Optimizing Supply Chain Efficiency and Profitability for a Startup: A Data-Driven Analysis

December 12, 2024

DS210 – B1; Professor Chator

Luis Felipe Alfaro

For this project I focused on analyzing the supply chain data of a cosmetics startup, aiming to optimize operations and improve profitability. The analysis is implemented through a comprehensive codebase that processes and analyses a dataset with many attributes including product price, costs, lead time, shipping costs, and customer demographics. The code begins with revenue analysis, calculating product-level revenues and identifying the most profitable products **(Figure 1)**. The skincare product line emerged as the top performer, generating over $1 million in revenue. This provides a clear understanding of which product categories drive the company's financial success. Following this, the code performs a supply chain efficiency analysis by evaluating shipping times and costs, highlighting correlations and inefficiencies like bottlenecks caused by extended lead times.

```
Most Profitable Products:
skincare: $1052073.30
haircare: $632896.21
cosmetics: $600580.45
```

Figure 1: Output of code, displaying the most profitable products ranked by revenue.

Then, to provide deeper insights, I also used predictive modeling through linear regression to predict revenues based on price and quantity sold, as well as the relationship between lead times and costs. The model reveals that a $1 increase in price leads to an estimated $5 increase in revenue, assuming demand remain stable. This helps forecast outcomes and improve decision-making.

The analysis also uses K-means clustering to group products based on cost and lead time, providing information into segmentation and enabling targeted strategies for each cluster. This segmentation allows of tailored strategies, such as streamlining logistics for costly clusters or negotiating better terms with suppliers. This clustering approach provides a foundation for differentiated supply chain management strategies.

An outlier detection module was employed to flag unusual cost patterns, allowing identification of inefficiencies. The outliers were found using z-scores, but upon further investigation, no outliers were found. This confirms the dataset's integrity while highlighting specific products or suppliers for routine monitoring to prevent inefficiencies.

For enhanced data interpretation, I employed visualization techniques, generating a scatter plot with a trend line, of cost vs lead time, to uncover patterns and relationships. To facilitate better analysis, I also included a feature engineering, creating derived metrics like cost per lead time, which allows for identification of the more costly products. The output of the code, in the terminal, provides important summary information and the longer data calculations I obtained are saved into CSV files, ensuring the results are easily accessible and reusable.

The analysis yielded important findings. First, skincare products dominate profitability generating the most revenue. Therefore, the underperforming categories - haircare and cosmetics - should be reevaluated for potential cost reductions to increase profitability. A weak negative correlation between cost and lead time suggests that increasing costs does not always reduce lead time, indicating inefficiencies in supply chain management **(Figure 2)**.

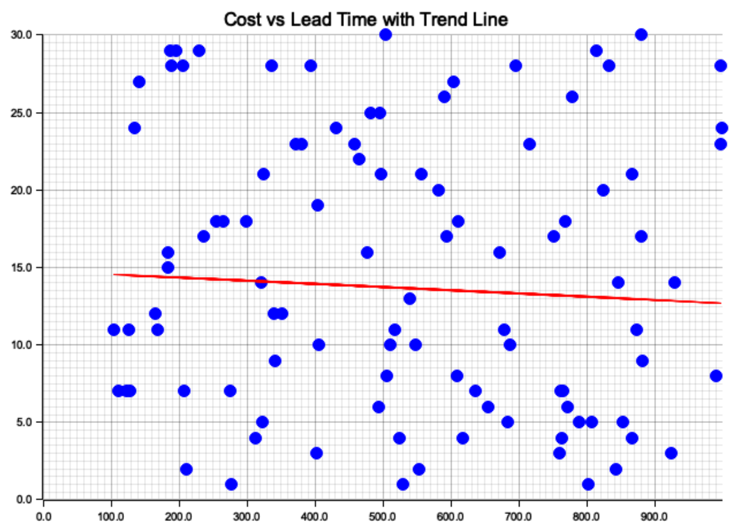

Figure 2: Graph of cost vs lead times, trend line included (-0.2).

Locations such as Kolkata and Mumbai emerged as cost-intensive, suggesting the need for alternative logistics providers or more efficient transportation strategies **(Figure 3)**.

```
High-Cost Shipping Locations:
Kolkata: $144.04
Mumbai: $137.48
Bangalore: $103.47
Chennai: $93.78
Delhi: $76.05
```

Figure 3: Output after running the code. Displays the costliest shipping locations for the startup.

Moreover, the analysis revealed significant disparities in product profitability, with certain products dominating revenue generation, while others underperformed. The clustering analysis further segmented products into meaningful groups, enabling tailored strategies to improve performance. Products in the high-cost, long-lead time cluster need specific attention like lead time improvement through renegotiated supplier contracts.

Based on these findings, I recommend reducing shipping costs in high-cost regions by exploring alternative carriers or transportation methods. More resources should be allocated to profitable categories like skincare, while reassessing underperforming categories. Consider reducing costs or discontinuing unprofitable products to streamline operations. Products with extended lead times, require process improvements. This may involve enhancing internal processes to reduce delays. Real-time analytics should be used to monitor lead times, costs, and revenues. This will enable proactive adjustments to supply chain strategies, ensuring sustained efficiency and profitability. For high cost, long-lead-time product clusters, implement targeted interventions such as investing in cost reduction technologies or finding closer suppliers to decrease lead times.

This project provides a comprehensive data-driven framework for analyzing and improving the supply chain of a cosmetics startup. Through identifying key revenue drivers, inefficiencies, and cost-saving opportunities, the company is equipped with actionable insights to streamline operations and enhance profitability. Implementing the recommendations outlined will allow the startup to achieve operational excellence and remain competitive in a dynamic market. Advanced analytics like clustering and regression were implemented directly in Rust, demonstrating its versatility as a tool for data science applications. These results identified key revenue drivers, inefficiencies and provided a roadmap for actionable improvements.