Supply Chain Shipping Data Analysis

Analyzing delivery timelines, delays, and performance factors across our global supply chain operations.

Title: E-Commerce Shipping Data Analysis
Author: Srinivas Senthil Kumar
Date: August 2025
Tools: Python, Pandas, Seaborn
Contact: srinivassenthilkumar082@gmail.com

Table of Contents

Introduction

- An International E commerce shipping company wants to discover key business insights and to use the advanced Machine learning techniques to study their customer database.
- I have attached the dataset (Train.csv) for your reference.

Objective

Our main objective is on identifying patterns in shipping delays and customer behavior to help logistics teams optimize delivery strategies and improve customer satisfaction.

Dataset Overview

Source of data: Kaggle

Key Columns:

- **Shipping Mode:** Road, Ship, or Flight
- **Discount Offered**
- **Delivery Status:** On time or Delayed
- **Product Importance & Priority**
- **Warehouse and Customer Location**
- **Customer Rating**

File: Train.csv

Rows: 10999 | Columns: 12

Exploratory Data Analysis (EDA)

OWL Ai

Key Questions Explored

1. What percentage of shipments are delayed?
2. Which shipping mode has the highest delay rate (Road vs Ship vs Flight)?
3. Do higher discounts lead to more delivery delays?
4. Do high-priority products get delayed more or less?
5. Which warehouses have the most delivery delays?
6. What is the average customer rating, and does it relate to delivery status?
7. Are customer queries being answered?
8. If product importance is high, does it get delivered on time or receive better ratings?

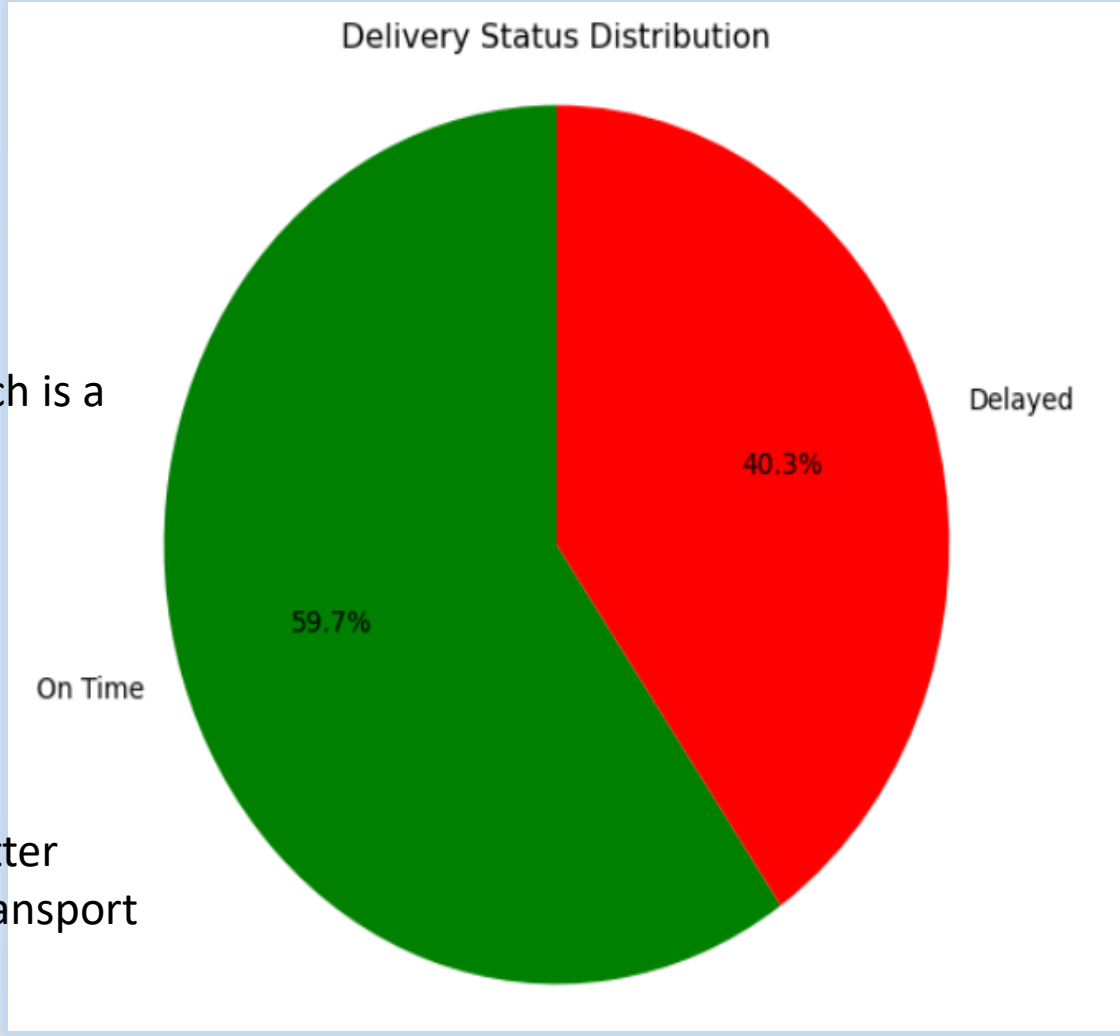
1. % On Time delivery vs delayed

Business Insights:

- Over 40% of orders are delayed — impacting customer satisfaction
- 4 out of 10 deliveries are late – which is a high impact
- Impacted customers and resulted in customer dissatisfaction

Recommendation :

- Improve delivery processes with better route planning or better mode of transport and real-time tracking systems.



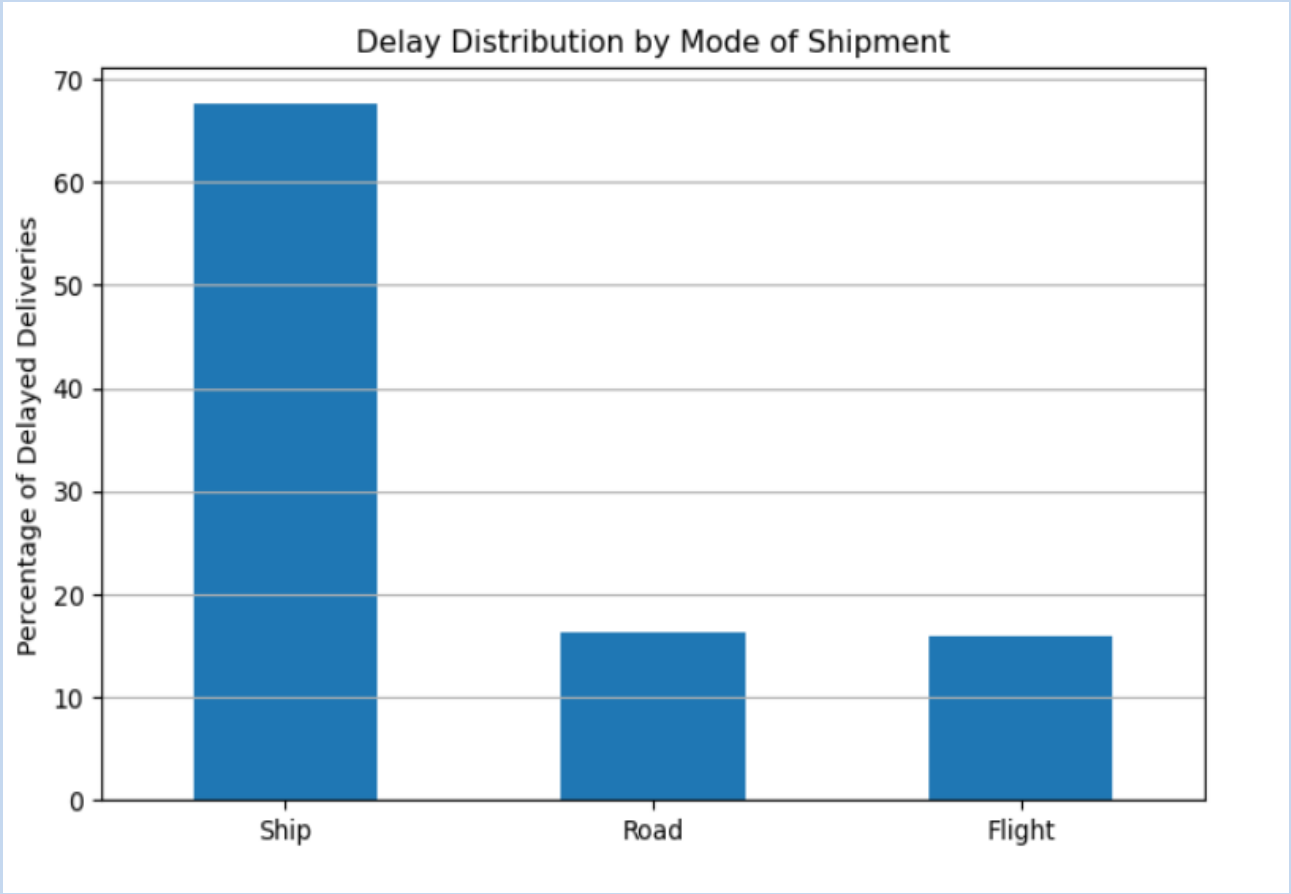
2. Over Two-Thirds of Cases

Business Insights:

- Ship mode accounts for ~67.7% of all shipments — and is likely responsible for most delays
- Ship is the slowest and least reliable – causes bottlenecks

Recommendation :

- Inspect partnerships with faster couriers (Air freight) for critical regions or products.



