**Final Project Summary — Business Analytics Pricing Optimization**

**Case Background & Problem Context**  
The project focused on retail pricing optimization, addressing the limitations of intuition-based pricing that cause revenue leakage and unpredictable demand. Literature emphasized pricing as a key growth lever (Nagle & Müller, 2018; McKinsey, 2021).

**Objectives**

* Build a data-driven model to optimize price points.
* Forecast demand elasticity to maximize revenue.
* Ensure scalability and auditability across products.
* Deliver a deployable ML solution with clear business alignment.

**Dataset & Preparation**  
We used the Kaggle *Retail Price Optimization Dataset* (Suddharshan, 2023), integrating unit price, product attributes, and transaction details. Steps included: cleaning missing values, feature engineering (revenue = price × demand), and train-test split (80/20).

**Methodology**  
Quantity-first modeling simulated revenue outcomes. Multiple algorithms were tested: Linear, Ridge, Lasso, Random Forest, Gradient Boosting, XGBoost, and MLP. Preprocessing, cross-validation, and hyperparameter tuning were conducted using Python (scikit-learn, XGBoost), tracked in GitHub.

**Results & Evaluation**

* **XGBoost** delivered the lowest RMSE and most realistic revenue curve.
* Gradient Boosting and Random Forest were strong contenders.
* Linear and Ridge were discarded due to unrealistic negative revenue outputs.  
  The optimal revenue peak was found at approximately **$199**.

**Business Value Delivered**

* **Revenue growth** through evidence-based pricing.
* **Forecast reliability** for promotions and planning.
* **Risk reduction** with transparent, auditable analytics artifacts (metrics.json, best\_model.joblib).
* **Scalability** across products and markets.

**Critical Reflection**

* Dataset limited to a single retail context.
* Ensembles outperformed linear models for structured data (Hastie et al., 2009; Shmueli et al., 2017).
* Continuous monitoring is required due to market drift risk.

**AWS Deployment Activities**

* Packaged the trained XGBoost model and inference script into model.tar.gz.
* Uploaded artifacts to S3.
* Created a SageMaker Model by linking artifacts and specifying a container image.
* Configured an Endpoint (ml.m5.large) for real-time inference.
* This deployment ensures the model is accessible as an API, aligning with production business use cases. SageMaker endpoints allow scalable, managed hosting, removing the burden of infrastructure management.

**Conclusion**  
The project successfully transformed raw retail data into a deployable, business-ready ML solution. By identifying $199 as the revenue-maximizing price point, we demonstrated the power of business analytics to drive growth, predictability, and competitive edge.