Capstone Project: E-Commerce Operational Efficiency Analysis

Author: Giorgos Pantsios **Date:** January 2026 **Certification:** Google Data Analytics Professional Certificate

Executive Summary

This analysis evaluates the operational efficiency of the company's two primary order fulfillment methods: Amazon Fulfillment and Merchant Fulfillment. By analyzing over 128,000 sales records, the study identified a critical performance gap. Amazon Fulfillment demonstrates superior reliability with a cancellation rate of **12.79%**, whereas Merchant Fulfillment lags with a cancellation rate of **17.47%**. This 4.68% variance represents a significant operational inefficiency. The report recommends shifting high-velocity inventory to Amazon's network to stabilize revenue and targeting a reduction in Merchant cancellations to under 14% through improved inventory synchronization.

Phase 1: Ask

1.1 Statement of Business Task

Background: Our client is an e-commerce retailer selling through Amazon. While sales volume is significant, the company is experiencing unoptimized revenue streams due to a high volume of unfulfilled transactions. The specific reasons driving order cancellations and returns remain unclear, making it difficult for management to intervene effectively.

Business Goal: The primary objective of this analysis is to identify operational inefficiencies within the order fulfillment process. I will analyze the "Unlock Profits with E-commerce Sales Data" dataset to determine the relationship between **product categories, fulfillment methods (Amazon vs. Merchant), and order status**. The insights will be used to provide recommendations that reduce cancellation rates and increase net revenue.

1.2 Key Stakeholders

- **Primary:** Supply Chain Management Team, Sales Director.
- **Secondary:** Inventory Managers.

1.3 Metrics to Measure

- **Cancellation Rate (%):** Orders cancelled before shipment.
- **Return Rate (%):** Shipped orders returned by customers.
- **Net Revenue:** Revenue realized after deductions.

Phase 2: Prepare

Data generation, collection, and verification.

2.1 Data Source

Dataset Name: [Unlock Profits with E-commerce Sales Data](#) **Source:** Kaggle (Open Source / CC0) **Description:** The dataset contains approximately 128,976 records of sales transactions, including Order ID, Date, Status, Fulfilment Type, SKU, Category, and Amount.

2.2 ROCCC Analysis

- **Reliable: Medium.** The dataset contains 128,976 entries, providing a significant sample size for analysis. However, as a third-party upload, it lacks official verification from Amazon.
- **Original: Low.** This is secondary data uploaded to Kaggle, not a direct database connection to the company's ERP system.
- **Comprehensive: High.** The dataset includes critical fields required for the business task: Order Status, Fulfillment Type, Amount, and SKU Categories.
- **Current: Low.** The data covers transactions from 2022. While useful for trend analysis, it may not reflect current 2026 market conditions.
- **Cited: Low.** The dataset is public domain (CC0) but does not cite the specific originating internal report.

Phase 3: Process

3.1 Data Cleaning & Manipulation

Tools Used: Python (Pandas library) **Steps Taken:**

1. **Dropped Irrelevant Columns:** Removed Unnamed: 22 (null artifact) and index (redundant).
2. **Handled Missing Values (Imputation):** identified 7,500+ records with missing Amount. Cross-referenced with Status and determined the majority were "Cancelled." Imputed values with 0.0 to preserve the records for "Cancellation Rate" calculations without skewing revenue totals.
3. **Data Type Conversion:** Converted the Date column to Datetime objects to enable monthly trend analysis.
4. **Standardization:** Renamed all columns to snake_case (e.g., ship_city) for consistency.

Phase 4: Analyze

1. Executive Summary

This analysis compares the performance of two primary fulfillment channels: **Merchant Fulfillment** and **Amazon Fulfillment**. The data indicates that Amazon Fulfillment not only handles a significantly higher volume of orders but also maintains a superior reliability metric with a lower cancellation rate compared to Merchant Fulfillment.

2. Data Overview

The following table summarizes the order volume and cancellation metrics for both fulfillment methods:

Fulfillment Method	Total Orders	Cancelled Orders	Cancel Rate (%)
Amazon	89,698	11,471	**12.79%**
Merchant	39,277	6,861	**17.47%**