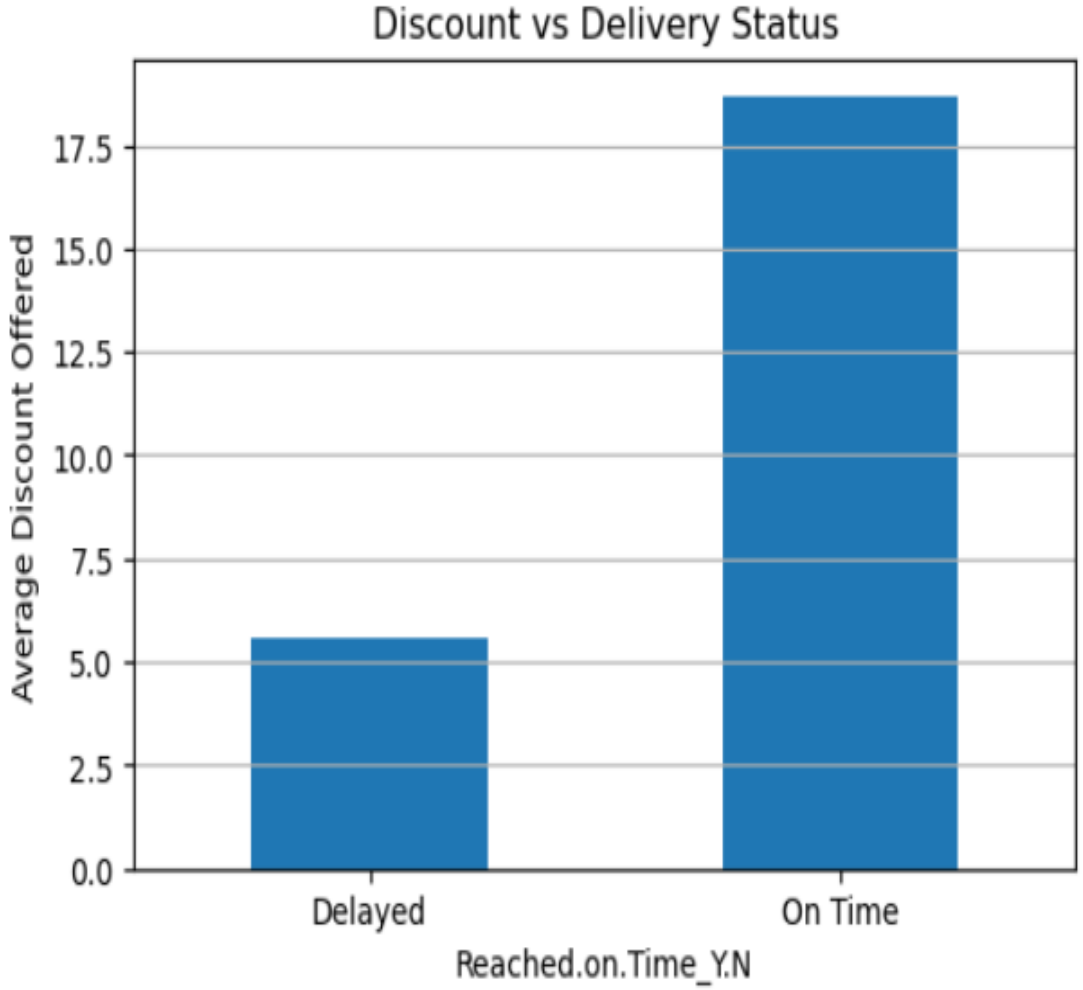
3. Discount Offered vs Delivery

Business Insights:

- Discounts don’t cause delays — in fact, on-time orders had bigger discounts.
- Higher discounts are often offered when products are delayed.
- Discounting may be used as a reactive compensation tactic.

Recommendation:

- Instead of relying on discounts, fix the root causes of delays.



