Optimisation of Retail Product Selection

MAST90014 Optimisation for Industry
Group Project

Author
Steven Nguyen
Shenru Duan
Kwan Sheng Ong
Qihan Zou
Haocong Chen

School of Mathematics & Statistics
The University of Melbourne
May 19, 2023

Contents

1 Introduction **2**

2 Problem Definition **2**

3 Data **3**
3.1 Kaggle Dataset 3
3.2 Synthetic Data 4

4 Base Model **5**
4.1 Modelling Demand 5
4.2 Modelling Captured Customers 5
4.3 Model Specification 5

5 Base Model - Product Selection **7**
5.1 Model 7
5.2 Experimental Results 7

6 Adding Bulk Purchase Freight Costs **9**
6.1 Results 9

7 Model with Free Shipping **11**
7.1 Model 11
7.2 Free shipping model with different shipping factors 12
7.3 Low Budget, Low Inventory 13
7.4 High Budget, Low Inventory 14
7.5 Low Budget, High Inventory 15
7.6 High Budget, High Inventory 16

8 Model with Discounts **17**
8.1 Results 18
8.2 Results - Return on Investment 19

9 Conclusions **21**

1 Introduction

Retailers are often posed with the problem of choosing and pricing products appropriately within the context of a large market with a variety of competitors. The number of units of a given product sold in a given period may depend on a variety of factors such as its demand in the market, its price relative to competitors, quality compared to its competitors, and more.

Immediately we see that there is a clear conflict driven by price increasing raw revenue per unit sold but also potentially reducing the demand for that product. Retailers also regularly decide when to add discounts to products to move inventory by attracting potential new customers, however this comes at the cost of decreasing their raw revenue per unit sold. Choosing which products to sell, and when to discount the products from their RRP (recommended retail price) given limited resources such as money and inventory is therefore a potentially interesting optimisation problem.

This problem will be primarily motivated by the following Kaggle dataset which contains timeseries data listing the retailer's products and associated data such as freight prices, unit prices, and how many customers bought the product in the time period. While there are a variety of problems that can be tackled based on this dataset, we consider primarily the problem of *product selection* for a single period based on fixed demands. These demands are computed based on the fixed prices and product ratings in the data. We develop a simple model based on this and then consider multiple scenarios and extensions involving

- Varying inventory and budget,
- Setting discounts on prices,
- Adding free shipping

Our main goal is to understand using our model, the best strategies of choosing products to sell based on varying conditions.

2 Problem Definition

Formally, we can define our problem as follows. First, let I be the catalogue of items that we are able to buy and sell from manufacturers.

- Denote $C = \{0, 1, 2, ..., \bar{c}\}$ to be the set of product categories.
- For each $c \in C$, let I_c be the products in category c.
- Then let $i_{ck} \in I_c$ denote product k in category c.

Then each item $i_{ck} \in I$ has the following as data -

- A recommended retail price (RRP) p. This is the base price the product is sold at.
- A bulk purchase price e, which is how much the retailer purchases the product for from suppliers.
- A product rating r, taking values between 0 and 5.
- A weight w, in grams.

Now for each category $c \in C$ we define the customer base as follows -

- We define the base customers, N_c^b, as the customers who are willing to purchase an item from category c at RRP, without any free shipping.
- We define market customers, N_c^m, as potential customers in the market who are only captured through discounts or free shipping.

We define the relative demand dem_{ck} of product k in category c to be a function f of the price and rating such that -

$$\begin{aligned} dem_{ck} &\in [0, 1] \quad && \forall c \in C, k \in I_c \\ \sum_{k \in I_c} dem_{ck} &= 1 && \forall c \in C \end{aligned}$$

In other words, the relative demand represents the proportion of customers in the market for category c that would buy product i_{ck}.

Our objective is to maximise profit with the main decision variables are the following -

- x_{ck}, the number of items to buy (and sell) of item i_{ck}
- d_{ck}, the discount we are to set on item i_{ck}.
- $fship_{ck}$, whether we are to offer free shipping on item i_{ck}.

This can be considered a sort of knapsack problem with with the additional goal of investigating when to offer free shipping and how much to discount products to capture a some of the market customers without eating into our profit. A couple of base assumptions we will make to simplify our problem are -

- The relative demand for each product i_{ck} in category c is fixed as a function of the price and rating.
- We assume customers do not buy more than one of one kind of product.
- Inventory will be looked at in terms of units and not the actual volume of the item in a warehouse.

We will introduce more assumptions and modelling decisions further when we specify our model in the sections below.

3 Data

3.1 Kaggle Dataset

The original Kaggle dataset contains product category, product ID, daily sales quantity, product price, freight price, number of customers, product volume, product rating score and three competitors' prices and freight prices [1]. For our model, we focused on seven product categories, product price, freight price, product weight, product scores, the sum of customers on different days for each product and three competitors' rating scores and prices (as shown in figure 1). In our simulation, we assume that the product rating score is independent of the price. We also assume that our max rating score is 5.

Out[27]:

product_category_name	product_id	comp_1	comp_2	comp_3	freight_price	product_score	product_weight_g	ps1	ps2	ps3	unit_price	customers
bed_bath_table	bed1	74.000000	39.240000	39.240000	10.256316	4.0	350	3.9	4.0	4.0	39.240000	915
	bed2	74.000000	74.000000	39.240000	12.940000	3.9	1383	3.9	3.9	3.9	74.000000	968
	bed3	74.000000	84.900000	39.240000	4.410000	3.3	1550	3.9	3.3	4.0	84.900000	530
	bed4	77.933333	44.154444	39.990000	12.055000	4.2	800	3.9	4.2	4.0	44.154444	515
	bed5	89.900000	163.398710	45.950000	8.760000	4.4	9750	3.9	4.4	4.0	163.398710	385
computers_accessories	computers1	94.900000	94.900000	77.900000	10.390000	4.0	173	4.0	4.0	3.5	66.342143	890
	computers2	114.491154	109.900000	77.900000	13.415000	3.5	180	4.2	3.5	3.5	77.900000	864
	computers3	139.990000	134.900000	78.712281	14.596667	4.2	922	4.2	4.2	3.5	132.970000	529
	computers4	114.491154	139.990000	77.900000	27.253036	4.2	6550	4.2	4.2	3.5	114.491154	968
	computers5	114.491154	119.900000	77.900000	10.869000	3.5	207	4.2	3.5	3.5	77.155000	763

Figure 1:

3.2 Synthetic Data

Alongside our Kaggle dataset we also generate the following data -

- The price to purchase items from manufacturers we generate by assuming randomly generated % margins on each of our products.
- We generate our market customers N_c^m randomly based on the base customers.

Various other parameters in our model we do not have access to we make assumptions and test accordingly on our model under different scenarios.

