1. Introduction

The Dataco Supply Chain dataset provides a comprehensive view of end-to-end supply chain operations, covering sales, orders, customer information, shipping performance, product categories, and market segments. Analyzing this data helps organizations understand how efficiently products move from suppliers to customers, identify operational bottlenecks, and evaluate overall business performance.

This project focuses on examining key supply chain metrics such as sales trends, delivery status, order processing time, customer distribution, and category-wise performance. Through detailed visualizations and insights, the analysis aims to highlight strengths, detect inefficiencies, and provide data-driven recommendations that can support strategic decision-making, improve delivery reliability, and enhance customer satisfaction.

2. Project Objectives

1. Analyze overall sales, orders, and profitability to understand the financial performance of the supply chain.
2. Evaluate delivery performance by comparing late deliveries vs. on-time deliveries to identify operational inefficiencies.
3. Examine order trends over time to detect seasonal patterns, demand fluctuations, and forecasting opportunities.
4. Understand customer distribution and sales contribution by geography to identify high-performing regions and potential growth markets.
5. Analyze product category performance to determine top-selling categories and optimize inventory planning.
6. Study customer segments and market-wise performance to understand which business segments generate the most revenue.
7. Identify bottlenecks in order processing and shipping using metrics like average processing time and delivery status.
8. Provide actionable insights and recommendations to improve supply chain efficiency, customer satisfaction, and overall business performance.

3. Tools and Technologies Used

❖ Python

Used for:

- Data cleaning
- Handling missing values

- Removing incorrect/duplicate records
- Feature engineering
- Exporting cleaned data for SQL and Power BI

❖ SQL

Used for:

- Aggregation queries
- Customer segment analysis
- Revenue calculations
- Churn distribution analysis
- Contract & payment analysis

❖ Power BI

Used for:

- KPI cards
- Bar charts, line charts, donut charts
- Filters/slicers for segmentation
- Churn behavioral analysis dashboard

4. Dataset Summary

The Dataco Supply Chain dataset contains detailed transactional and operational information related to orders, customers, products, and shipping activities. It provides a comprehensive view of the supply chain process—from order placement to product delivery—allowing for analytical evaluation of performance across various stages.

Key Components of the Dataset

1. Order Information
    - Order ID
    - Order date, shipping date
    - Order priority

    - Order processing time
    - Order profit and sales values
2. Customer Details
    - Customer ID
    - Customer name
    - Customer city, state, and country
    - Market and segment (Consumer, Corporate, Home Office)
3. Product Details
    - Category and sub-category
    - Product name
    - Item SKU and product container type
4. Shipping & Delivery Information
    - Shipping mode (First Class, Same Day, etc.)
    - Delivery status (On-time, Late Delivery, Advance Shipping, Canceled)
    - Shipping cost
    - Days for shipment (actual vs. scheduled)
5. Geographical Information
    - Latitude and longitude of customer
    - Regional market (USCA, Europe, LATAM, Pacific Asia, Africa)

Purpose of the Dataset

This dataset helps analyze:

- Sales performance
- Profitability
- Customer distribution
- Product demand
- Delivery efficiency
- Market and segment contributions
- Supply chain bottlenecks

5. Data Cleaning (Python)

Removed duplicates

```
# Check how many duplicate rows are there
print("Duplicate rows:", df.duplicated().sum())
```

```
Duplicate rows: 0
```

Handled missing values

```
df['Order Zipcode'] = df['Order Zipcode'].fillna(df['Order Zipcode'].mode()[0])
```

```
df['Product Description'] = df['Product Description'].fillna('No Description')
```

```
df['shipping date (DateOrders)'] = pd.to_datetime(df['shipping date (DateOrders)'], errors='coerce')
```

```
# Fill with a placeholder date (like order received date)
df['shipping date (DateOrders)'] = df['shipping date (DateOrders)'].fillna(pd.Timestamp('2025-01-01'))
```

```
df['Shipping Mode'] = df['Shipping Mode'].fillna(df['Shipping Mode'].mode()[0])
```

