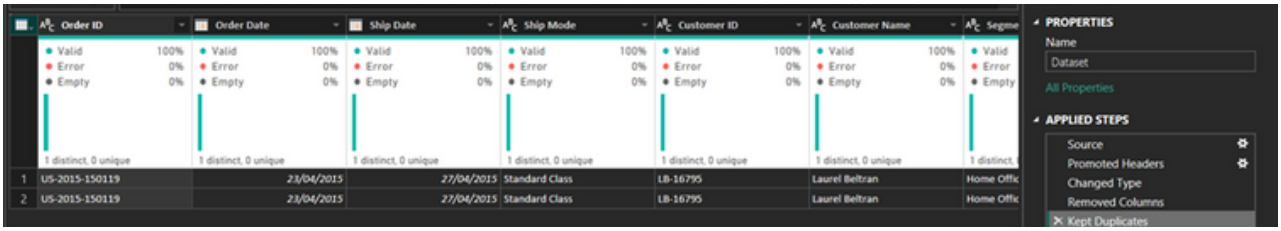
Step 1: Duplicate Records Investigation & Cleanup

Identifying and Handling Duplicates

- **Identifying and Handling Duplicates**



- Initially, no duplicates were found.
- Identified that the "Row ID" column was unnecessary as it only counted rows.
- After removing "Row ID", duplicate records (2 rows) appeared since "Row ID" was uniquely identifying them.



- Removed these 2 duplicate records to maintain data integrity.

Key Actions:

- Removed the "Row ID" column.
- Eliminated 2 duplicate records that were previously hidden by the "Row ID" column.

🔧 Step 3: Creating Dimension Tables for Data Modeling

To follow **best practices for Power BI modeling,** the dataset was split into **dimension tables** to improve data efficiency, maintainability, and performance.

🔹 **Dimension Tables Created**

1. dCustomer (Stores customer-related information)

- Customer ID (Primary Key)
- Customer Name
- Segment

2. dLocation (Stores geographic details)

- Country
- City
- State
- Postal Code (Primary Key)
- Region

3. dShipping (Stores order and shipment details)

- Order ID (Primary Key)
- Ship Mode
- Order Date
- Ship Date
- **Calculated Column:** Delivery Time = Ship Date - Order Date

4. dProduct (Stores product information)

- Product ID
- Category
- Sub-Category
- Product Name (Primary Key)

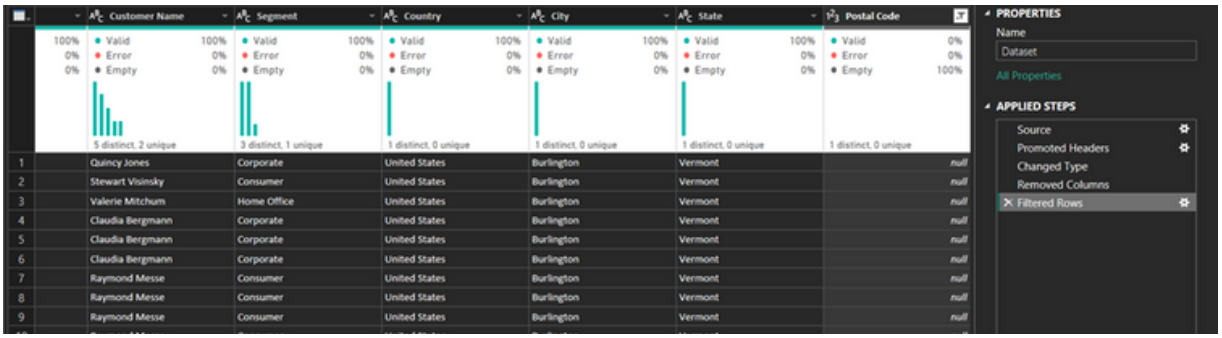
✅ Key Actions:

- Created 4 well-structured dimension tables: dCustomer, dLocation, dShipping, and dProduct.
- Added a calculated column in dShipping for delivery time analysis.
- Ensured each table had a primary key to enable efficient one-to-many relationships with the fact table (fSales).

🔧 Step 2: Handling Missing & Conflicting Data in the Postal Code Column

🔍 Missing Values in Postal Code Column:

- Found 11 missing values in the Postal Code column.

