Superstore Final Project:

Insights and Implications using R-Studio

Silvia Ji, Koa Mahuna, Gianni Pink, Kitty Pham, and Jack Taylor

College of Business Administration, Loyola Marymount University

BSAN 6050: Customer Relationship Management

December 18, 2020

Summary

Summary

In the first part of the analysis, we aimed to identify the factors that drive Customer Lifetime Value, as well as Sales and Profit. We determined the two most significant factors to be shipping costs and discounts. Similar to Customer Lifetime Value, higher shipping costs can contribute to higher sales revenues per unit due to an increased price. For our second insight, we can begin by analyzing the frequency tables present in the second block of our report. These frequency tables give us information about the popularity of items and a proxy for the importance of certain product categories. Additionally these tables can help us determine which regions of our business should be primarily tended to. Discerning the value of a product is essential to constructing an appropriate promotional campaign. We need to be able to identify whether a product is beneficial to us due to the fact that it sells in high quantities or because it sells at a high price or some combination of those two. In the third block, we are using descriptive analysis and visualization to explore how customers are represented in each market and what we should be selling in each region. Based on our data, we discovered that the top 3 countries were China and Japan from North Asia, and Canada. After analyzing our most profitable countries, we found that the top three countries to invest in are Lithuania from the European Union, Sweden from Northern Europe, and Thailand from Southeast Asia. We determined that despite negative or low profits, we can be the top competitors in these countries and drive sales. In the fourth block we aimed to find ways of how we could improve customer service and company efficiency on our e-commerce platform through a recommendation system. We sought to accomplish building a recommendation system by conducting a market basket analysis in R. Such an analysis would tell us how frequently different products or categories are purchased together. We found that our most frequently purchased items are binders, storage, art, paper, chairs, phones, furnishings, accessories, labels, and envelopes. After exploring the results of our item matrix and conducting a frequent itemset generation, which is the process of finding all frequent item-sets with a support greater than or equal to a predetermined minimum support count, we then generated our rules. Lastly, we envisioned possible implications for each insight.

Descriptive Insights

Our first step is to simply observe the data and figure out what information we have available to us for analysis. After identifying the variables that we can consider we will be able to construct meaningful questions that can provide us with the ability to make predictions and suggestions for the business. During this initial observation phase we decided to identify the quantitative columns as these are the columns we can perform analysis on. The meaningful quantitative columns were: sales, quantity, discount, profit, and shipping cost. Additionally, we wanted to identify categorical columns, or the columns that we could use to slice our data into more specific and precise sections. The further we can isegment our data, the more helpful our insights will be. If we do not segment our data we will only be able to contrive very general and broad suggestions. Therefore, in order to be most useful we need to have distinct segments. The types of columns that we can use to segment our data include: segment, location, category of purchase, and personal customer identifiers.

By referencing some of the frequency tables we have created in our report, we can identify some trends and patterns in our data. For example, block 2 provides us with a number of useful frequency tables. We can see that the office supplies category is responsible for the highest quantity of sales. However, the technology category provides us with the most expensive items. Furthermore, although most of our sales take place in Central America; Asian regions are responsible for the most expensive basket purchases according to the average amount of money spent.

Modelling Based Insights

Introduction

The dataset provided to our group allowed us to derive a number of useful insights. These insights could be leveraged in order to make smarter business decisions and create a more profitable business. Additionally, a competitive advantage could be gained or increased by considering the dataset. For these reasons, it is extremely important that a manager understands what each block of our report means and how that information can be translated into actionable steps for their company. In order to facilitate and

assist this process, we have provided the following section of our report. This section will provide an analysis of the important points from each block of our report followed by examples of how these points can be translated into profit.

Insight 1

In block 1 of the analysis, we aimed to identify the factors that drive Customer Lifetime Value, as well as Sales and Profit. To accomplish this, we took a closer look at countries, regions, shipping costs, discounts, and profits. To further analyze potential business implications, we lastly performed a case study focusing on table sales, as it was our most profitable product category overall. To find out which factors were significant in driving Customer Lifetime Value, Sales and Profits, we closely studied their correlation to each other. Before calculating Customer Lifetime Value, we identified major differences within individuals in our customer base. Two of the main differences we found are the country a customer resides in, as well as the geographic region their country is located in.

Our first finding was that the country a customer resides in does in fact have an impact of their Customer Lifetime Value. However, as the business is active in 147 countries, we found it more feasible to work with regions for a broader focus. The active regions are Africa, Oceania, EMEA, North, West, Central, South, North Asia, Central Asia, Canada, Southeast Asia, East, and the Caribbean. Notice that some of the markets, such as EMEA and Africa, may overlap. However, it does not impact the outcome of the study as the goal of our analysis was to see which specific region or combination of regions had customers with the highest Customer Lifetime Value. We found customers in Canada and North Asia had particularly high Customer Lifetime Values when compared to customers who reside elsewhere. Countries in North Asia include, but are not limited to China, Korea, and Japan. Coincidentally, these three countries have also seen a recent upsurge in general consumerism with the emergence of a larger middle class in the past decades. In addition, international companies' market shares have also increased recently in these countries.

