FINAL PROJECT REPORT

Retail Sales Performance & Operational Efficiency Analysis

Project Title -

Sector - Retail / E-commerce Analytics

Team Members:
- Viraj Chafale
- Aryan patel
- Aaditya Rana
- Augustya Purohit
- Gogulamudi Jayadeep

Institute:
Newton School of Technology

Date:

---

2. Executive Summary

Problem
Retail management lacks consolidated visibility into regional sales performance, category contribution, customer segment behavior, and shipping efficiency. Strategic decisions regarding logistics optimization and revenue focus areas are therefore taken without a unified analytical foundation.

Approach
The team acted as Retail Analytics Consultants and analyzed a structured dataset of **9,800 transaction records** using Google Sheets. Data cleaning, standardization, and feature engineering were performed, followed by KPI framework development and pivot-based exploratory analysis. An interactive dashboard was created to provide decision-ready visibility across regions, categories, segments, and shipping performance.

Key Insights

- Total sales recorded: **$2,166,852.12**
- Technology category is the top revenue contributor (**$739,253**, ~34% share).
- West region leads geographically (**$657,738**, ~30% share).
- Consumer segment dominates revenue (**~$1.036M**, ~48% contribution).
- Average shipping time is **9.49 days**, indicating logistics optimization opportunity.

Key Recommendations

- Prioritize Technology category for revenue expansion.
- Optimize shipping operations to reduce delivery cycle.
- Strengthen presence in South and Central regions.
- Develop targeted strategies for Corporate and Home Office segments.

---

3. SECTOR & BUSINESS CONTEXT

Sector Overview

The retail and e-commerce sector is increasingly driven by data-centric decision making. With expanding product catalogs and geographically distributed customers, organizations rely on sales analytics to identify growth opportunities, optimize logistics, and improve customer satisfaction.

Current Challenges

- Uneven regional sales performance
- Logistics inefficiencies impacting delivery time
- Difficulty identifying high-value customer segments
- Lack of consolidated performance monitoring

Why This Problem Was Chosen

The provided retail dataset offers transaction-level visibility into customer behavior, product performance, and shipping operations. Analyzing this data enables identification of revenue concentration, operational gaps, and strategic growth opportunities relevant to real-world retail businesses.

---

4. PROBLEM STATEMENT & OBJECTIVES

Formal Problem Definition

> "Which regions, categories, and customer segments are the primary drivers of retail revenue, and where do logistics inefficiencies create opportunities for operational improvement?"

Project Scope

- Sales trend evaluation (2015–2018)
- Category and sub-category performance
- Regional revenue comparison
- Customer segment analysis
- Shipping efficiency analysis
- KPI framework and dashboard creation

Success Criteria

- Identification of top revenue drivers
- Detection of logistics inefficiencies
- Development of KPI framework
- Creation of interactive dashboard
- Actionable business recommendations

---

5. DATA DESCRIPTION

Dataset Source
Retail transaction dataset containing order-level sales information across the United States.

Data Structure

- Rows: **9,800**
- Columns: **18**
- Order Date Range: **02-Jan-2015 to 30-Dec-2018**
- Geography: **United States**

Key Columns

- Order ID — unique order identifier
- Order Date — order placement date
- Ship Date — shipping date
- Ship Mode — delivery method
- Segment — customer segment
- Category — product category
- Region — sales region
- Sales — revenue value

Data Limitations

- Missing values present in Customer Name and Region.
- Ship Date contained mixed formats prior to cleaning.
- No profit or cost data available.
- Dataset represents a historical snapshot, not real-time data.

---

6. DATA CLEANING & PREPARATION

(All primary steps executed in Google Sheets as per capstone requirement)

Missing Values Handling

- Customer Name nulls replaced with **"Unknown Customer"** to preserve transaction integrity.
- Region nulls marked as **"Unknown"** and reconciled where possible using City/State mapping.

Date Standardization

Ship Date contained mixed formats. All values were converted into a uniform date format to ensure accurate chronological sorting and shipping duration calculation.

Feature Engineering

The following derived columns were created to enhance analytical depth:

- **Order Year** — extracted from Order Date for year-wise analysis
- **Order Month** — extracted for seasonality detection
- **Shipping Days** — calculated as (Ship Date − Order Date) to measure fulfillment efficiency

Assumptions

- Sales values are assumed to be in a consistent single currency.
- Shipping duration calculated using calendar day difference.
- Missing customer identifiers do not affect revenue totals.

