ML Engineer Assignment: Fuel Price Optimization

Business Context

A retail petrol company operates in an open market where competitors freely change their daily prices. The company can set its own retail price once per day, at the start of the day. The objective is to maximize daily profit by choosing the right price, using historical data and a daily feed of market inputs.

Objective

Develop an ML system that:

- Recommends the optimum daily retail price to maximize total profit.
- Handles data ingestion, transformation, and feature computation efficiently.
- Can be easily extended and deployed in a real-world scenario.

Data Provided

1. oil_retail_history.csv — ~2 years of daily historical data. Columns:
    - date – calendar date
    - price – company's historical price
    - cost – company's cost per liter
    - comp1_price, comp2_price, comp3_price – competitor prices
    - volume – liters sold
2. today_example.json — example of what is known at the start of a new day. Keys:

Your Task

- date
- price (company's last observed price)
- cost (today's cost)
- comp1_price, comp2_price, comp3_price (today's competitor prices)

1. Explore the dataset to understand demand dynamics, seasonality, and price relationships.

2. Design and implement a strategy to recommend the daily price that maximizes total profit.
    - You are free to choose any approach: statistical, econometric, machine learning, optimization, or hybrid.
    - You may also propose business guardrails (e.g., limit daily price changes, ensure competitiveness).
3. Validate your approach using the historical data and explain its effectiveness.
4. Demonstrate how your system would output a recommended price for the provided today_example.json.

Data Engineering

Design a data ingestion and transformation pipeline that:

- Reads daily data (simulate batch or streaming ingestion).
- Cleans and validates incoming records.
- Computes derived variables such as price differentials, lag features, moving averages, etc.
- Stores processed data in a format suitable for ML training.
- Demonstrate this pipeline using the provided oil_retail_history.csv.

Machine Learning

- Develop a modeling strategy to recommend the daily price that maximizes total profit.

- You may choose a regression, optimization, or reinforcement learning approach.
- Show how your model incorporates business rules such as:
  - Maximum allowed price change per day.
  - Competitor price alignment.
  - Minimum profit margin thresholds.
- Train, validate, and evaluate your model using the historical dataset.
- Include appropriate metrics and visualizations to justify effectiveness.

System Design

- Package your pipeline and model into a reusable function or module:
- Input: today_example.json
- Output: Recommended price and expected volume/profit

Optional

- Expose the function as an API using FastAPI or Flask.
- Include a simple configuration for batch scheduling (e.g., Airflow, Prefect, or Cron simulation).

Deliverables

- Code (notebook or scripts) that:
  - Reads the dataset.
  - Implements your chosen strategy.
  - Implements the data pipeline.
  - Trains and evaluates your ML model.
  - Outputs a recommended price given a today_example.json.
- Summary Document (2–3 pages) describing:
  - Your understanding of the problem.
  - Key assumptions.
  - Data pipeline design and technology choices.
  - Chosen methodology and reasoning.
  - Validation results.
  - Example output for the provided today_example.json.
  - Recommendations for possible improvements or extensions.