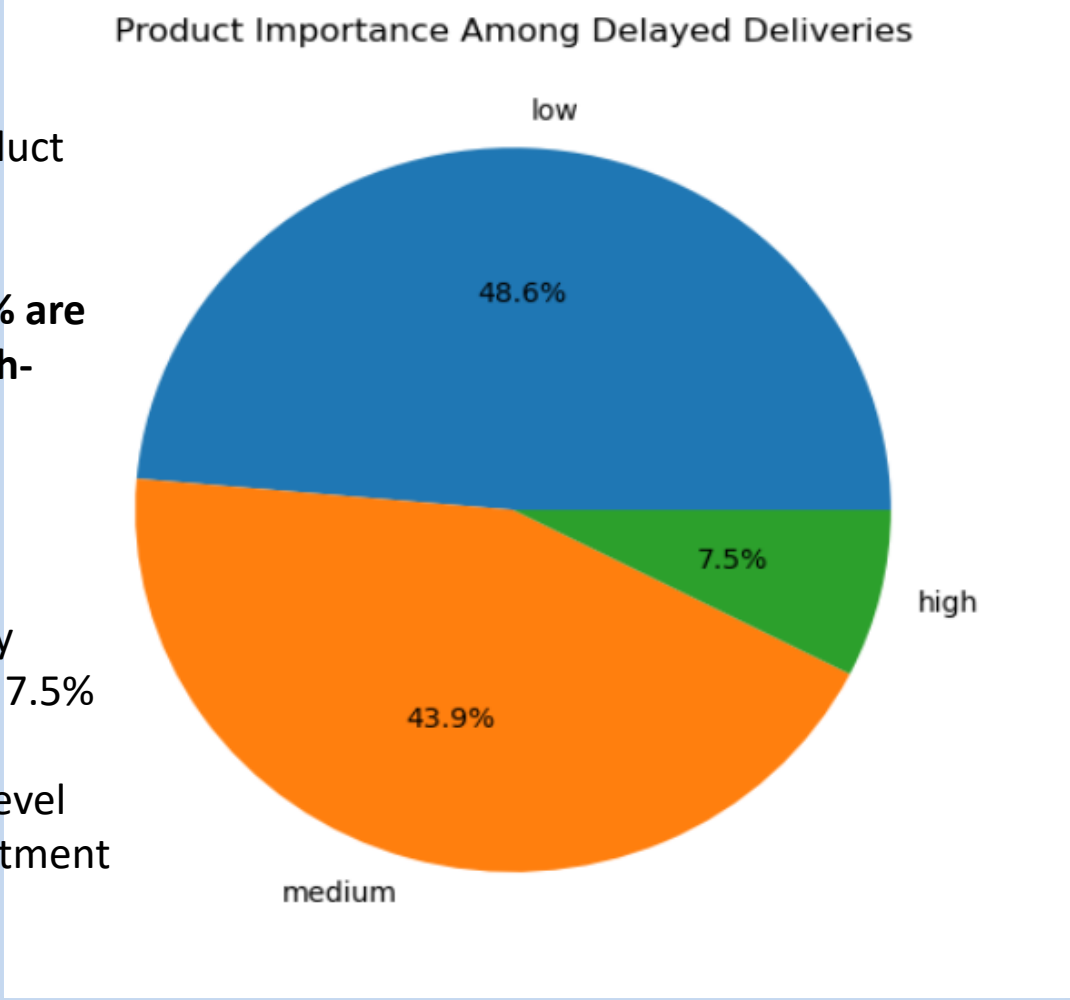
4. High Product Importance Doesn't Guarantee Faster Delivery

Business Insights:

- Delivery rates are similar across all product importance levels.
- Even among **delayed orders**, nearly **49% are low-importance**, and only **7.5% are high-importance**.

Recommendation:

- Shows some **priority handling** is already happening in the supply chain , but still 7.5% delays for high-importance items exist
- Implement priority queues or Service Level Agreement tiers — a formalized commitment
- Focus on service quality and delivery experience across all tiers.