---

7. KPI & METRIC FRAMEWORK

KPI	Value	Formula	Why it Matters
Total Sales	$2,166,852.12	SUM(Sales)	Overall revenue health
Total Orders	4,821	COUNTUNIQUE(Order ID)	Demand volume
Average Order Value	$449.46	Sales / Orders	Customer spend quality
Avg Shipping Time	9.49 days	AVG(Shipping Days)	Logistics efficiency

Category Revenue

- Technology: **$739,253** (Highest)
- Furniture: **$659,817**
- Office Supplies: **$634,252**

Regional Revenue

- West: **$657,738** (Leader)
- East: **$610,064**
- Central: **$452,670**
- South: **$362,530** (Lowest)

EXPLORATORY DATA ANALYSIS

Trend Analysis

Sales between 2015–2018 show relatively stable transaction activity with periodic monthly fluctuations, indicating seasonal retail demand patterns. No strong long-term upward trend is observed, suggesting a mature and competitive market environment.

Category Performance

Technology contributes approximately **34% of total revenue**, clearly outperforming Furniture and Office Supplies. This indicates strong customer preference toward technology products and highlights it as the primary revenue engine.

Regional Analysis

The West region generates nearly **30% of total sales**, establishing it as the dominant geographic market. The South region significantly underperforms, indicating potential for targeted market expansion.

Segment Analysis

Revenue contribution by segment:

- Consumer: **$1.036M (~48%)**
- Corporate: **$633K**
- Home Office: **$386K**

The Consumer segment is the primary demand driver, while Corporate customers show strong purchasing capacity per order.

Shipping Analysis

- Average shipping time: **9.49 days**
- Most used mode: **Standard Class**

This indicates a cost-efficient but slower delivery strategy that may impact customer satisfaction.

---

9. ADVANCED ANALYSIS

Customer segmentation reveals that:

- Consumer segment dominates order volume and revenue.
- Corporate customers show strong purchasing power.
- Home Office segment remains under-penetrated.

This indicates opportunity for targeted B2B and small-business marketing strategies.

10. DASHBOARD DESIGN

Dashboard Objective

Provide a single-view monitoring system for sales, segments, regions, and logistics performance.

View Structure

- Top: KPI cards
- Middle: Sales trend & category comparison
- Bottom: Region and segment analysis

Filters Implemented

- Region slicer
- Category slicer
- Order Year slicer

---

11. INSIGHTS SUMMARY

- Technology category contributes the largest revenue share (~34%).
- West region is the primary revenue driver (~30%).
- Consumer segment generates nearly half of total sales (~48%).
- South region remains significantly underpenetrated.
- Average order value of **$449** indicates strong basket size.
- Standard Class dominates shipping usage, reflecting cost-focused logistics.
- Average shipping time (~9.5 days) presents customer experience risk.
- Revenue distribution shows moderate geographic concentration.
- Category mix indicates technology-led growth structure.
- Data required preprocessing before reliable analysis.

12. RECOMMENDATIONS

1. Double-down on Technology category
→ Impact: Immediate revenue lift

2. Optimize logistics to reduce 9.49-day average shipping
→ Impact: Better customer satisfaction

3. Expand in South and Central regions
→ Impact: Market penetration growth

4. Target Corporate & Home Office segments
→ Impact: Higher AOV potential

---

13. IMPACT ESTIMATION

If executed:

- Logistics improvement could reduce delivery time by **15–25%**
- Category focus could lift revenue by **8–12%**
- Regional expansion could unlock **$100K+ incremental sales**
- Segment targeting could improve AOV by **5–10%**

---

14. LIMITATIONS

- Missing values required imputation
- No profit/cost visibility
- Correlation does not imply causation
- External factors not included
- Historical snapshot only

---

15. FUTURE SCOPE

- Time-series forecasting for demand prediction
- Profitability and margin analysis
- Customer lifetime value modeling
- Inventory optimization analysis
- Machine learning–based demand forecasting
- Deeper ship-mode efficiency benchmarking
-

---

16 .CONCLUSION

This analysis demonstrates that retail performance is strongly influenced by category mix, regional presence, customer segmentation, and logistics efficiency. The developed KPI framework and interactive dashboard provide management with a unified decision-support system. By optimizing shipping performance and focusing on high-performing categories and regions, the organization can unlock meaningful revenue growth and operational efficiency improvements.

17 .APPENDIX

Column Name	Description
Row ID	Unique row identifier
Order ID	Unique identifier for each order
Order Date	Date when order was placed
Ship Date	Date when order was shipped
Ship Mode	Shipping method used
Customer ID	Unique customer identifier
Customer Name	Name of the customer
Segment	Customer segment (Consumer, Corporate, Home Office)
Country	Country of sale (United States)
City	Customer city
State	Customer state
Postal Code	Postal code
Region	Sales region
Product ID	Unique product identifier
Category	Product category
Sub-Category	Product sub-category
Product Name	Name of the product
Sales	Revenue generated from the order

18 .Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing
Viraj Chafale	✔️		Lead	Contributor	Secondary
Aryan Patel			✔️	✔️	✔️
Aaditya Rana		✔️	✔️	✔️	
Augustya Purohit					
Gogulamudi Jayadeep					