2023 INFORMS Blue Summit Supplies Data Mining Challenge
Pricing Analytics and Optimization for eCommerce Decision-making

Abstract:

In the rapidly evolving world of eCommerce, strategic pricing plays a crucial role in driving profitability and competitive advantage. Accurate pricing not only influences consumer behavior but also impacts a company's bottom line. In industries with fierce competition and dynamic pricing, such as eCommerce, finding the optimal price for a product becomes a complex yet essential task. This task becomes even more significant considering the large volume of data generated through various eCommerce activities, providing a rich source for data-driven decision making. Leveraging this data for pricing analytics and optimization can result in improved decision-making processes, increased profitability, and enhanced business growth. Recognizing the importance and the challenges of Pricing Analytics and Optimization for eCommerce Decision-making, we have partnered with Blue Summit Supplies (BSS), a company that sells office supply products online. Based in Huntsville, Alabama, USA, BSS is continuously seeking innovative methods to maximize its profit. One of the significant challenges BSS faces is determining the optimal selling price for a large quantify of products in a market where prices can change daily or even hourly. It is highly desirable to understand the relationship between product prices and profits that involves analyzing historical data and experimenting with real-time prices, a process that can take weeks. Given the dynamic pricing strategies of market competitors, it's crucial for the company to analyze and respond to the market quickly and make prompt optimal pricing decisions. Through collaboration with BSS, the data challenge is aimed to call for novel data-driven solutions for the pricing analytics and optimization problem using BSS eCommerce data.

1. Problem Statement

The primary objective of this competition is to predict the optimal daily market prices for eCommerce products sold by BSS, with the goal of maximizing the profit. At present, BSS employs manual methods to analyze market variables and competitors' prices to make pricing decisions. The manual methods, while functional, are not highly efficient and time-consuming. Given the technological advancements and the vast amount of data available, we believe there is a significant opportunity to develop innovative data-driven methods to quickly analyze market trends and make informed pricing decisions.

The dataset for this competition comprises a set of market and sales variables from office products sold on the BSS eCommerce platform. The provided dataset spans over two and a half years, providing a comprehensive look at the market dynamics of more than 200 office products. The profitability of each product is determined by several factors. These factors include eCommerce revenue and various internal costs such as cost of goods sold (COGs), fulfillment fees, referral fees, and advertisement costs. BSS strategically adjusts product prices to maximize profitability, considering the prices set by competitors and customer demand responses. However, the dynamic market prices make it challenging for the company to accurately predict the optimal price that will maximize profitability. The existing manual pricing decision process is not only time-consuming but also lacks the robustness that data-driven decision-making can potentially provide. Consequently, there is a significant need to develop a data-driven process. This process should ideally determine the market price that yields the maximum profitability by analyzing the eCommerce market data. In summary, this data challenge seeks innovative solutions to effectively address this problem with two main goals:

- Determine the optimal selling price for an eCommerce product by leveraging more than two years of historical data comprising various sales and market variables.
- Predict the expected daily profit for that product given a determined optimal selling price.

The developed solution should ideally lead to more profitable pricing decisions and strategies in a dynamic eCommerce market.
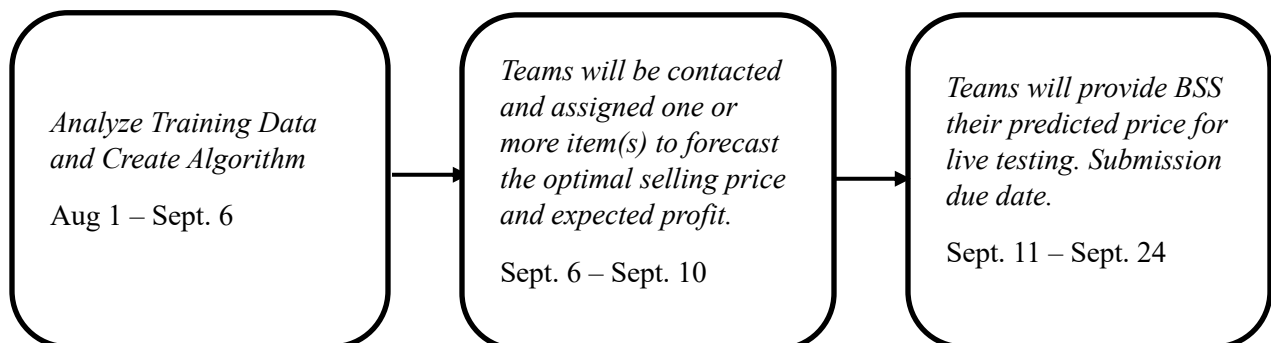
2. Timeline

The data challenge will be conducted from August 1st to September 24th and will be divided into two phases: the Training Phase and the Online Testing Phase.