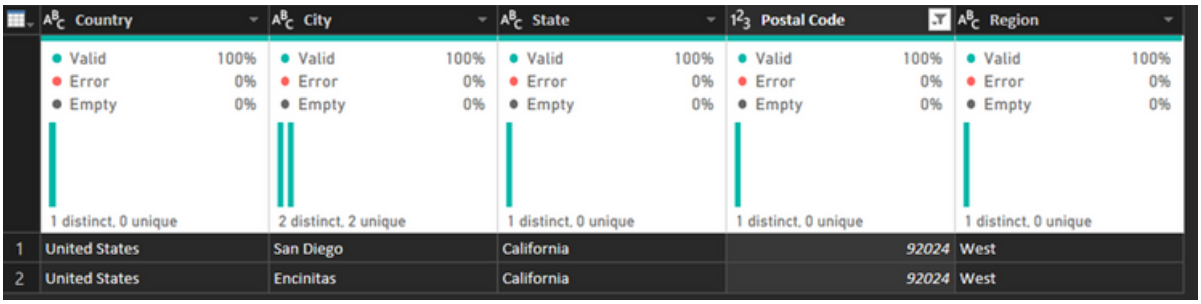


- Upon filtering, noticed that all benged to Burlington, Vermont.
- Replaced the missing postal codes with 5401, the official postal code for Burlington, VT.

🔍 Resolving Many-to-Many Issues in Postal Code & Location Mapping

Identified that two different cities in California shared the same postal code (92024):

- San Diego
- Encinitas



- This caused a many-to-many relationship issue when linking the dLocation table to fSales.
- Corrected the San Diego zip code from 92024 to 92124 (which is its correct zip code).
- This correction enabled a one-to-many relationship between dLocation and fSales.

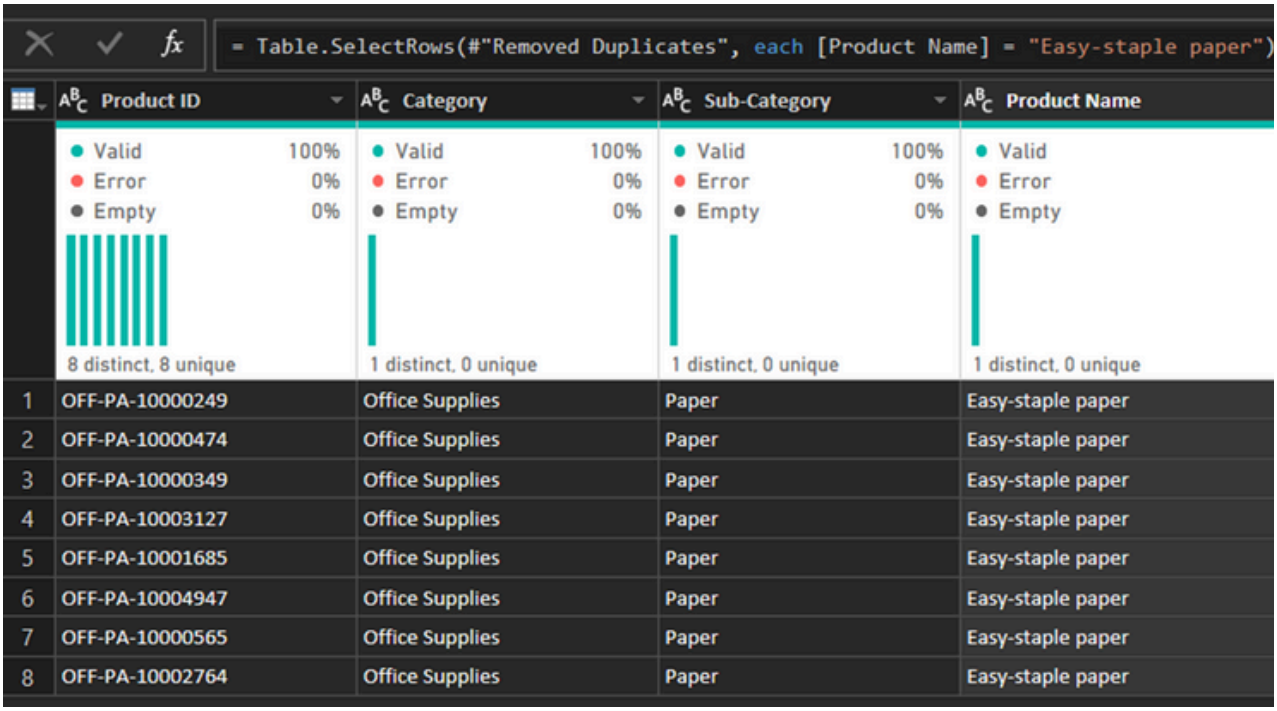
✅ Key Actions:

- Replaced 11 missing postal codes in Burlington, VT.
- Corrected San Diego's postal code from 92024 to 92124.
- Ensured a one-to-many relationship between dLocation and fSales.

🔧 Step 4: Resolving Issues with the Product ID Column

During the data exploration phase, it was observed that the **Product ID** column did not serve as a **true primary key**. The following inconsistencies were identified:

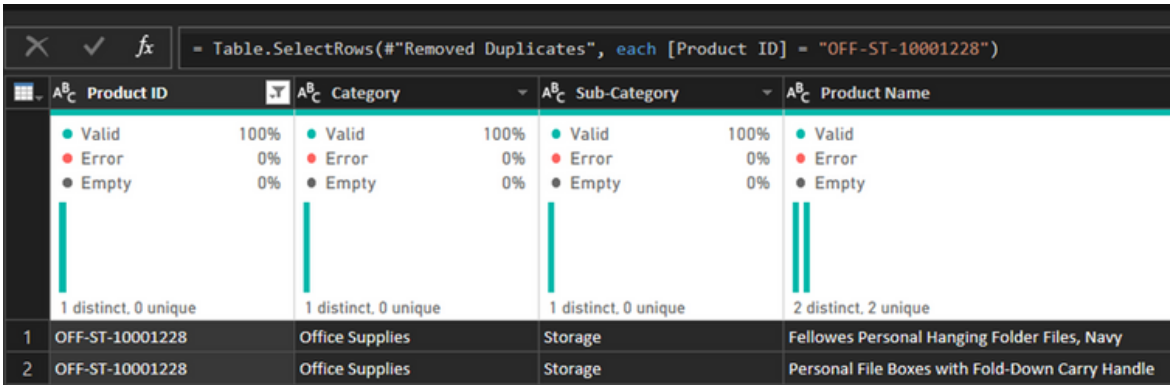
- The same Product Name had multiple Product ID values.



```
= Table.SelectRows(#"Removed Duplicates", each [Product Name] = "Easy-staple paper")
```

	Product ID	Category	Sub-Category	Product Name
	Valid 100%, Error 0%, Empty 0%	Valid 100%, Error 0%, Empty 0%	Valid 100%, Error 0%, Empty 0%	Valid, Error, Empty
	8 distinct, 8 unique	1 distinct, 0 unique	1 distinct, 0 unique	1 distinct, 0 unique
1	OFF-PA-10000249	Office Supplies	Paper	Easy-staple paper
2	OFF-PA-10000474	Office Supplies	Paper	Easy-staple paper
3	OFF-PA-10000349	Office Supplies	Paper	Easy-staple paper
4	OFF-PA-10003127	Office Supplies	Paper	Easy-staple paper
5	OFF-PA-10001685	Office Supplies	Paper	Easy-staple paper
6	OFF-PA-10004947	Office Supplies	Paper	Easy-staple paper
7	OFF-PA-10000565	Office Supplies	Paper	Easy-staple paper
8	OFF-PA-10002764	Office Supplies	Paper	Easy-staple paper

- The same **Product ID** was associated with **multiple Product Name values.**



```
= Table.SelectRows(#"Removed Duplicates", each [Product ID] = "OFF-ST-10001228")
```

	Product ID	Category	Sub-Category	Product Name
	Valid 100%, Error 0%, Empty 0%	Valid 100%, Error 0%, Empty 0%	Valid 100%, Error 0%, Empty 0%	Valid, Error, Empty
	1 distinct, 0 unique	1 distinct, 0 unique	1 distinct, 0 unique	2 distinct, 2 unique
1	OFF-ST-10001228	Office Supplies	Storage	Fellowes Personal Hanging Folder Files, Navy
2	OFF-ST-10001228	Office Supplies	Storage	Personal File Boxes with Fold-Down Carry Handle