Besides geographical location, other factors that can affect Customer Lifetime Value include shipping costs and discounts. We found that the higher shipping costs increase Customer Lifetime Value, while discounts decrease Customer Lifetime Value. This is logical as these factors also contribute to higher profits per customer, which translates to a higher Customer Lifetime Value. Similarly, we confirmed that higher shipping costs contribute to higher sales revenues on average. This indicates that most customers are willing to purchase products regardless of their shipping costs.

To put our findings into perspective, we focused on our most profitable product segment, which are tables. We wanted to further explore which above mentioned factors specifically contributed to increased table sales. We again determined the two most significant factors to be shipping costs and discounts. Similar to Customer Lifetime Value, higher shipping costs can contribute to higher sales revenues per unit due to an increased price. However, a Harvard business school case cites that 44% of shopping cart abandonment is due to high shipping fees. Offering free shipping for certain order sizes can entice customers to purchase in greater quantities. Thus, we suggest obtaining data on shopping cart abandonment rates compared to shipping costs for further analysis. Discounts decrease the profits per unit sold, especially in Lithuania, Sweden, and Thailand as they decrease the price per unit. However, discounts are a good way to attract new customers that can turn into repeat customers, or attract repeat customers to buy other products in addition to the discounted item that they normally would not have

purchased. Thus, we recommend continuing to offer discounts, but also find strategies to offset profit losses incurred from markdowns.

Insight 2

For our second insight, we can begin by analyzing the frequency tables present in block 2 of our report. These frequency tables give us information about the popularity of items and a proxy for the importance of certain product categories. Additionally these tables can help us determine which regions of our business should be primarily tended to. Our tables tell us that we have a higher number of sales per product in our office supplies category. Also, we can see that within each category there are products driving the majority of sales. These products would be chairs, bookcases, binders, storage, appliances, and all of the technology products. These products sell in high quantities, at high prices, or both. Additional pieces of information can be gathered from our tables of regional differences. These tables can highlight high performing regions for us such as the high number of sales present in Central America or the high average sale price of Asia. Finally, we can look at profit by year in order to further deepen our understanding of how our profit varies from year to year. We can note that 2013 averaged higher profit than any other year despite the fact that 2014 saw greater sales. There was a much smaller increase from 2011 to 2012 than there was from 2012 to 2013. Then, following the high profits in 2013 we see a slight decline in profits in the following year.This disparity would need to be addressed in greater depth in order to glean more useful information.

```
# Groups:   Category [3]
   Category        Sub.Category mean.Sales     N
   <chr>           <chr>             <dbl> <int>
 1 Furniture       Bookcases         608.   2411
 2 Furniture       Chairs            437.   3434
 3 Furniture       Furnishings       122.   3170
 4 Furniture       Tables            879.    861
 5 Office Supplies Appliances        576.   1755
 6 Office Supplies Art                76.2  4883
 7 Office Supplies Binders            75.1  6152
 8 Office Supplies Envelopes          70.2  2435
 9 Office Supplies Fasteners          34.4  2420
10 Office Supplies Labels             28.2  2606
11 Office Supplies Paper              69.0  3538
12 Office Supplies Storage           223.   5059
13 Office Supplies Supplies          100.   2425
14 Technology      Accessories       244.   3075
15 Technology      Copiers           679.   2223
16 Technology      Machines          524.   1486
17 Technology      Phones            508.   3357
```

```
   Region         mean.Sales     N
   <chr>               <dbl> <int>
 1 Africa              171.   4587
 2 Canada              174.    384
 3 Caribbean           192.   1690
 4 Central             254.  11117
 5 Central Asia        368.   2048
 6 East                238.   2848
 7 EMEA                160.   5029
 8 North               261.   4785
 9 North Asia          363.   2338
10 Oceania             316.   3487
11 South               241.   6645
12 Southeast Asia      283.   3129
13 West                226.   3203
```

Discerning the value of a product is essential to constructing an appropriate promotional campaign. We need to be able to identify whether a product is beneficial to us due to the fact that it sells in high quantities or because it sells at a high price or some combination of those two. There are some products that fall outside of these categories as well. For example, a product that we sell at a loss may seem like it is not something we should keep however, if it is responsible for lots of foot traffic to the store it may be worth it. This additional foot traffic may generate more sales on other goods as well. The profit from these extraneous sales may cover the loss we take on that good. In cases like these, we would want to hold on to the unprofitable good because it brings us money in other ways.