4 Base Model

4.1 Modelling Demand

We calculate dem_{ck} in the following way -

$$dem_{ck} = \text{softmax}(-p_k(5 - r_k))$$

where given a real valued vector $\mathbf{x} \in \mathbb{R}^n$, the softmax function is defined as

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{k=0}^{n} e^{x_k}}$$

We represent our relative demand in this way because it is a natural way of mapping real numbers to the unit interval while still preserving their relative magnitudes. In our case we can see that the demand for a product is less as the price increases, but becomes lower as the rating gets closer to the max rating.

4.2 Modelling Captured Customers

We model 'captured' customers from free shipping or discount offers in the following way.

- Offering free shipping on any product gives a fixed percentage ρ_{fship} of the market customers N_c^{market}.
- For each % of discount we offer on a product, this gives $\rho_{discount}$ % of the market customers N_c^{market}.

Hence the total number of customers/amount sold of item i_{ck} is

$$dem_{ck}\left(N_c^{base} + (\rho_{fship} fship_{ck} + \rho_{discount} d_{ck}) N_c^{market}\right)$$

4.3 Model Specification

This model is implemented in `final_model_clean.ipynb` in our supplementary code.

Data:

C : The set of product categories, $C = \{0, 1, 2, \ldots, \bar{c}\}$.

p_{ck} : recommended retail price (RRP) for product k in category c.

e_{ck} : the prime cost (unit price to purchase the product) for product k in category c.

r_{ck} : the rating for product k in category c. $r_k \in [0, 5]$. The rating is based on the historical record of reviews from customers.

w_{ck} : the weight, in grams, of product k in category c.

N_c^{base} : It represents the number of customers who are estimated to have a strong desire (base number of customers) to purchase the product for $c \in C$.

N_c^{market} : It represents the estimated number of customers who may buy commodity $c \in C$ in the market. They do not necessarily choose to purchase, but their shopping tendency will be affected by various events such as merchandise discounts and promotions (eg. free shipping).

dem_{ck} : measures the relative demand of a product k within category c. It can be thought of as how likely a customer is to buy the product.

Parameters: These can be considered as data as well as they are fixed constants during optimisation, but we denote them as parameters here because we would like to adjust them. Note: All parameters that have been used are defined in Jupyter. If the parameter here did not include a specific numerical value, please refer to Jupyter.

> $budg$: It is the upper bound that represented the maximum budget (in AUD) that the store can be used to purchase products. The initial value is set as $budget = 1 \times 10^9$ (AUD).
>
> W_c : The warehouse capacity allowed for category c. Each product is assumed to be weighted with one unit of storage.
>
> $(scost)_c$: the unit cost per gram of shipping a product in category c.
>
> $\rho_{discount}$: called `DISCOUNT_FACTOR` in the code, this denotes the % of customers captured from N_c^{market} for each unit of discount.
>
> ρ_{fship} : called `SHIPPING_FACTOR` in the code, this denotes the % of customers captured from N_c^{market} when free shipping is offered.

Decision Variables:

> x_k : Integer variable, it represented the number of product $k \in K_c$ be purchased by the store.
>
> d_k : Continuous variable $\in [0, 1]$, it represented the degree of discount for product $k \in K_c$.
>
> f_k : Binary variable, it represented whether the product $k \in K_c$ is free shipping or not. $f_k = 1$ for free shipping, $f_k = 0$ for not free shipping.

Model: The objective function and constraints: This model is built to maximize a store's net profit, that is, to maximize revenue minus the total costs as follows:

Model:

$$\begin{aligned}
\max \quad & \sum_{c \in C} \sum_{k \in I_c} x_{ck} \left(p_{ck}(1 - d_{ck}) - e_{ck} - scost_c w_{ck} fship_{ck} \right) && && (1)\\
\text{s.t.} \quad & \sum_{c \in C} \sum_{k \in I_c} x_{ck} e_{ck} \leq budg && && (2)\\
& \sum_{k \in I_c} x_{ck} \leq inv_c && \forall c \in C && (3)\\
& x_{ck} \leq dem_{ck} \left(N_c^{base} + (\rho_{fship} fship_{ck} + \rho_{discount} d_{ck}) N_c^{market} \right) && \forall c \in C, k \in I_c && (4)\\
& x_{ck} \in \mathbb{Z}_{\geq 0} && && (5)\\
& d_{ck} \in (0, 1) && && (6)\\
& fship_{ck} \in \{0, 1\} && && (7)
\end{aligned}$$

Explanation:

> (1): The objective function that maximized the total profit for the store.
>
> (2): This constraint represents our limited budget.
>
> (3): This constraint restricts that the store cannot store products past the inventory space.
>
> (11): This constraint represents that how much we can sell depends on the number of customers, as described in section 4.2.

5 Base Model - Product Selection

5.1 Model

First we set both our discount factor and free shipping factor to be 0. Our main goal here is to consider firstly which products to buy and in what quantities in isolation of the free shipping and discount variables.

5.2 Experimental Results

To simplify our approach here we will mainly look at the case where we only have one category of product, as the behaviour of our model is likely to be the same as we introduce more categories (it is simply introducing more products).

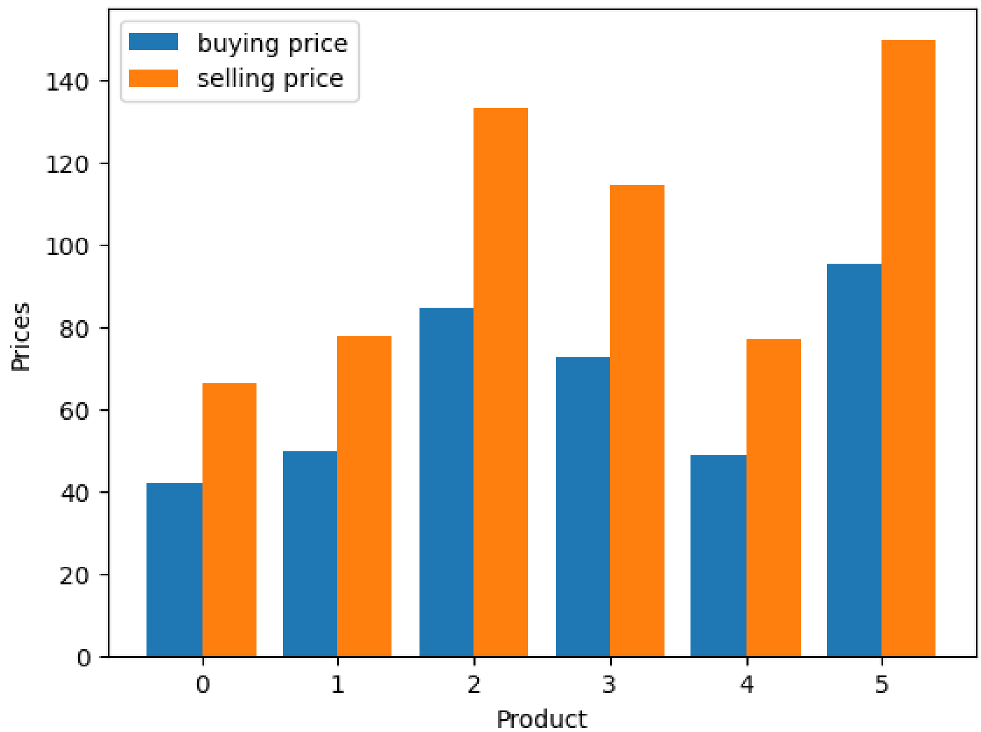
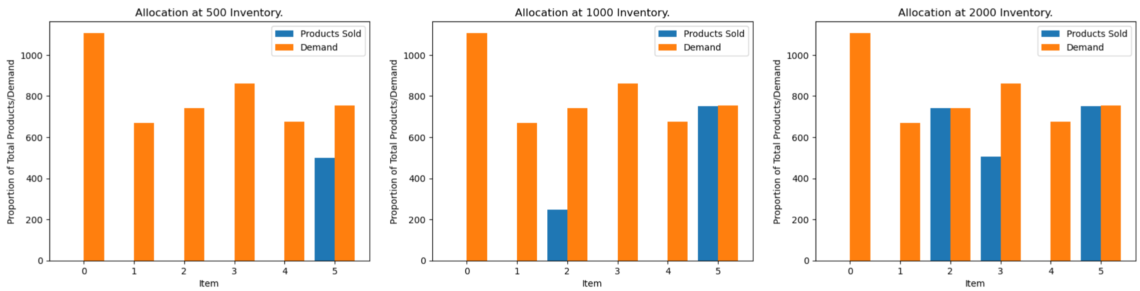