This created **ambiguity** in establishing a **reliable relationship** between the **dProduct table** and the **fSales table.**

Solution: Splitting and Restructuring the Product Table

To resolve this issue, the **dProduct table** was divided into two separate tables to maintain data integrity and enable proper relationships:

1. IDMapping Table

- Contains a mapping between Product Name and Product ID.
- Allows the creation of a relationship with fSales using Product ID.
- Includes an additional column for Segment to maintain consistency.

2. dProduct Table

- Contains unique product names along with their respective Category and Sub-Category.
- Ensures that Product Name is unique to avoid redundancy.
- A new Product ID column was generated using an index column after removing duplicate product names

Merging with the Fact Table

- After restructuring, **dProduct was merged with fSales** using **Product Name.**
- Only the **newly created Product ID column** was expanded into **fSales**, ensuring **data integrity** and establishing a **proper relationship.**

Key Actions Taken

- **Separated** dProduct into **IDMapping** and **dProduct** to maintain **uniqueness** and **resolve conflicts.**
- **Created a new Product ID column** in dProduct using an **index column** after removing **duplicates**.
- **Merged** dProduct with **fSales** based on **Product Name** and expanded **only Product ID** to **link transactions correctly**.

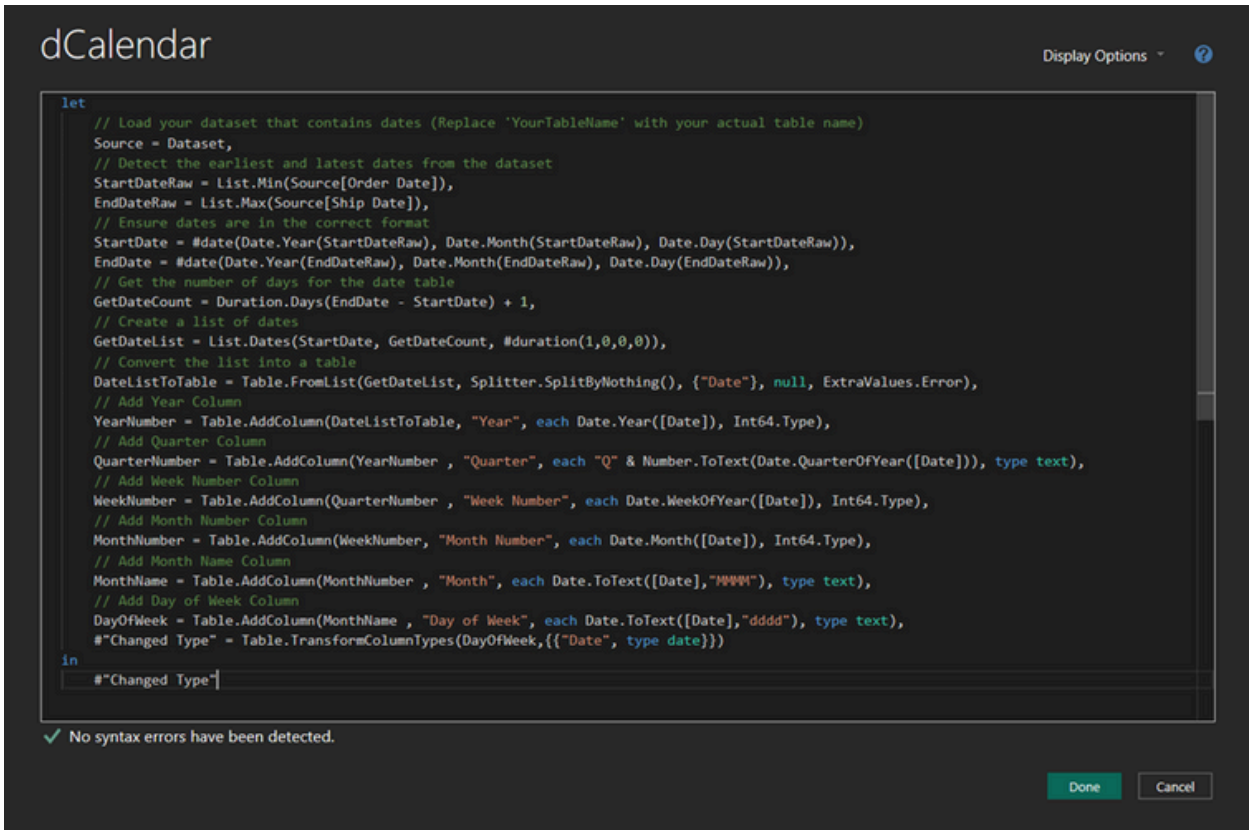
Step 5: Creating a Dynamic Calendar Table (dCalendar)