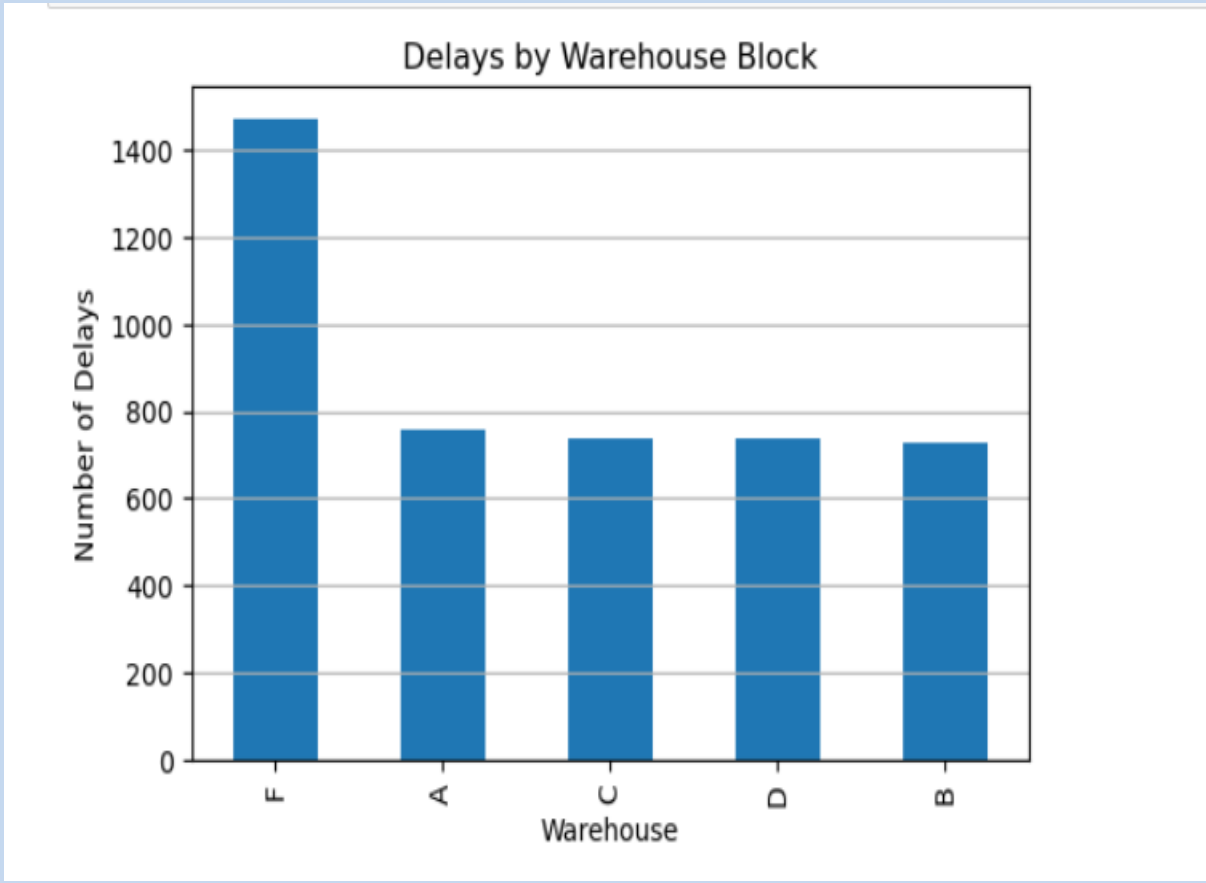
5.Warehouse F Has the Most Delays Operational Bottleneck Suspected

Business Insights:

- Warehouse **F** shows the **highest number of late deliveries**, followed by A and C.
- Operational issues or capacity overload in specific warehouses may lead to delays.

Recommendation:

- Audit performance at Warehouse F (staffing, layout, dispatch speed).
- Balance order volume across warehouses or reallocate resources.
- Assess staffing levels and equipment functionality
- Review order processing workflows and pick paths
- Evaluate carrier pickup schedules and dock utilization

