Data Scientist Interview Challenge

Context: Pricing and Shipping Optimization for a Medical E-Commerce Platform
Estimated Time Commitment: 6–10 hours
Submission Format:

- 1–2 page summary document (PDF)
- Code (GitHub repository or ZIP file)
- 60-second video demo

- 1-page technical reflection

Overview

Your client is an e-commerce business looking to improve its profitability. They suspect they are leaving money on the table by not pricing dynamically. Your task is to analyze their sales data and recommend a pricing strategy that maximizes profit by modeling price elasticity and forecasting demand changes.

The goal is not to build a complete system at scale but to understand how you think about data science, engineering trade-offs, and rapid prototyping. You are encouraged to use AI-assisted development tools such as GitHub Copilot, ChatGPT, or others to speed up your process. If you do, please indicate where and how you used them.

Dataset

Use the dataset (or any other similar dataset if this one isn't working): "Unlock Profits with E-commerce Sales Data" from Kaggle

https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data

This dataset contains ~120k records with fields such as SKU, category, style, channel, etc.

Output 1: Summary Document

Submit either a **1–2 page memo,** or a **5-6 slide deck** as a PDF that summarizes your work. Be sure to succinctly summarize your work and make clear recommendations. Your document may touch on the following aspects:

1) Data Preprocessing and Feature Engineering
    - How you cleaned and standardized data
    - Description of preliminary analysis and features chosen for examination
2) Demand Curve Estimation

    - Estimate price elasticities: how do units sold relate to price per SKU/category?
    - Fit regressions or non-linear models to derive elasticity coefficients.

3) Dynamic Pricing Simulation
    - Propose a pricing adjustment framework.
    - Simulate demand and calculate profit outcomes.
    - Optimize price points for each SKU/category.

4) Segmentation Analysis
    - Segment by relevant columns: analyze how elasticity varies.

5) Evaluation & Scenario Testing
    - Compute total baseline revenue vs. simulated revenue.
    - Run scenarios: promotions, peak demand.

Output 2: Code

Submit your code via a **GitHub repository** or a **ZIP file**.

Include a README.md with:

- A summary of your solution
- Setup instructions
- How to run your code

Clearly comment and structure your code for readability

Mention any AI-assisted coding tools used

Output 3: Demo

Submit a 1-minute video or Loom recording explaining your work.