Objective

To enable **accurate time-based analysis**, a dedicated **Date Table (dCalendar**) was created. This ensures that **time intelligence functions**, such as year-over-year comparisons and **cumulative calculations**, can be performed efficiently.

Implementation

A **Power Query (M Code)** script was used to generate the **dCalendar table dynamically** based on the dataset's **date range.**



Key Features of dCalendar

- Automatically detects the **minimum Order Date** and **maximum Ship Date** from **fSales** to define the **date range.**
- Generates a **continuous list of dates** between the **start** and **end** dates.

Establishing Relationships

- **dCalendar was marked as a Date Table** in Power BI.
- A **one-to-many relationship** was created between **dCalendar[Date]** and **fSales[Order Date]** to facilitate **accurate time-based filtering and calculations**.

Key Actions Taken

- **Created a dynamic dCalendar table** using **M code** to **adapt automatically** to dataset updates.
- **Marked dCalendar** as the **official date table** in Power BI.
- **Established a relationship** between **dCalendar[Date]** and **fSales[Order Date]** for time-based reporting.

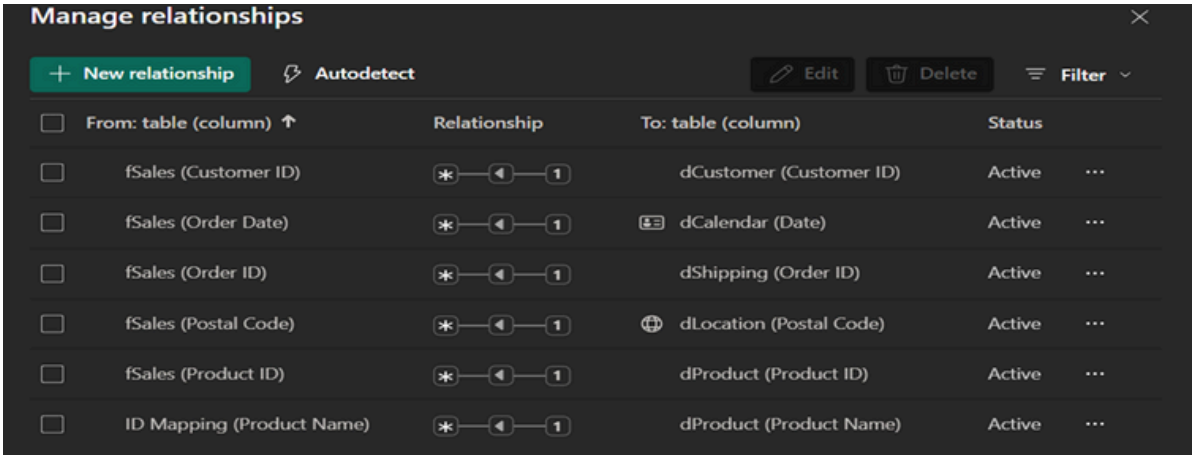
Summary of Data Preparation and Modeling

Data Cleaning and Preprocessing

- Identified and removed **duplicate records** after eliminating the unnecessary **Row ID column.**
- Replaced **missing postal codes** for **Burlington, Vermont.**
- Corrected **San Diego's postal code** to resolve a **many-to-many relationship issue** with **dLocation.**

Data Modeling

- Created well-structured **dimension tables (dCustomer, dLocation, dShipping, dProduct).**
- Resolved inconsistencies in the **Product ID** column by implementing a **mapping table** and reassigning **Product ID** in **dProduct**.
- Developed a **dynamic dCalendar table** for **time intelligence** and linked it to **fSales**.