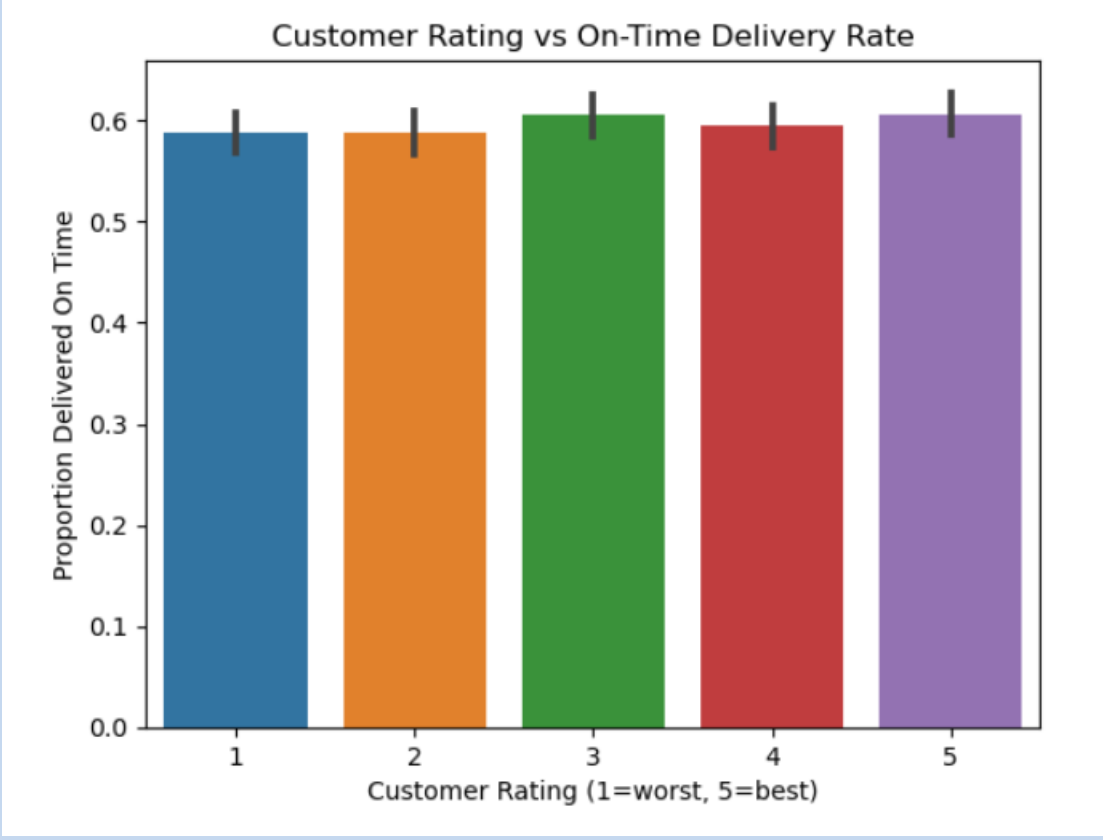


6. Proportion of on-time deliveries

Correlation Identified
Higher customer ratings (4-5 stars) correlate with on-time deliveries, while lower ratings (1-2 stars) show slightly fewer on-time deliveries.

Moderate Effect
While delivery timeliness influences satisfaction, the correlation is weaker than expected, suggested other factors also significantly impact customer ratings.

Strategic Implication
To improve overall satisfaction, delivery performance must be addressed alongside product quality, customer support, and pricing considerations.



7.Customer Care Calls vs Delivery Status

Key Observation

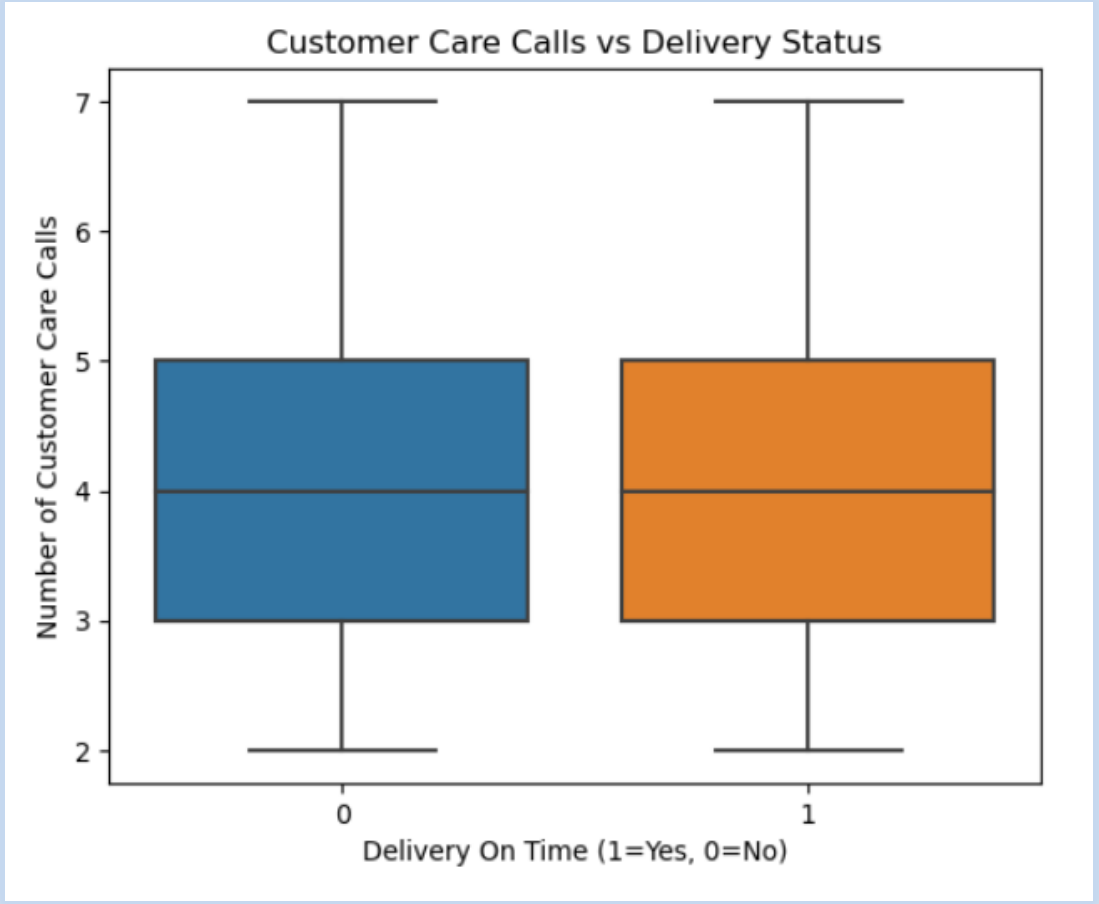
Delayed shipments generate a little wider range and higher average number of customer service calls compared to on-time deliveries.