3. Performance Analysis

Volume Distribution

- **Dominant Channel:** Amazon is the primary fulfillment driver, accounting for roughly **69.5%** of the total order volume analyzed ($89,698$ out of $128,975$ total orders).
- **Scale:** Amazon fulfills approximately **2.3x** the volume of Merchant-fulfilled orders, suggesting it is the preferred or more scalable method for this dataset.
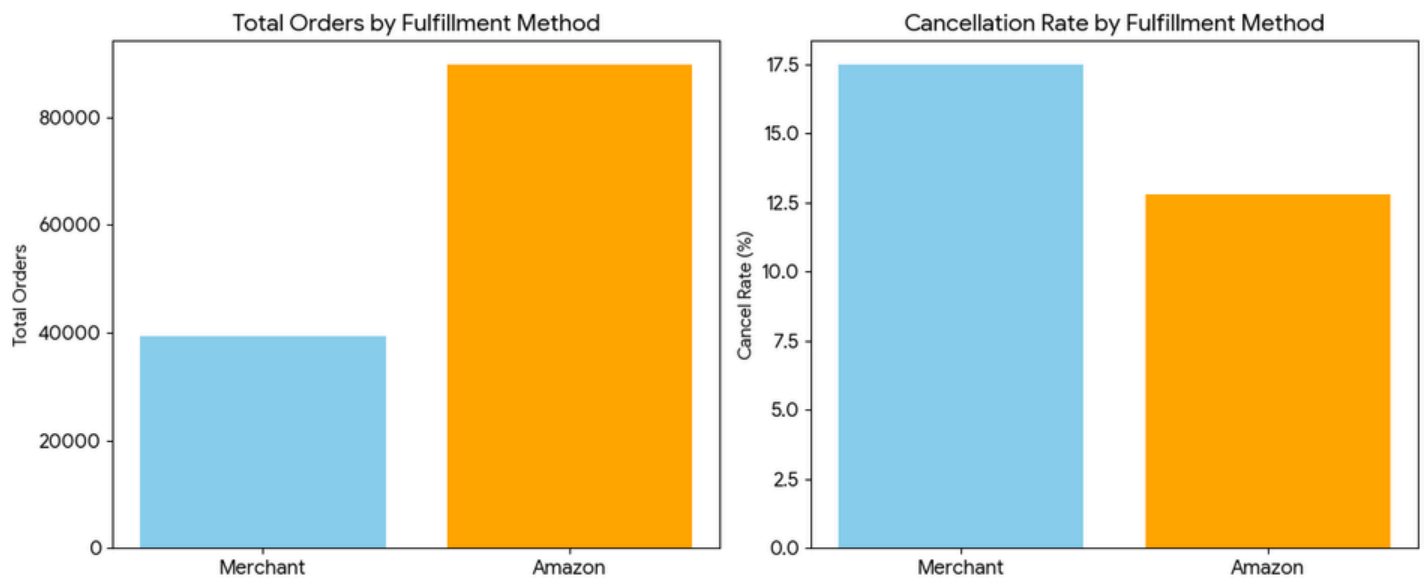
Reliability & Cancellation Risk

- **Stability:** Amazon demonstrates greater stability with a cancellation rate of **12.79%**, which is **4.68 percentage points lower** than the Merchant rate.
- **Merchant Risk:** Merchant fulfillment shows a higher volatility with a **17.47%** cancellation rate. For every 100 orders fulfilled by the Merchant, nearly 18 are cancelled, compared to roughly 13 for Amazon.

4. Conclusion

The data suggests a strong correlation between Amazon Fulfillment and higher order success rates. While Merchant fulfillment remains a vital component, it carries a higher risk of order cancellation. Transitioning high-volume or critical inventory to Amazon Fulfillment could potentially reduce the overall cancellation rate and improve customer satisfaction.

Phase 5: Share

Total Orders by Fulfillment Method — Cancellation Rate by Fulfillment Method

Phase 6: Act

1. Recommendations Based on the identified 4.68% variance in cancellation rates between the two fulfillment methods, the following actions are recommended to recover lost revenue:

Short-Term Strategy (Immediate Impact)

- **Prioritize Amazon Fulfillment:** Immediately shift inventory allocation for top-selling SKUs to Amazon Fulfillment. Leveraging Amazon's lower cancellation rate (12.79%) will instantly maximize the completion rate of high-volume orders.
- **Audit "Out of Stock" Triggers:** The high Merchant cancellation rate (17.47%) strongly suggests inventory data discrepancies. Implement a daily synchronization check between the warehouse management system and the sales platform to prevent customers from ordering out-of-stock items.

Long-Term Strategy (Process Improvement)

- **Root Cause Investigation:** Conduct a secondary analysis on the Cancelled_Orders subset. Differentiate between cancellations initiated by the seller (stockouts) versus the customer (delivery delays) to refine the operations policy.
- **Performance KPIs:** Establish a new Key Performance Indicator (KPI) for the Merchant Fulfillment team, setting a target to reduce the cancellation rate from 17.47% to **<14.0%** within the next quarter.

2. Next Steps

- Present these findings to the Supply Chain Management Team and Sales Director.
- Initiate a pilot program moving the top 50 riskiest Merchant-fulfilled SKUs to Amazon FBA to measure net revenue lift.