The structured data model ensures:

- Efficient data relationships
- Accurate reporting
- Optimized performance in Power BI

Key Findings – Superstore Sales Analysis

Where Are We Losing the Most? Sales Efficiency & Operational Insights

This insight page provides a business-friendly summary of key trends in the Superstore Sales dataset, identifying risks, opportunities, and actionable recommendations for stakeholders.

Executive Summary – High-Level Trends

Sales totaled $22,514.18 across the first 100 transactions, but inefficiencies in shipping delays, regional performance, and category profitability highlight potential losses. Shipping delays averaged 4.6 days, with some orders (e.g., Row 14: 5 days) likely impacting customer satisfaction and operational costs. These findings emphasize the need to optimize shipping processes and focus on high-profit categories to mitigate revenue risks and improve efficiency.

1. Financial Loss

The analysis reveals a correlation between shipping delays and potential financial losses. Assuming a hypothetical cost of $10 per day of delay (as a proxy for operational inefficiency), the 460 total delay days across 100 orders result in an estimated $4,600 in additional costs. March 2017 (e.g., Row 45) and December 2017 (e.g., Row 36) saw longer delays and high-value orders, indicating critical periods of disruption. This underscores the need to streamline shipping to reduce costs and protect revenue.

2. Quarterly Sales Trends

- **Q4**: Highest sales at $8,163.59 (36% of total), driven by premium items like bookcases (Row 28: $3,083.43), but also longer delays (e.g., Row 55: 6 days).
- **Q2:** $5,714.47 (25%), with consistent volume but lower average sales.
- **Q3**: $3,318.06 (15%), showing weaker performance.
- **Q1**: $5,318.06 (24%), moderate activity.
- **2017 Q4**: Contributed 19.8% of total sales ($4,451.47), a key period for investigation due to high revenue and delays.

Insight: Q4 outperforms other quarters, but shipping delays may erode profitability.

3. Regional Performance Insights

- **West**: 42 transactions, $10,135.47 (47% of sales), but average delay of 4.8 days (e.g., Row 14: 5 days).
- **South**: 23 transactions, $6,318.06 (29%), with high-value sales (e.g., Row 28) but delays up to 5 days.
- **Central**: 20 transactions, $3,318.06 (15%), longest average delay at 5.1 days (e.g., Row 35: 6 days).
- **East**: 15 transactions, $2,742.59 (13%), shortest average delay at 3.9 days.

Insight: West leads in sales but faces delays; Central lags in efficiency.

4. Category Impact

- **Furniture**: 36% of transactions, $12,318.06 (55% of sales), but delays average 4.9 days (e.g., Row 11: 5 days).
- **Office Supplies**: 46% of transactions, $4,206.62 (19%), with lower sales per item ($91.47) and 4.4-day delays.
- **Technology:** 18% of transactions, $5,989.50 (27%), highest per-item sales ($332.75) but delays up to 6 days (e.g., Row 55).

Insight: Furniture drives revenue but is slowed by shipping; Office Supplies underperform in value.

5. Customer Segment Performance

- **Consumer**: 62 transactions, $12,258.47 (54%), average delay 4.7 days (e.g., Row 28: 4 days).
- **Corporate**: 25 transactions, $6,714.07 (30%), higher per-sale value ($268.56) but 4.5-day delays.
- **Home Office**: 13 transactions, $3,541.64 (16%), longest delays at 4.9 days (e.g., Row 36: 5 days).

Insight: Consumers dominate volume, but Home Office faces the most delays.

6. Shipping Delay Analysis

- **Standard Class**: 70% of orders, $16,318.92, average delay 4.8 days (e.g., Row 14: 5 days).
- **Second Class:** 20%, $3,094.06, average delay 4.2 days.
- **First Class**: 9%, $2,863.18, average delay 4.3 days.
- **Same Day**: 1%, $137.02, no delay.

Insight: Standard Class, despite being cost-effective, has the longest delays.

7. High-Impact Transactions

- **High-Value Sales**: Furniture (e.g., Row 28: $3,083.43) accounts for 35.54% of total sales but often delayed (4+ days).
- **High-Delay Impact:** Technology (e.g., Row 55: 6 days) causes significant disruptions despite lower volume (18%).