Operational Impact

- Increased support team workload
- Higher customer service costs
- Potential for escalations and complaints
- Resource diversion from other service areas

Hidden cost of delays

Each percentage point improvement in on-time delivery can reduce customer service load and associated costs.



8. Product Importance vs On-Time Delivery

Unexpected Finding

- High-importance products do not show significantly better on-time delivery rates compared to low or medium importance items.

Operational Gap

- Current logistics processes treat all product categories similarly, regardless of stated importance or priority level.

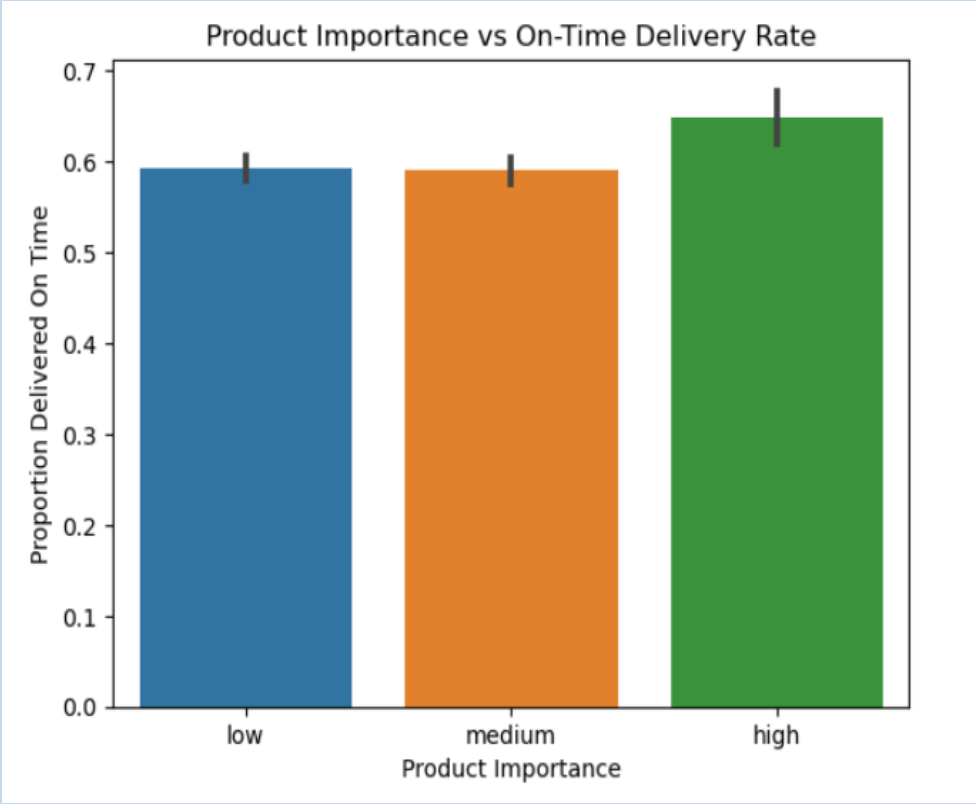
Strategic Recommendation

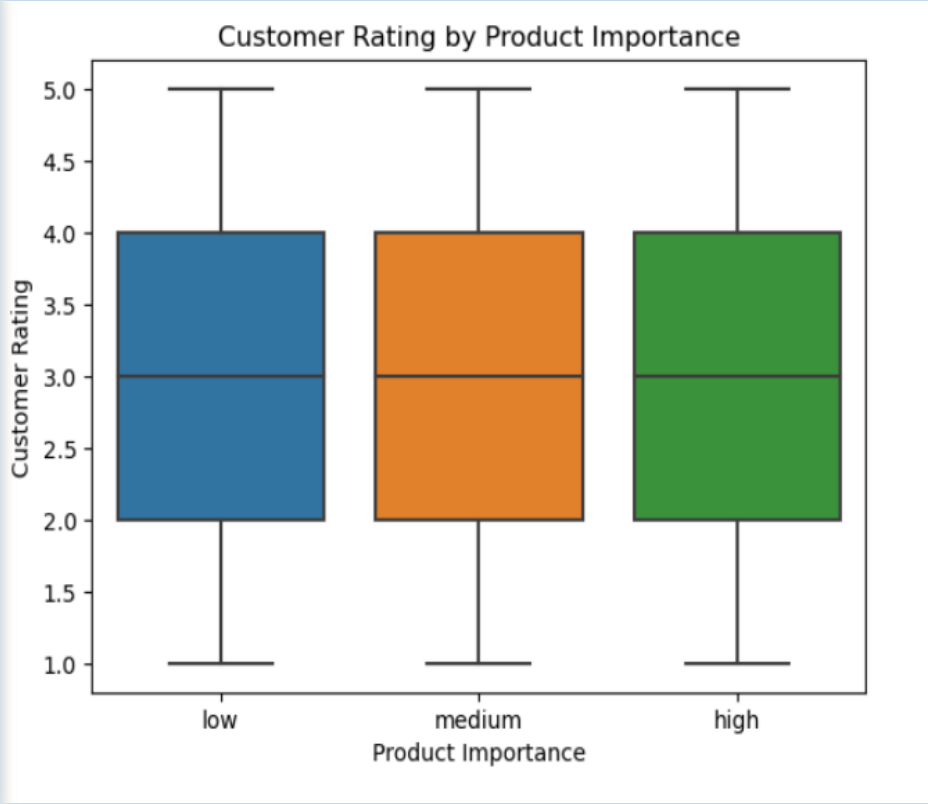
- Implement true prioritization in the supply chain to ensure high-importance products receive preferential handling and expedited shipping.
- This represents a significant opportunity to align operational execution with business priorities and customer expectations.

Conclusion:

If a company claims premium product handling, they may need to improve prioritization for **high-importance** items.

Product Importance vs Customer Rating





- Companies shouldn’t assume high-priority products yield higher customer satisfaction — they must still ensure quality and service for all items.

Key Performance Insights Summary

Insight Category	Key Finding	Business Implication
On-Time Delivery Rate	59.67% on time, 40.33% delayed	Performance is well below industry standard
Delay by Shipment Mode	Ship mode caused 67.7% of delays	Consider mode optimization for critical shipments
Product Importance & Delay	High-importance items only 7.5% of delays	Effective prioritization exists but can be improved
Discount & Delay	Delayed orders had lower discounts (5.55 vs 18.66)	Discounted orders receive better service (unexpected)
Delay by Warehouse	Warehouse F had most delays (1,472 orders)	Targeted improvement opportunity identified

Recommendations

- **Switch from Ship to Air/Road for Critical Routes:**