Figure 2: Buying and selling prices for our sample product category

In our first experiment we run the model under the assumption of limited inventory and effectively endless budget. In this scenario, intuitively what should happen is that since the budget does not matter, we can simply buy up the products that have a high margin, regardless of whether or not it has a high purchasing price from our supplier. We can see in figure 5 that without budget constraints this is indeed true and the model simply chooses the products with high margin.

The more interesting problem however stems from when we have limited budget. In this case we would like to choose products which have a high margin but may not be able to purchase them in volume compared to a cheaper product with lower margin. Moreover there is also a hard cap on the amount of demand on the product as specified in our model. In this case there is no obvious pattern. To observe any patterns we fix our inventory to be uncapped and look at how our product stock changes with different budgets.

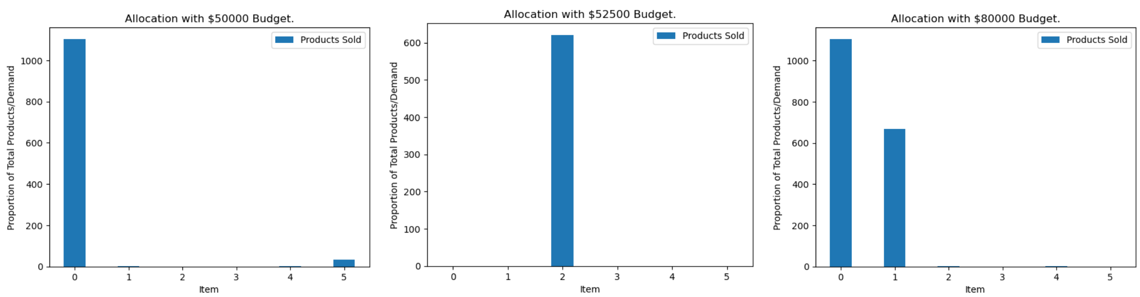
The result in figure 4 shows how shows how the products we sell can vary wildly with the amount of budget we have. Roughly speaking we observe that with very low budgets, it is best to not choose multiple kinds



(a) (b) (c)

Figure 3: Allocation under (a) 500 (b) 1000 (c) 2000 product inventory

of products as there is limited money to spend. However as we gradually increase our budget, we can start buying cheap products in bulk. At some point we reach a maximum based on the demand and then start stocking other products. As we reach enough budget we can start buying another product in bulk instead (see the change between (a) and (b)). With enough budget we can start bulk buying multiple kinds of products - this is what we see in (c).



(a) (b) (c)

Figure 4: Allocation under different budgets.

6 Adding Bulk Purchase Freight Costs

Now although each item we buy from a manufacturer incurs a cost by itself, we may also consider the cost of transporting these items to our warehouses. Freight costs are based on the weight of products, and transporting lots of product may incur extra costs through extra logistical support and vehicles required to transport around the items. To model this we suppose that for each product category, for every x grams of total product weight in our bulk purchase, we incur an additional \$$y$ cost. We model this with the following -

Extra Parameters:

- Let $bulk_c$ be the maximum weight of product type c that can be transported at once.
- Let $tcost_c$ be the cost of transporting and managing product type c.

Variables:

- Let x_{ck} be the number of units of item i_{ck} that we stock.
- Let z_c be a dummy variable encoding the number of transport units required to transport all products in category c.

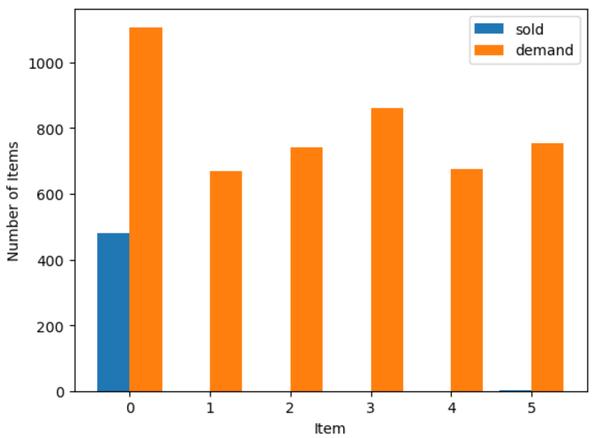
Model:

$$
\begin{aligned}
\max \quad & \sum_{c \in C} \sum_{k \in I_c} \left(x_{ck} \left(p_{ck}(1 - d_{ck}) - e_{ck} - scost_c w_{ck} fship_{ck} \right) \right) - z_c tcost_c && && (8) \\
\text{s.t.} \quad & \sum_{c \in C} \sum_{k \in I_c} x_{ck} e_{ck} \leq budg && && (9) \\
& \sum_{k \in I_c} x_{ck} \leq inv_c && \forall c \in C && (10) \\
& x_{ck} \leq dem_{ck} \left(N_c^{base} + (\rho_{fship} fship_{ck} + \rho_{discount} d_{ck}) N_c^{market} \right) && \forall c \in C, k \in I_c && (11) \\
& z_c \geq \left(\sum_{k \in I_c} w_{ck} x_{ck} \right) / bulk_c && \forall c \in C && (12) \\
& x_{ck}, z_c \in \mathbb{Z}_{\geq 0} && && (13) \\
& d_{ck} \in (0, 1) && && (14) \\
& fship_{ck} \in \{0, 1\} && && (15)
\end{aligned}
$$