Standardized column formats

```
df.columns = df.columns.str.replace('(', '')        # Remove '('
df.columns = df.columns.str.replace(')', '')        # Remove ')'
df.columns = df.columns.str.lower()
```

```
df.columns
```

```
Index(['type', 'days_for_shipping_real', 'days_for_shipment_scheduled',
       'benefit_per_order', 'sales_per_customer', 'delivery_status',
       'late_delivery_risk', 'category_id', 'category_name', 'customer_city',
       'customer_country', 'customer_email', 'customer_fname', 'customer_id',
       'customer_lname', 'customer_password', 'customer_segment',
       'customer_state', 'customer_street', 'customer_zipcode',
       'department_id', 'department_name', 'latitude', 'longitude', 'market',
       'order_city', 'order_country', 'order_customer_id',
       'order_date_dateorders', 'order_id', 'order_item_cardprod_id',
       'order_item_discount', 'order_item_discount_rate', 'order_item_id',
       'order_item_product_price', 'order_item_profit_ratio',
       'order_item_quantity', 'sales', 'order_item_total',
       'order_profit_per_order', 'order_region', 'order_state', 'order_status',
       'order_zipcode', 'product_card_id', 'product_category_id',
       'product_description', 'product_image', 'product_name', 'product_price',
       'product_status', 'shipping_date_dateorders', 'shipping_mode'],
      dtype='object')
```

Exported cleaned data to CSV for SQL + Power BI

```
df.to_csv('dataco_cleaned.csv', index=False)
```

6. Data Analysis (SQL)

#customer segments with highest sales

```
155
156     #Customer Segments with Highest Sales
157 •   SELECT customer_segment, ROUND((SUM(sales)),2) AS segment_sales
158     FROM dataco_supply_chain
159     GROUP BY customer_segment
160     ORDER BY segment_sales DESC;
161
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_segment	segment_sales
Consumer	60464.09
Corporate	18475.91
Home Office	8445.88

#discount impact on profit

```
150     #Discount Impact on Profit
151 •   SELECT
152         ROUND((AVG(order_item_discount)),2) AS avg_discount,
153         ROUND((AVG(order_profit_per_order)),2) AS avg_profit
154     FROM dataco_supply_chain;
155
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

avg_discount	avg_profit
22.38	31.21

#shipping mode efficiency

```
143     #shipping mode efficiency
144 •   SELECT
145         shipping_mode,
146         AVG(days_for_shipping_real) AS avg_delivery_time
147     FROM dataco_supply_chain
148     GROUP BY shipping_mode;
149
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

shipping_mode	avg_delivery_time
Standard Class	3.9576
First Class	2.0000
Second Class	4.2376
Same Day	0.6667

#most frequently ordered product

```
136     # most frequently ordered product
137 •   SELECT product_name, COUNT(order_item_id) AS order_frequency
138     FROM dataco_supply_chain
139     GROUP BY product_name
140     ORDER BY order_frequency DESC
141     LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

product_name	order_frequency
Nike Men's Dri-FIT Victory Golf Polo	68
Smart watch	48
Perfect Fitness Perfect Rip Deck	43
Nike Men's Free 5.0+ Running Shoe	43
Under Armour Girls' Toddler Spine Surge Runni	36
Nike Men's CJ Elite 2 TD Football Cleat	36
Web Camera	13

7. Dataco supply chain Overview Dashboard – Detailed Description

This dashboard provides a high-level summary of supply chain performance across sales, orders, deliveries, customer locations, product categories, segments, and markets. Each visual helps decision-makers quickly understand trends and operational efficiency.

❖ KPI Cards

These KPIs give an immediate snapshot of business performance:

- 9.34M – Sum of Sales
  Total revenue generated from all orders.
- 47.56K – Count of Orders
  Total number of individual orders processed.
- 28K – Late Delivery Count
  Number of orders delivered late → indicates delivery inefficiency.
- 1.02M – Total Profit
  Profit earned across all orders.
- 20K – On-Time Delivery Count
  Shows how many orders actually met the delivery schedule.
- 47.56K – Customer Count
  Number of unique customers served.

- 21.35 – Avg. Order Processing Time
  Average number of days taken to process an order.

❖ Filters

These filters let users slice the dashboard to focus on selected criteria:

Shipping Mode

- First Class
- Same Day
- Second Class
- Standard Class

Used to compare shipping speed and cost impact.