Training Phase (August 1st to September 6th): During the training phase, contestants will use the provided training dataset to develop their models. This dataset contains a set of sales and market variables of the eCommerce products sold by BSS. Contestants can analyze this dataset and develop pricing analytics and optimization models to 1) determine optimal selling prices for an eCommerce product and 2) predicting the expected daily profit for the product based on the optimal selling price.

Online Testing Phase (September 6th to September 24th): In the testing phase, the original training dataset will be updated with current market data. Contestants will use the updated dataset to apply their models and predict the optimal price and expected profit of a product. Each participating team will be assigned a specific product. The team's task is to predict the optimal daily price and profit for their assigned product based on its near real-time market data from September 11th to September 24th. *BSS will update the testing dataset two times (Sept. 11 and Sept.18). Therefore, each contestant will have to data to perform a t+6 forecast. That is, teams will forecast the price and profit for a one-week look ahead.* The predicted optimal prices will be applied in the real world on the BSS eCommerce platform. In other words, BSS will use your predicted prices to set their product prices on the BSS eCommerce platform for a week. This will initiate a two-week period of online live testing to evaluate the effectiveness of each team's proposed pricing solution and compare if the proposed data-driven solution can increase sales profit compared to existing manual methods.

This hands-on, real-world testing approach offers contestants a unique opportunity to see the direct impact of their work. It also provides BSS with practical, data-driven insights to enhance their pricing strategies. The ultimate goal of this data challenge is to leverage the power of data analytics and machine learning to improve eCommerce pricing decisions, leading to enhanced profitability and competitive advantage.

Analyze Training Data and Create Algorithm

Aug 1 – Sept. 6

→

Teams will be contacted and assigned one or more item(s) to forecast the optimal selling price and expected profit.

Sept. 6 – Sept. 10

→

Teams will provide BSS their predicted price for live testing. Submission due date.

Sept. 11 – Sept. 24

3. Dataset Description

Training Dataset

The training dataset contains a large collection of data for 227 different office products sold on the BSS eCommerce platform over a period of two and half years (01/01/2021 – 07/09/2023). Here's a brief description of the variables in the dataset:

- `Sku`: This is the product ID, a unique identifier for each product.
- `salesdate`: This represents the date on which a particular product was sold.
- `price`: This is the price at which the product was sold on a given day.
- `unitsordered`: This variable shows the number of units of a product ordered on a particular day.
- `sales`: This represents the total revenue generated from the sale of a product on a given day (it is calculated as the product's price times the number of units ordered).
- `cogs`: This stands for "Cost of Goods Sold", which is the direct cost incurred by the company to produce or purchase the product.
- `fba`: This is the eCommerce fee associated with selling the product. It includes the costs of storage, packing, and shipping handled by Amazon.
- `reffee`: This is the eCommerce platform fee associated with selling the product (15% of sales).
- `adspend`: This represents the advertisement cost associated with the product.
- `profit`: This is the profit obtained from selling the product, calculated as sales minus the sum of cogs, fba, reffee, and adspend (profit = sales – cogs – fba – reffee – adspend).
  - `comp_x_price`: This represents the price of a similar product sold by a competitor. Up to 5 competitors' price data are available for each product (67 items have 0 competitors, 65 items have 1 competitor, 56 items have 2 competitors, 28 items have 3 competitors, 9 items have 4 competitors, 2 items have 5 competitors).
- **`comp_data_min_price`**: This is the minimum price among all competitors for a similar product.
- **`comp_data_max_price`**: This is the maximum price among all competitors for a similar product.
- **`managed_fba_stock_level`**: This represents the available quantity of the product in stock.
- **`min_price`**: This is the minimum allowable selling price for the product
- **`max_price`**: This is the maximum allowable selling price for the product.