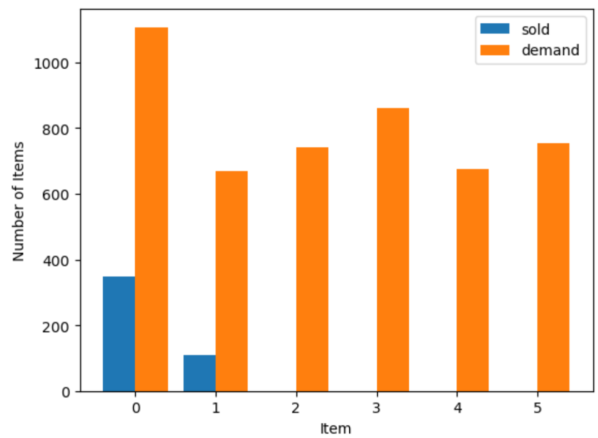
Effectively, this penalises bulk buying.

6.1 Results

Adding this additional part to our model results in changes to the distribution of products we select. Rather than stocking up lots of product in bulk when we have a low budget, it causes us to instead stock a looser spread of products. However the effect varies depending on the actual parameters set in the model. If the cost of transportation for the total weight of products is low relative to profit margin, we will still bulk buy.



(a)



(b)

Figure 5: Allocation with (a) no additional cost (b) additional bulk shipping cost. Here we set the extra cost to be $1000 per 10kg.

7 Model with Free Shipping

Free shipping is a common strategy for stores to attract more customers and sales, and more often than not it is even more effective than discounts. According to Shippo's State of Shipping Report, up to 62% of customers would not even purchase with a reatiler without free shipping. In addition, 47% of customers would spend up to a minimum amount on a purchase just to get the free shipping. [2] All this points to the importance of understanding the costs and benefit of free shipping to a retailer.

7.1 Model

Instead of a randomly generated margin to determined the unit costs of items, a softmax function is used to determine the margin of each item.

$$e_{ck} = p_{ck} \cdot \left(1 - softmax(p_{ck}^{0.3})\right)$$

For a higher unit price, this softmax function returns a higher corresponding margin. This is because often prices do not grow at the same rate as costs, resulting in higher priced items having a higher profit margin.

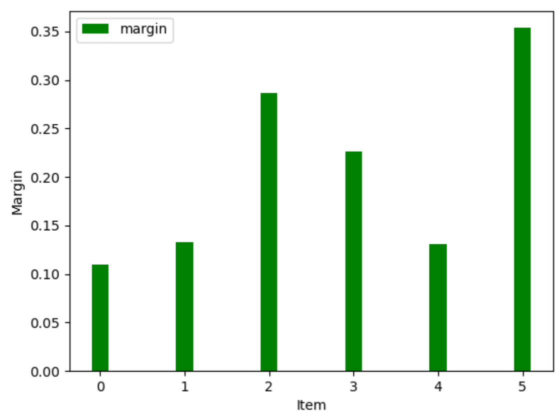


Figure 6: Bar chart of margin of each item in the category of computer accessories.

Since we are adding free shipping, the costs of free shipping has to be taken into account. In this model, we just used a fixed delivery rate ($0.01 per gram) and use the weight of each item from our data. Inventory is also now in terms of weight (g) instead of units to better simulate storage of differently weighted items.

The discount factor $\rho_{discount}$ is set to zero to solely study the effects of free shipping and what factors affect the choice to offer free shipping.

In constructing the free shipping model, the determination of the shipping cost requires careful consideration. Within the dataset, the variable 'freight price' exhibits the highest correlation with the shipping cost under investigation, making it the chosen reference variable. It is important to acknowledge that factors such as product quantity, volume, and weight can significantly impact the freight cost in the general shipping process. Additionally, to gain further insights, we assumed the freight charges imposed by three competitor entities as points of reference.
To establish the model, we employed a linear regression approach incorporating the aforementioned variables that potentially influence the freight cost. Subsequently, the minimum value obtained from this regression model was utilized to define the shipping cost for each transaction eligible for free shipping. The resultant formula is as follows:

$Freight\,price = 7.4 - 5.79082277 \times 10^{-2} \times \text{qty} - 5.16579875 \times 10^{-5} \times \text{volume} + 2.55371881 \times 10^{-3} \times \text{product weight} + 7.25291370 \times 10^{-2} \times \text{fp1} + 3.50075889 \times 10^{-1} \times \text{fp2} + 1.16122879 \times 10^{-1} \times \text{fp3}$
By utilizing this formula, we can determine the shipping cost associated with each free shipping transaction is \$7.4 .

7.2 Free shipping model with different shipping factors

After we made a line graph for the profit and the total number of products sold.Based on the graphical analysis of profit and sales under varying free shipping factors, it was observed that both profit and the number of sold items initially increase as the free shipping factor rises, reaching their peak values at around 0.9. At this point, the profit reaches its highest point, amounting to \$1587556, accompanied by a total of 22,803 items sold. we also observed that when the shipping factor increased from 0.1 to 0.9, the profit has been rising in an almost straight line. As for the relationship between the total number of products sold and the shipping factor, the same thing as profit is an upward trend. When the factor is larger, the more products we sell. Unlike profit, the image is not completely in a straight line, but there are more obvious turning points.

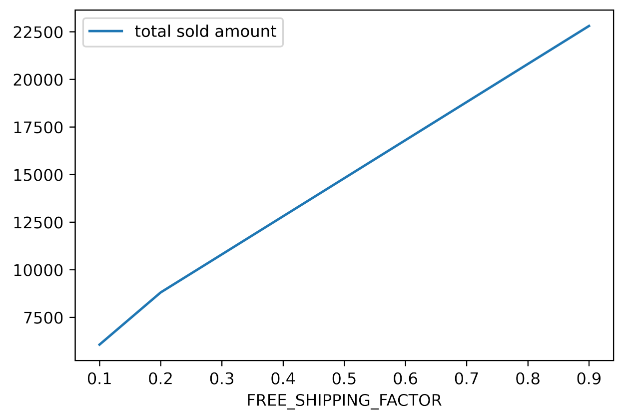


Figure 7: line graph for shipping factor and total sold amount

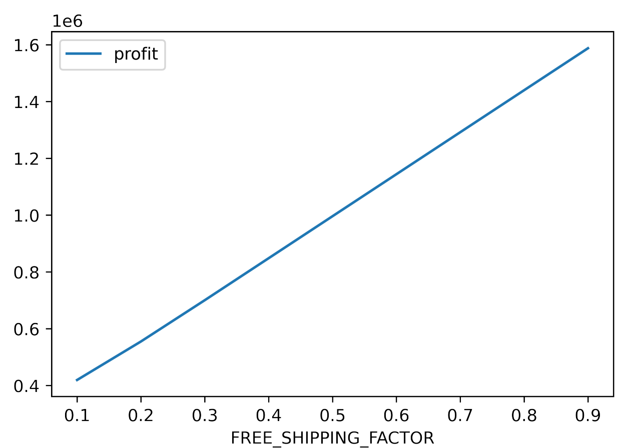


Figure 8: line graph for shipping factor and profit

In practical scenarios, offering free shipping for all customer transactions is not feasible. Therefore, it becomes necessary to select an appropriate shipping factor as a parameter in our model to optimize costs and maximize sales and profits. In this study, we investigate the impact of incremental changes in the