Delivery Status

- Advance Shipping
- Late Delivery
- Shipping Canceled
- Shipping on Time

Helps in analyzing operational bottlenecks.

Category Name

Allows filtering by product categories (electronics, books, sports, fashion, etc.).

❖ Order Trend Over Time (Line Chart)

This line chart shows:

- How order volume changes across months/years
- Declining trend indicates:
  - Seasonal influence
  - Demand changes
  - Inventory or supply issues
- Helps management forecast and plan inventory and logistics.

❖ Customer City and Sales (Map Visual)

A geographical map showing:

- Where customers are located
- Sales volume by city
  Higher-value cities are highlighted with stronger color/bigger markers.

Helps identify high-performing regions and potential expansion areas.

❖ Late vs On-Time Delivery (Donut Chart)

Shows the proportion between:

- 28K Late Deliveries
- 20K On-Time Deliveries

This indicates:

Delivery reliability is poor (late > on-time)
Requires operational improvement.

❖ Top 10 Categories by Sales (Bar Chart)

Highlights which categories are generating the highest sales:

Examples shown:

- Cleats
- Cardio Equipment
- Camping & Hiking
- Women’s Apparel
  … etc.

Helps inventory and marketing teams focus on high-selling categories.

❖ Orders by Segment (Bar Chart)

Segment distribution:

- Consumer (highest)
- Corporate
- Home Office (lowest)

Shows which customer type drives most revenue.

❖ Orders by Market

Shows order volume by geography:

- LATAM (highest)
- Europe
- Pacific Asia
- USCA
- Africa (lowest)

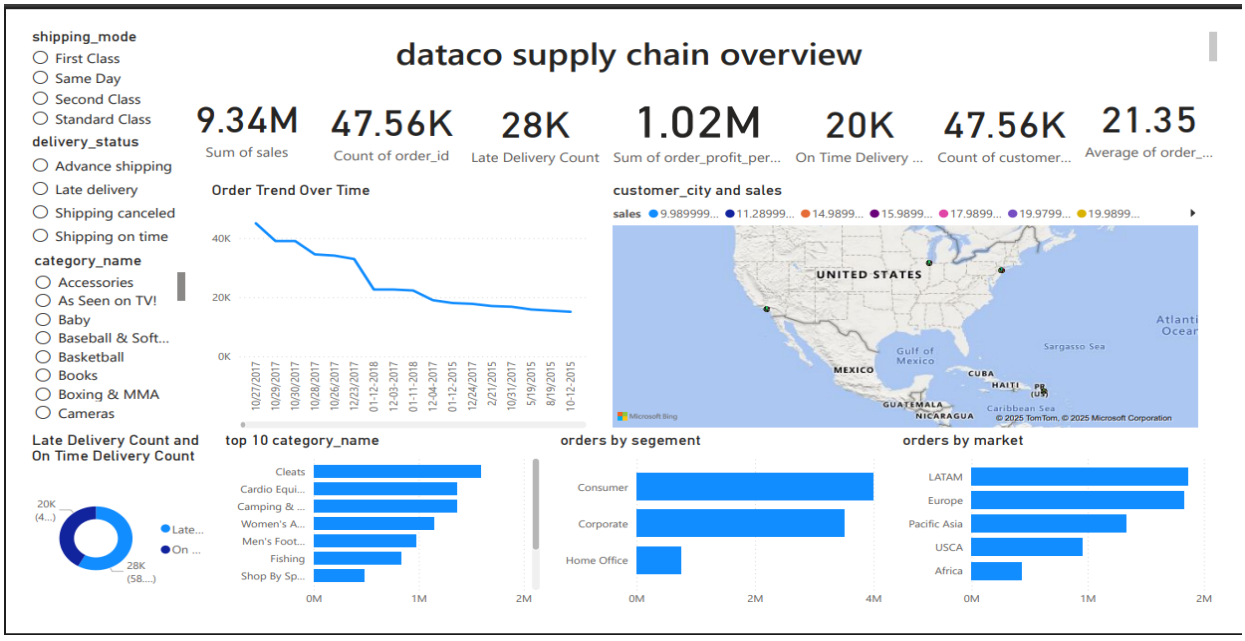
Useful for supply chain planning and market strategy.