Additional Note:

- Profit = Sales – Cost of Goods Sold (COGs) – eCommerce Fee (FBA) – Referral Fee (eCommerce platform fee – 15% of sales) – Ad Spend.
  - COGs, eCommerce Fee, and Referral fee are all stagnant variables that do not change.
  - Sales is a function of units sold x price.
- We have max and min prices because we have cost internal restrictions. The product prices are required to be within the range between max and min prices.
- The goal is to maximize profitability.

Testing Data

After developing an optimal pricing and profit prediction solution based on the training data, contestants will be provided with a testing dataset. The testing dataset will cover a two-week window for evaluation purposes. Each team will be assigned a specific product(s), for which they will need to predict the optimal daily price based on the product's real-time sales data from September 11th to September 24th.

During this two-week online live testing phase, BSS will set their product prices on their eCommerce platform according to the prices predicted by the contestants. This will allow us to validate whether the proposed data-driven approach can increase profit compared to the current manual pricing decision approach. Please note that for marketing strategy reasons, the predicted price should end in $0.05 or $0.09.

This data challenge offers an opportunity to leverage real-world data to develop and validate innovative solutions for pricing optimization in eCommerce. It's a chance to directly impact business decisions and profitability, while gaining hands-on experience in data analytics and machine learning.

4. Evaluation Criteria

We call for participants to develop innovative pricing analytics and optimization solutions given the valuable eCommerce dataset from the BSS platform. We will make a comprehensive evaluation of each solution. Four finalists will be selected based on both online live testing performance and the methodology clarity and soundness in submitted technical report.

Technical Report of the Proposed Solution:

- Data Processing: Does the report include thorough data preprocessing and exploration to understand the data?
- Methodology Innovation: if the solution introduces novel data analytics or optimization techniques or ideas? If any novel or interesting findings discovered based on the data mining of the BSS sales and pricing dataset.
- Generalization and Scalability: is the developed pricing optimization approach generalizable for practical real-world applications?
- Results Interpretation and Insights: Are the results clearly explained and interpreted? Does the report provide actionable insights for eCommerce decision-making?
- Presentation Clarity: Is the report well-organized and written in clear, concise language?

Numerical Evaluation:

- Daily Profit Prediction: evaluate your predicted daily profit versus actual profit given your model-optimized product prices using R2 and MAPE as metrics over the two-week testing phase window.
- Profitability Analysis: we will compare the profit statistics of your proposed data-driven solution and the existing manual method using by BSS to evaluate if profitability increases compared to the existing method.

5. Submission
- **Report Submission Deadline: September 24th**
- A written technical report of your developed solution for Pricing Analytics and Optimization. Maximum 5 pages (single column) includes introduction, data analysis, pricing decision methodology, results interpretation, conclusions, and references.
- The report should include the list of team members, their affiliations, and contact information (email and phone). One team member is identified as the main contract.
- Provided your predictions to the provided testing file towards the end of August for live testing.
- Reproducible code

6. Prizes

The submissions for this competition will be evaluated by a distinguished panel of judges, comprising members from the INFORMS Data Mining Society and representatives from Blue Summit Supplies (BSS). This panel will select four finalist teams based on their innovative approaches to pricing analytics and optimization, the effectiveness of their solutions in the testing phase, and the soundness and clarity of their solution report.

The selected finalists will be invited to present their work at the 17th INFORMS Workshop on Data Mining & Decision Analytics (DMDA) to a broad audience including academic researchers and industry practitioners. The DMDA Workshop will be organized on October 15th in conjunction with the 2023 INFORMS Annual Meeting in Phoenix, Arizona, USA. After the finalist presentations, the judging panel will decide the final winners. A total prize fund of $2,500 has been allocated for the top two winners:

- The First Prize winner will be awarded **$1,500**.
- The Second Prize winner will receive **$1,000**.
- All finalists will be recognized for their outstanding work with a finalist award plaque from the INFORMS Data Mining Society.

Competition Chairs

Hieu Pham
Assistant Professor
Department of Information Systems, Supply Chain, and Analytics, College of Business
The University of Alabama in Huntsville
hieu.pham@uah.edu

Shouyi Wang
Associate Professor
Department of Industrial, Manufacturing & Systems Engineering, College of Engineering
The University of Texas at Arlington
shouyiw@uta.edu