  Evaluate air freight partnerships for high-value or time-sensitive products.
- **Audit and Upgrade Warehouse F:**
  Investigate staffing, layout, and process bottlenecks in high-delay warehouses.
- **Implement Real-Time Tracking and Route Optimization:**
  Use predictive analytics and logistics tools to proactively address delays.
- **Introduce Tiered Service Levels:**
  Formalize **Service Level Agreements (SLAs)** to ensure better handling of high-priority shipments.
- **Rebalance Customer Support Resources:**
  Link support demand forecasts to delivery performance data to improve resource planning.
- **Improve Cross-Department Coordination:**
  Align operations, logistics, customer care, and discounting strategies for a cohesive delivery experience.

Limitations

- **Lack of Time-Series Data:**
  Shipment trends over time were not explored in depth due to lack of timestamps or daily logs.
- **Limited Context on External Factors:**
  Traffic, weather, and regional disruptions were not included but may significantly influence delivery times.
- **Assumption of Causation from Correlation:**
  Observations are correlational —predictive ML models could offer deeper insights.
- **Single Dataset Scope:**
  Data from only one company limits generalization; insights may not fully translate across industries or geographies.

Conclusion

- This analysis highlights key operational and customer experience challenges in the shipping lifecycle of an e-commerce business.
- Despite offering high-priority labels or discounts, many systemic inefficiencies still lead to delivery delays — notably from shipping mode choices and underperforming warehouses.
- Addressing these issues with targeted operational changes, customer-centric strategies, and improved data-driven logistics can significantly improve delivery reliability and customer satisfaction.
- By publishing this analysis, I aim to showcase how data analysis can uncover impactful business insights from real-world operations and guide strategic decisions across supply chain networks.