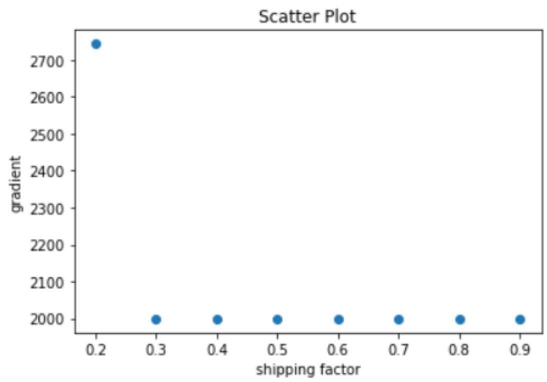


Figure 9: increasing sold items with 0.1 shipping factor increase

shipping factor on sales volume. By analyzing the scatter plot we generated, we observed a clear pattern in sales as the shipping factor increased. It is worth noting that we found that when the shipping factor is set to 0.2, the sales growth is the fastest. At this time, every additional 0.1 customers free shipping can increase the sales by about 2800. However, as the shipping factor exceeded 0.2, the sales growth rate began to stabilize, and the free shipping sales of customers who did not increase by 0.1 increased by about 2,000. Therefore, according to our analysis, setting the shipment factor at 0.2 is clearly a balance between sales volume, profitability and cost considerations. When the shipment coefficient is 0.2, we will display the supply and demand relationship in the form of a histogram. We performed a bar chart analysis to examine the

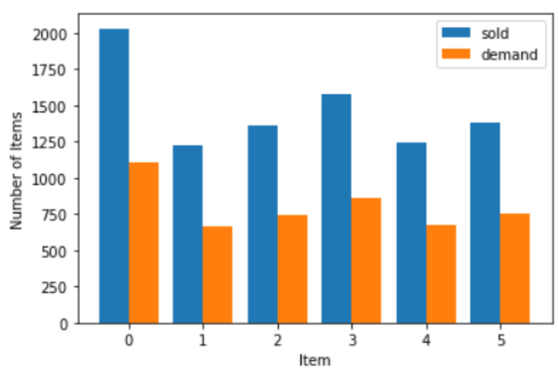


Figure 10: charts for demand and sold

demand and sales data when the profit and the number of items sold were at their maximum, especially with a free shipping factor of 0.2. It is evident from the graph that the number of items sold has greatly exceeded demand. A rough estimate suggests that roughly 2 times as many items are sold as there are demanded. The bar chart analysis highlights the significant impact of the implementation of the free shipping policy on consumer behavior, leading to a significant surge in sales volumes that exceed initial demand levels while saving costs and increasing revenue. This finding suggests that using a free shipping factor of 0.2 as an incentive can serve as an effective strategy to increase profitability and drive sales growth in the market.

7.3 Low Budget, Low Inventory

With both a low budget and low inventory space, we can only buy and store very few items. So of course the item with the highest profit margin is going be chosen since it will be the highest return of investment.

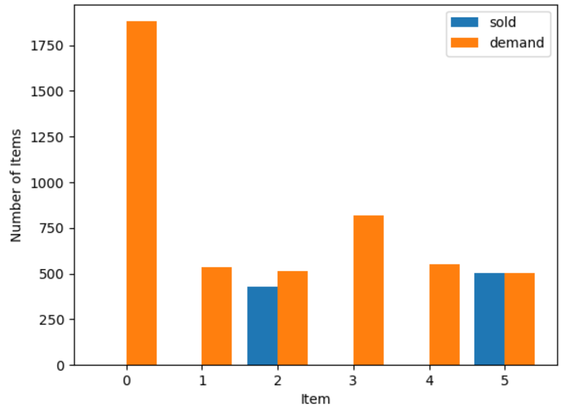


Figure 11: Bar chart of number of sales and demand of each item in the category of computer accessories with low budget and low inventory space.

7.4 High Budget, Low Inventory

With the same inventory space, but higher budget, we can see an interesting change in the choice of inventory.

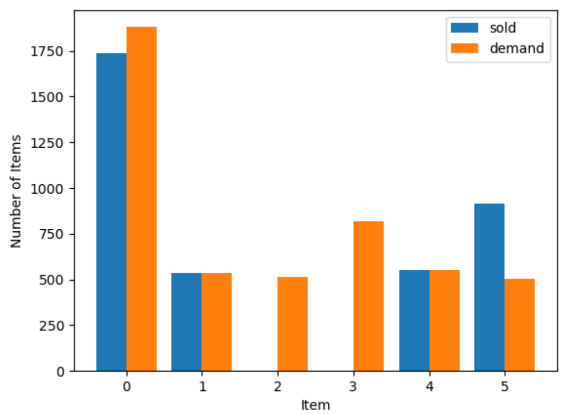


Figure 12: Bar chart of number of sales and demand of each item in the category of computer accessories with high budget and low inventory space.

Instead of stocking up item 2 and 3, we now stock item 0, 1 and 4, even if they have lower profit margin when compared to item 2 and 3. This indicates another factor affecting the choice of inventory. A quick look at the weight of each item quickly explains this puzzling phenomena.

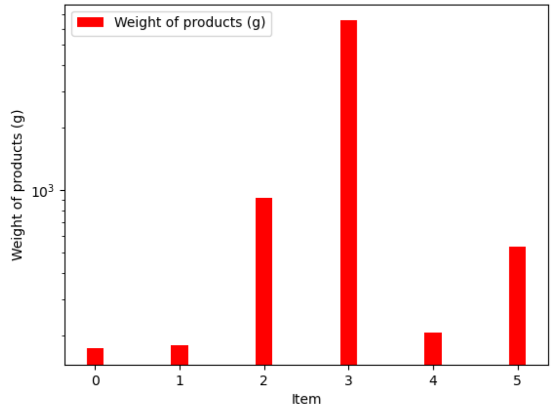


Figure 13: Bar chart of the weight of each item in the category of computer accessories.

As shown by Figure 13, both item 2 and 3 have significantly higher weight when compared to other items. Since we have low inventory space, we are not able to store as many item 2 and 3. This makes the items with a smaller weight more profitable as we can store and sell more of them, even if they have smaller margin when compared to item 2 and 3.

In the next section when we increase the inventory space, decreasing the inventory cost of each item, we will again see a shift towards stocking up higher profit margin items.