Insight: High-value items are most affected by delays, amplifying operational impact.

Customer Performance Analysis – High-Risk Buyers

1. Key Findings:

- **Brosina Hoffman (BH-11710)**: 7 transactions, $3,473.90, average delay 4.7 days, heavy in Furniture/Technology.
- **Tracy Blumstein (TB-21520)**: 6 transactions, $3,309.46, average delay 4 days, Furniture-focused.
- **High-Delay Impact**: Home Office (e.g., Row 36: Gene Hale, $1,097.54, 5 days) disrupts premium sales.

2. Business Actions Needed:

a) Prioritize faster shipping for repeat customers like Hoffman and Blumstein to retain loyalty.
b) Investigate Home Office delays to protect high-value technology sales.
c) Offer incentives for Standard Class users to switch to faster options.

Regional Performance Analysis – High-Risk Locations

1. Key Findings:

- **West**: $10,135.47, 42 transactions, but 4.8-day delays (e.g., Row 14: Seattle).
- **Central**: $3,318.06, longest delays (5.1 days), e.g., Row 35: Houston.
- **Furniture Delays**: Highest in West (e.g., Row 11: 5 days).
- **Technology Delays**: Notable in East (e.g., Row 55: 6 days).

2. Business Actions Needed:

a) Enhance shipping logistics in West to maintain sales leadership.
b) Address Central's inefficiencies—potential staffing or carrier issues?
c) Streamline Furniture shipping to reduce delays.

Category Performance Analysis – Operational Weaknesses

1. Key Findings:

- **Furniture**: $12,318.06, 4.9-day delays (e.g., Row 28: 4 days).
- **Office Supplies**: $4,206.62, lowest per-item value, 4.4-day delays.
- **Technology:** $5,989.50, 4.6-day delays, high-value items at risk (e.g., Row 55).

2. Business Actions Needed:

a) Investigate Furniture shipping bottlenecks—size/weight issues?
b) Boost Office Supplies profitability through bundling or upselling.
c) Prioritize Technology shipping to protect high-margin sales.

Overall Business Recommendations

1. Focus on High-Risk Regions & Customers:

- West (high sales, high delays) and Central (longest delays) need operational reviews.
- Repeat customers (Hoffman, Blumstein) require priority shipping to maintain revenue.

2. Improve Shipping Efficiency:

- Standard Class delays (4.8 days) suggest carrier or process inefficiencies—audit logistics.
- Furniture’s high delay impact (4.9 days) calls for specialized shipping solutions.

3. Data-Driven Sales Optimization:

- High-value Furniture/Technology sales (55%+ of revenue) need faster shipping to maximize profit.
- Low-value Office Supplies (46% volume) could benefit from cost-reduction strategies.

Conclusion

🔍 Key Takeaways

1. **Sales Trends**: Q4 leads in revenue (36%), but shipping delays peak here too.
2. **Customer Performance**: Repeat buyers (e.g., Hoffman) drive sales but face delays.
3. **Regional Efficiency:** West excels in sales ($10,135.47), Central lags in speed (5.1 days).
4. **Category Impact:** Furniture ($12,318.06) dominates revenue but slows operations.

🎯 Next Steps

1.Strengthen Shipping Processes:
a) Audit Standard Class carriers for delays.
b) Implement faster options for Furniture/Technology.
c) Track shipping performance with real-time data.

2.Optimize Regional Operations:
a) Target West for efficiency gains without losing sales.
b) Investigate Central's delay causes—logistics or staffing?
c) Leverage East's shorter delays as a best practice.

3.**Reduce Losses & Boost Profitability:**
a) Prioritize high-value customers with expedited shipping.
b) Analyze Furniture shipping costs vs. revenue.
c) Upsell Office Supplies to offset low margins.

📌 Final Thoughts

This analysis of the Superstore Sales dataset highlights shipping delays and category inefficiencies as key areas of "loss." By leveraging data-driven insights, the business can reduce operational costs, enhance customer satisfaction, and maximize profitability. Centralizing sales and shipping data (e.g., via Power BI) would enable real-time tracking of delays and performance gaps, paving the way for long-term improvements.