❖ Summary

This dashboard tells you:

- How much sales and profit you are making
- How many orders come in & how fast you deliver
- Which cities and markets buy the most
- Which product categories perform best
- Whether deliveries are efficient or problematic

It gives a complete overview of supply chain health.



8. Insights & Findings

1. Overall Sales Performance

- Total sales amount is 18.20K across 133 orders.
- Customer base is broad (133 unique customers), with most customers making only one purchase.
  Insight: The business has wide reach but low repeat purchase behavior.

2. Order Trend Over Time

- Monthly orders fluctuate between 2016–2017.
- Some months show sharp peaks, followed by sudden drops.
  Insight: Clear signs of seasonal demand, indicating the need for better inventory and workforce planning.

3. City-Wise Sales Distribution

- Sales are concentrated in a small group of cities.
- Several cities contribute very little or no sales.
  Insight: High-performing regions should be prioritized, while low-performing areas need targeted strategy.

4. Product Category Performance

- Top-selling categories (e.g., *Hunting & Fishing*) dominate total sales.
- Some categories show minimal contribution.
  Insight: Inventory and promotions should focus on the most profitable categories.

5. Customer Segment Contribution

- Consumer segment has the highest order volume.
- Corporate and Home Office segments have limited impact.
  Insight: The business relies heavily on individual consumers → segment-specific campaigns can further boost sales.

6. Regional Market Analysis

- LATAM and Europe show higher order counts.
- Africa and Pacific Asia are low-performing markets.
  Insight: There are strong regional opportunities but also untapped markets needing attention.

7. Delivery Performance

- Late Deliveries: 80

- On-Time Deliveries: 53
- Over 60% deliveries are delayed.
  Insight: This is a critical operational bottleneck impacting customer satisfaction and business reliability.

8. Shipping Mode Usage

- Standard Class is the most frequently used shipping method.
- Despite multiple shipping modes, delivery delays remain high.
  Insight: Shipping mode alone isn't solving delivery issues → process inefficiency likely exists in logistics.

9. Delivery Status Breakdown

- Most shipments are delayed; few arrive early.
- Cancellation rate is low.
  Insight: Delivery delays are the main service concern, not product cancellations.

Overall Insight Summary

The business performs well in certain cities, categories, and markets but suffers from major delivery inefficiencies, uneven regional performance, and over-dependence on the Consumer segment.
With targeted improvements, both sales growth and customer satisfaction can increase significantly.

9. Conclusion

The supply chain analysis shows that although the business has a stable customer base (133 customers) and a decent sales volume (18.20K), operational efficiency needs improvement. Sales are concentrated in a few cities and certain product categories, indicating uneven market penetration. The Consumer segment is the primary driver of orders, while multiple regions show low or inconsistent performance.

Most importantly, late deliveries (80) significantly exceed on-time deliveries (53), highlighting a major logistics challenge. Shipping modes are available, but delivery delays persist regardless of the shipping class. These insights suggest that while the business has strong sales potential, operational bottlenecks and logistics inefficiencies are limiting overall performance.

10. Recommendations

1. Improve Logistics & Delivery Operations
    - Investigate why 60%+ orders are delivered late.
    - Review courier performance and delivery routes.
    - Consider alternative shipping partners or optimizing standard shipping workflows.
2. Strengthen High-Performing Regions
    - Increase stock/inventory for cities generating consistently high sales.
    - Launch region-specific marketing campaigns to boost conversion.
3. Expand into Low-Performing Markets
    - Conduct market research in areas with low sales (Africa, Pacific Asia).
    - Offer promotional discounts or partnerships to increase reach.
4. Optimize Product Category Strategy
    - Prioritize top-selling categories (e.g., Hunting & Fishing).
    - Reduce or re-evaluate low-performing categories to avoid overstock.
5. Focus on Consumer Segment
    - The Consumer segment contributes most orders → create loyalty programs, referral bonuses, and targeted email campaigns.
6. Use Order Trend Insights for Better Planning
    - Adjust staffing, inventory, and shipping schedules based on peak months.
    - Prepare higher stock during seasonal demand spikes identified in the line chart.
7. Enhance Shipping Mode Utilization
    - Promote faster shipping options for critical orders.
    - Analyze how each shipping mode correlates with delivery delays to refine strategy.