7.5 Low Budget, High Inventory

With high inventory space, the storage cost (in terms of weight) of the items is less of a concern when compared to their profit margin.

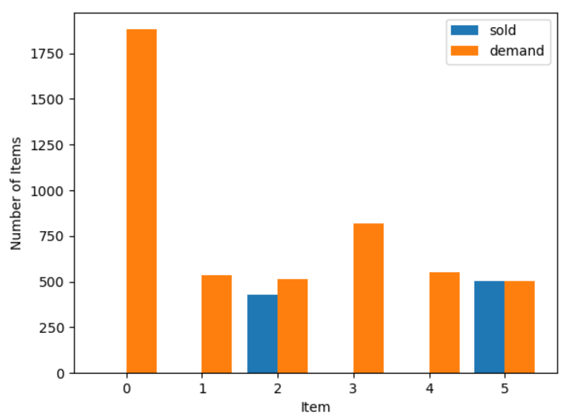


Figure 14: Bar chart of number of sales and demand of each item in the category of computer accessories with low budget and high inventory space.

As stated in the previous subsection on high budget and low inventory, Figure 14 shows we now choose to stock item 2, which has higher profit margin instead of say item 1.

7.6 High Budget, High Inventory

With high amount of budget and inventory capacity, we were able attract and serve more demand through offering free shipping for certain higher margin items.

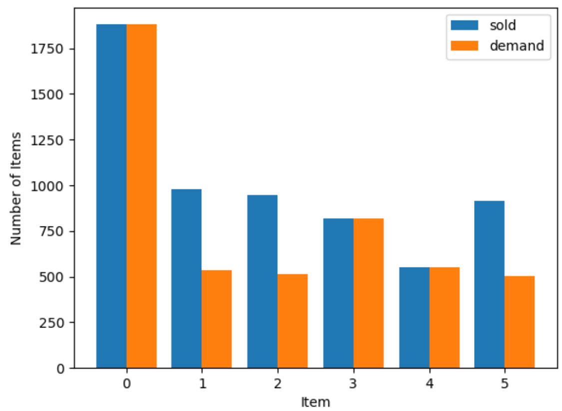


Figure 15: Bar chart of number of sales and demand of each item in the category of computer accessories with high budget and high inventory space.

However, even if Item 3 has higher margin than Item 1, free shipping was not offered to attract more customers. This is because its higher weight will result in a higher delivery cost, ultimately resulting in a lower margin than Item 1 if free shipping was offered.

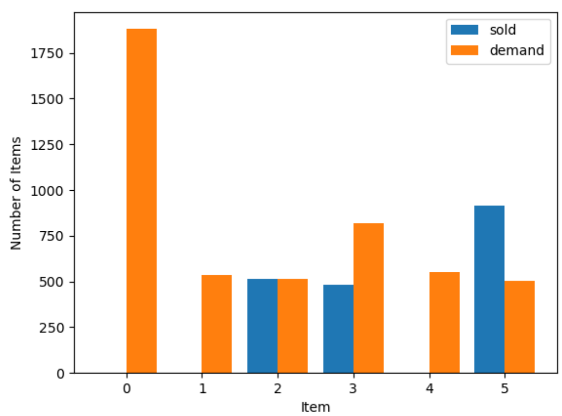


Figure 16: Bar chart of number of sales and demand of each item in the category of computer accessories with medium budget and high inventory space.

This is further supported with a slightly lower budget. In Figure 16 we can see even without fulfilling all possible demand, we have offered free shipping for Item 5 to attract a higher number of customers. Essentially, Item 5 with and without free shipping can be thought of as two "different" items with different demands and different profit margins. After stocking up the third highest profit margin item, the fourth is Item 5 with free shipping, that is why Item 5 is given free shipping before stocking up any of the other items. In summary, Item 5 is quickly given free shipping to attract more customers due to its high margin and low weight that results in low inventory cost and delivery cost.

8 Model with Discounts

Based on the basic model, we first studied the influences of budget on discounts. To make the simulation reflects real-life scenarios more accurately, we set the lower bound of the discount rate to 0.1. We fixed DISCOUNT-FACTOR equals to 5 and the budget range from 1×10^5 to 1×10^9. From plot 17, we can tell that as the budget continues to increase, our objective function value (store net profit) also increases until it reaches a maximum value. Beyond this point, the net profit remains constant and does not further increase. In figure 18, only the maximum value of the discount is plotted to provide a more direct and clearer view of changes in the discount value. Same to the changes in the net profit, the discount values increase until it reaches a maximum value and remains constant beyond that point.

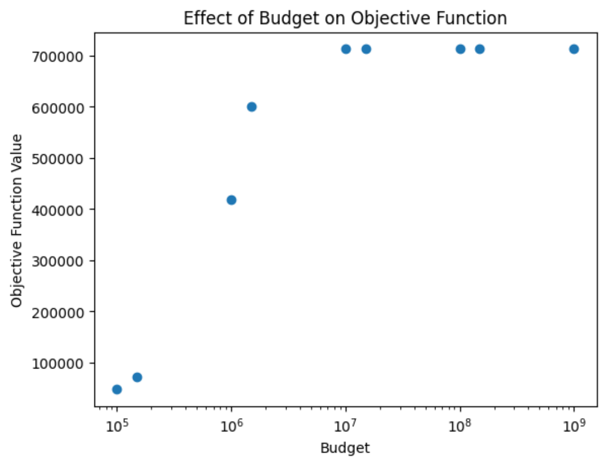


Figure 17:

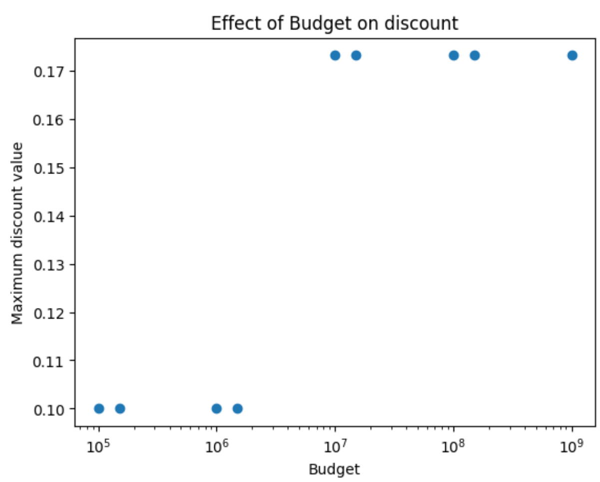


Figure 18:

Then we studied the influence of different DISCOUNT-FACTOR on the influence of discount values. For fixed budget (1×10^9), the discount factor increases from 1 to 7. The larger the discount factor, the more potential customers will bring, and the more net profit will be made (as shown in figure 19. The discount value slowly increases as the discount factor increases, until reached the max value of the discount (≈ 0.190), and then suddenly decreases for the larger discount factors. The decreases in discount may cause by two reasons. First, when the discount factor is large enough, the discount is not important to influence the number of potential customers. Second, too large discounts will decrease net store value even if it can bring extra customers. Therefore, there is a sudden decrease after the extreme discount values.

