Predicting e-commerce order cancellations

Problem statement

How to reduce the cost associated with e-commerce order cancellations by xx% in Q4, 2022 by predicting order cancellations.

Context

The company's canceled orders in Q2, 2022 amounted to 9% of total sales. Order cancellations not only 'steal' revenue the company would otherwise have kept, but also disrupt the fulfillment process and result in additional costs. The benefits of being able to predict which orders will be canceled are as follows:

1. It would help reduce costs associated with processing canceled orders (by placing them at the end of the fulfillment queue and/or by implementing proactive measures to prevent cancellations).
2. It would allow for improved operational efficiency: fulfillment process optimization and better inventory management.
3. Understanding what factors affect users' decision to cancel the order may help the business improve its product offerings, marketing strategy, and customer service.

The clients for this project would be:

- Fulfillment operations and supply chain management teams, since knowing which orders are at risk of cancellation would allow them to optimize the fulfillment process and inventory management.
- Customer relationships team, since they will be able to proactively reach out to customers at risk of canceling and address their concerns.
- Potentially, Sales team, since recovering orders that would have been canceled otherwise would increase overall sales.

Criteria for success

The cost associated with e-commerce order cancellations would have been reduced by xx% in Q4, 2022.

Scope of solution space

Creating a model to predict order cancellations based on historical data of Q2 2022, which includes order date, product features, price, shipment method, shipment destination, promotions, and sales channel.
Deliverables: a GitHub repo with notebooks for data wrangling, EDA, pre-processing, modelling, project report, and slide-deck.

Constraints

We don't want to hurt customer experience (for example, by removing the cancelation option or implementing cancellation fees).

Stakeholders

Chief Operations Officer
Supply Chain manager
Customer Support Manager
Sales Director

Data sources

Amazon Sale Report for Q2, 2022: https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data/data
I assume Sales team of any e-commerce company would have this kind of report available.

Target feature: Status

Features to use in analysis
- Fulfillment
- Sales Channel
- Ship-service-level
- Style
- SKU
- Category
- Size
- Amount
- Ship-city
- Ship-state
- ship-postal-code
- Promotion-ids
- B2B
- Fulfilled-by
- Date

Additional data that could have been helpful in this analysis:
1. Customer data: demographics, whether the customer is new or returning, history of previous order cancellations, etc (could be acquired from CRM system).
2. Online behavior data: like the number of website visits preceding purchase, time spent on the website, number of product page visits, previous product reviews, etc (could be acquired from web analytics platforms like Google Analytics or Adobe Analytics).