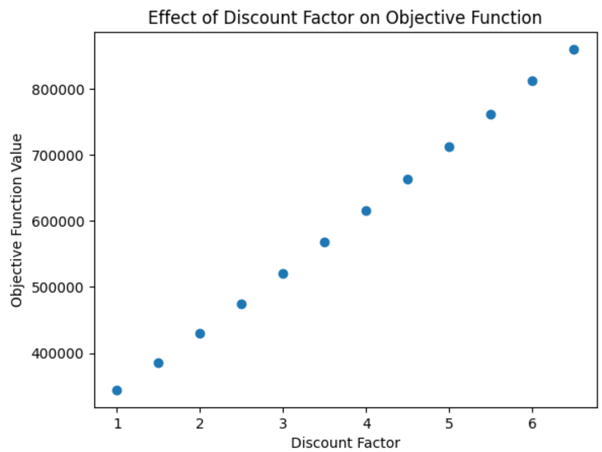


Figure 19:

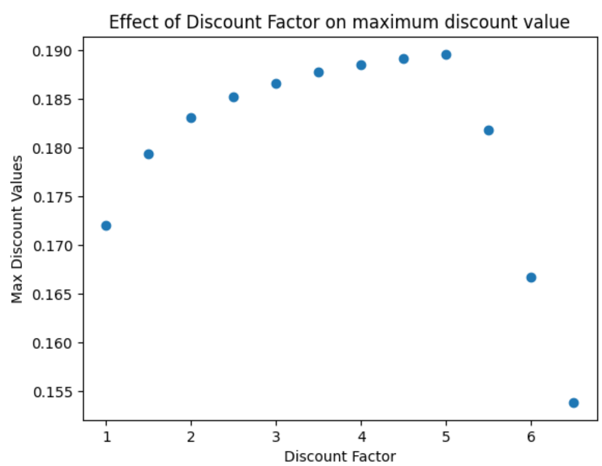


Figure 20:

8.1 Results

In conclusion, as expected, the higher discount factor increases objective values as expected. The higher budget will cause the discount factor to increase but it will approach to a constant. The increase in discount factor will bring higher objective value. But the discount value will decrease after meeting its extreme value.

8.2 Results - Return on Investment

Adjusted for the number of customers, we can see that a retail business of this type requires a large amount of capital to have a worthwhile return on investment, as shown It is obvious that higher budgets mean we are able to operate to get more raw revenue, but it would also be meaningful to understand how revenue changes as we increase budget. In other words we want to know how much return on investment we get operating the retail business under these conditions. To start we set the scenario as follows -

- We use all 9 categories in our dataset.
- We compute the number of base customers as the total number of customers observed in the market, then the market customers by multiplying the number of customers by a random constant $c > 1$.
- We compute the price per unit cost to ship a product to customs based on the average freight price divided by weight in the data.
- We assume a fixed discount factor of 1 and a free shipping factor of 20% for all product categories. In other words, free shipping captures 20% of the market customers, and each point of discount is worth 1% more customers in the market.

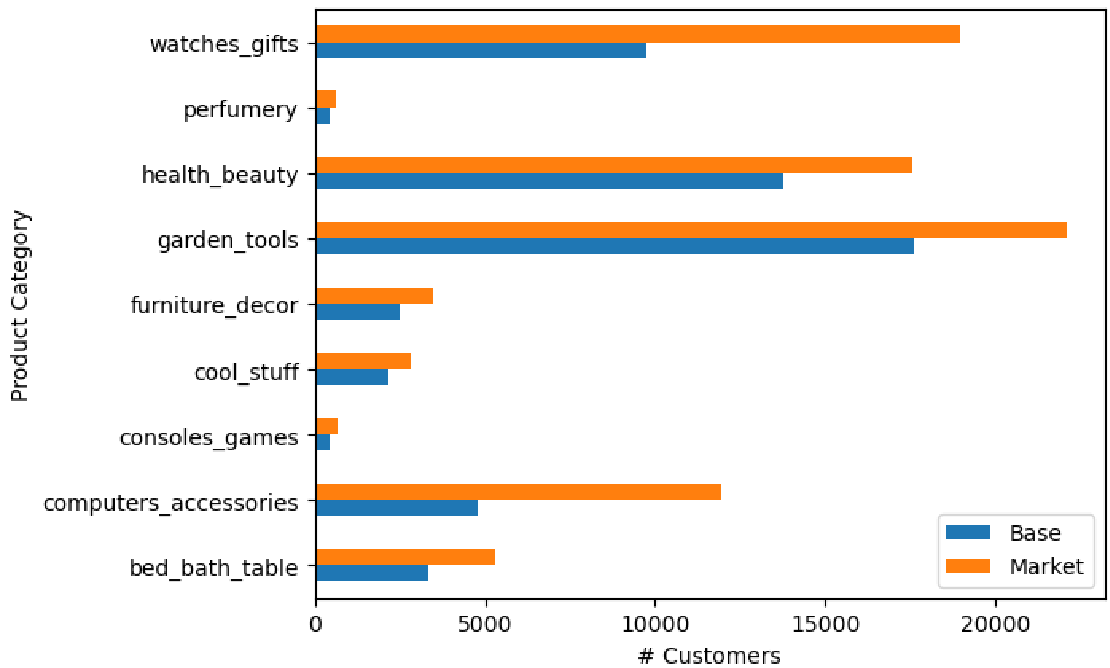


Figure 21: Total number of customers.

We can see the total number of customers in figure 21 Note that this may be low/unrealistic for this kind of business in general, but will serve our purpose.

We implement our full model and then perform optimisation under varying budgets. The results are shown in figure 22.

What we can see is that our % return on investment as we vary the amount of money we have available to pay for shipping and buy stock naturally increases to a certain point, but actually *decreases* and plateaus as we cap out our budget. In the context of our model, the plateau is due to the fixed demand and the additional customers from discount and free shipping also eventually capping out when the model decides it

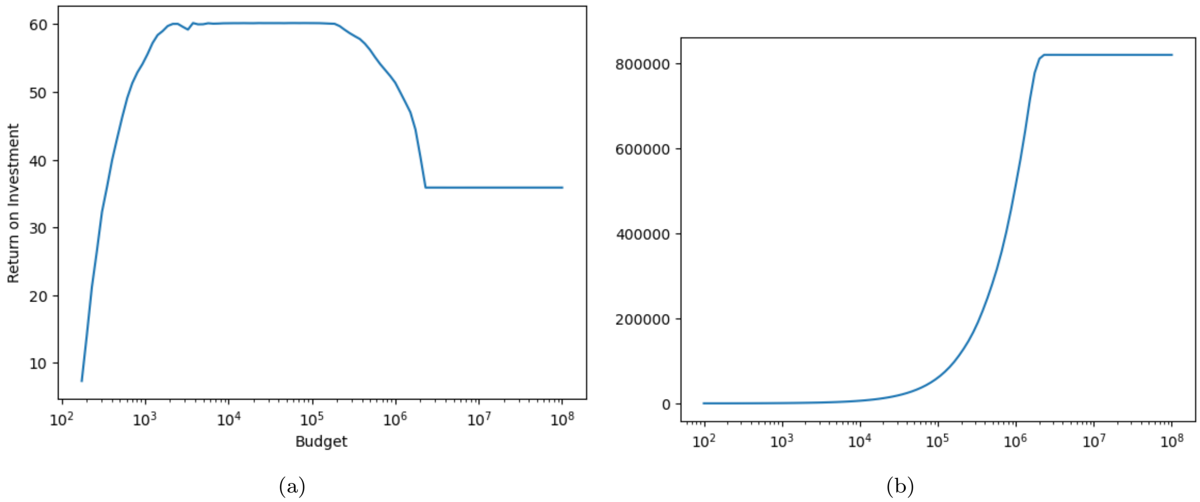


Figure 22: Return on investment (a) and profit (b) as we change our capital.

is no longer profitable.

The steady increase is due to the fact that as we increase our ability to stock higher margin products, our net margin overall naturally increases. However since our aim is to maximise raw profit and not a % return, eventually as we add more budget, the model will make the decision to stock up lower margin products as we have fulfilled demand for higher margin products already. This results in the overall margin decreasing.

If we examine our raw profit graph the pattern also makes sense in this context as well - it increases due to higher budget, and then caps out. Now what may matter the most is when the business is the most efficient under our model. We can see that the most gains are made when the amount of capital we have is relatively low, suggesting that it is difficult to get going with this kind of business under our assumptions, but increasing capital has sizable benefits.

9 Conclusions

Based on this model we can see that generally for a retailer, maximising profit intuitively leads to trying to find products which provide the highest margin, while using free shipping and discounts opportunistically to produce more customers while balancing the reduction in margin. Our model is particularly interesting when the resources are limited and knowing which combinations of products to stock to maximise profit is particularly important. However there is still work to be done to more realistically simulate a large retailer. To better simulate an actual retail market, one of the considerations we should make is considering number of customers or demand as an actual function of our decision variables. Right now we are treating the demand for each product as data fixed outside of our model but this may not be realistic as it causes us to simply consider the margin of products in a vaccuum. In reality retailers may have large amounts of products advertised on their stores in order to attract more people even if they may not necessarily buy a certain product specifically. More work needs to be done to model the expected amount of customers buying an item, perhaps via some sort of stochastic programming approach.

Another way to simulate real market conditions is to add commercial competitors. We considered the presence of commercial competitors (three similar stores) that had items similar to the store that we optimized at the beginning of the construction of our model. However, each of the different commodities of these four stores including our store has a different rating and price (in some cases, the rating or prices are the same). We want to infer based on this how many customers will go to each store, and how many customers who might otherwise go to our store will be diverted to other stores. And then through calculation and inference to optimize the price and quality of our products to attract more customers to our stores. However, in order to avoid too tedious of a calculation, we decided to reduce this construction in the model to a more simple situation. We introduced a factor to both discounts and free shipping to simulate the capture of more customers through offering a better deal to the customers. In further studies, we should reintroduce the commercial competitors as part of the model optimization consideration instead of just using a simple, idealized factor to estimate the number of customers who choose to purchase products from our store. At the same time, we should conduct more in-depth market research and use more data on commercial competitors for research and comparison. Further modeling can also be conducted on their selling price, buying price, commodity rating, and corresponding sales volume and profit.

Furthermore, another limitation is that time periods are not included in our model. Hence, we can consider including the time periods for improving our model in further development. For example, we can consider there are four periods for a year (since there are four quarters in a year, we can also consider different time periods such as half a year and so on), and then we can build a more realistic model. Firstly, based on the warehouse's inventory ceiling for a certain kind of goods, we can consider buying the products when the purchase price is low and then storing them in the warehouse, which is conducive to optimizing our profits. At the same time, because the amount of some items sold is strongly correlated with the quarter, some items will sell particularly well in some quarters and less well in others. Therefore, based on the historical data, we can add a new parameter related to the quarterly correlation. We can make further data analysis based on this and buy more of a certain commodity as inventory before the good period of sales, increasing our potential profit. We should also consider buying less of the product before a period of bad sales for it and try to clear inventory before that period. In the period of poor commodity sales, as discussed in the report, we can consider discounts and free shipping and other promotional activities, so these products in the period of bad sales can be sold better.

At the same time, considering the influence of time, we may need to add another parameter, the shelf life of the product, to control when we should buy the product and sell the batch of products in advance. We can also consider depreciation or obsolescence of products, and should try our best to control the goods to be sold within their expiration date otherwise they need to be discarded. To do so we can model promotional activities in the case that a certain product is approaching its expiration date but there is a lot of inventory so that their inventory can be cleared.

Despite these limitations we nevertheless believe that this kind of simulation is useful for a small retailer business as it comes up with a variety of possible scenarios. The time extension mentioned can be implemented as an extension of our current model, while properly simulating market conditions and customer demand may prove to be difficult.

References

[1] Suddharshan S. *Retail price optimization*. Jan. 2022. URL: `https://www.kaggle.com/datasets/suddharshan/retail-price-optimization?resource=download`.

[2] Shippo. *State of Shipping Report 2023*. 2023. URL: `https://try.goshippo.com/2023-state-of-shipping/`.

[3] https://www.bcg.com/about/people/experts/rob-souza. *The Elements of a Retail Business Model*. https://www.bcg.com elements-retail-business-model. Oct. 2020. (Visited on 05/19/2023).

[4] Francis J Mulhern and Daniel T Padgett. "The relationship between retail price promotions and regular price purchases". In: *Journal of Marketing* 59.4 (1995), pp. 83–90.

[5] Robert Fildes, Shaohui Ma, and Stephan Kolassa. "Retail forecasting: Research and practice". In: *International Journal of Forecasting* 38.4 (2022), pp. 1283–1318.

[6] Dhruv Grewal and Anne L Roggeveen. "Understanding retail experiences and customer journey management". In: *Journal of Retailing* 96.1 (2020), pp. 3–8.

[7] Mark Armstrong and Yongmin Chen. "Discount pricing". In: *Economic Inquiry* 58.4 (2020), pp. 1614–1627.

[8] Christopher Bliss. "A theory of retail pricing". In: *The Journal of Industrial Economics* (1988), pp. 375–391.

[9] Daniel Hosken and David Reiffen. "Patterns of retail price variation". In: *RAND Journal of Economics* (2004), pp. 128–146.