Greater attention to detail will result in greater customer satisfaction. This is because as we can further identify the characteristics of a good we can more easily customize a promotional plan. This, in turn, will yield greater ease of use for our consumers. Once we are clear on what our products do for us; we can better serve our customers. The additional information helps us paint a more complete picture of what impacts sales and customers' preference. This relationship is not causal however, so it is important to understand that we have not accounted for all variables that may impact this relationship. This omitted variable bias will be responsible for some variation in our relationships. With this in mind, we should attempt to gather more data in order to further flesh out this insight. As we increase the amount of data we have, we can better understand how this interaction occurs. Additionally, as we have more pieces of the puzzle we will be able to better serve our customers and more effectively promote our products.

Insight 3

In Block 3, we are using descriptive analysis and visualization to explore how customers are represented in each market and what we should be selling in each region. The regions that we will be focusing on are Canada, Northern Europe, the European Union, North Asia, and Southeast Asia. Based on

our data, we discovered that the top 3 countries were China and Japan from North Asia, and Canada. We are comfortable in these countries, but we wish to understand our position better so we can customize our products.

In China, we perform well in the technology segment. Within that category, copiers and phones are the most profitable. In Japan, we also see similar results. We excel in the Technology category, which copiers and phones are also the most profitable. Machines and Accessories generally don't perform well in the most profitable countries in Asia. If we must downsize, the two sub-categories to continue to operate and improve in are phones and copiers for China and Japan. Moving onto Canada, we see great potential in both office supplies and technology. Storage and Appliances are the most profitable. We should communicate with the R&D team to continue operating and improving these sub-categories. Throughout the years, our phones and copiers are the best performers. The top performing countries suggest that overall, phones and copiers establish our global presence.

After analyzing our most profitable countries, we found that the top three countries to invest in are Lithuania from the European Union, Sweden from Northern Europe, and Thailand from Southeast Asia. We determined that despite negative or low profits, we can be the top competitors in these countries and drive sales. We will also go into more details with these investment opportunities.

The first country that we have found an investment potential in is Lithuania. Its GDP per capita is the highest in the Baltic states. Lithuania's economy (GDP) grew more than 500 percent since 1990. We are currently not performing well in Lithuania for the last few years. One segment that we can have our R&D team to investigate is Office Supplies. Within this category, supplies, labels, and envelopes can be greatly improved. We should become a specialized market in this area or increase our offering throughout the corporations of Lithuania. We should also review our supply chain to cut costs. Right now is the best time to improve our presence in this country.

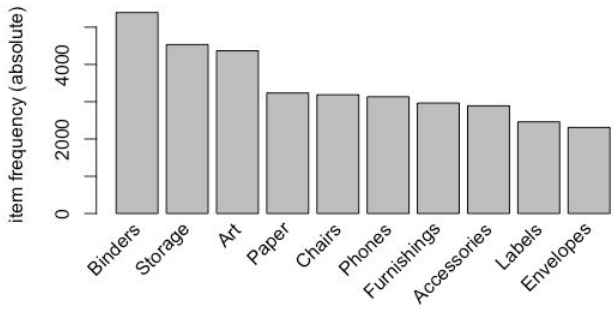
The next country is Sweden. It is ranked number 17 in the world in terms of GDP per capita. Our offering in Sweden needs to be improved greatly in order to compete with strong competitors. We should

focus on the technology segment. Despite negative profits, we discovered that Machines have the strongest investment potential. We need to analyze why this sub-category is not generating positive profits right now, who our competitors are, and what we can do to improve our image for retailers selling our machines. With high disposable income and great businesses, we should capture this market in Sweden.

Lastly, we want to look at Thailand in Southeast Asia. Thailand is a developing country. It is an interesting case, as we can both manufacture our products and sell our products in this country. Known for their technology sector, the Thai market is also receptive to our technological offerings. Our profit can greatly improve in Thailand. Looking at the years individually and together, the phone sub-category has continuously outperformed others. We think that the other subcategories are lowering our profit. We should reconsider offering them in this market. Overall, we can greatly improve our supply chain and products in these countries. Their economy has great potential in the areas that we wish to improve in.

Insight 4

In block 4 we aimed to find ways of how we could improve customer service and company efficiency on our e-commerce platform through a recommendation system. We sought to accomplish building a recommendation system by conducting a market basket analysis in R. Such an analysis would tell us how frequently different products or categories are purchased together. Exploring the dataset through R, we found that we offer exactly 17 unique subcategories of products. Using R also allowed us to convert our data into a transactional format. We did so specifically by extracting the necessary columns (order id and sub-category) into a csv and then transformed them using the read.transactions function. Doing so led to our first insight. We found that our most frequently purchased items are binders, storage, art, paper, chairs, phones, furnishings, accessories, labels, and envelopes.



This gave us an indication of what we would likely find in our basket output, and could be useful for business decisions, which will be discussed later. Through the first part of our model, we were also able to discern that most of the time, our customers will only buy one or two subcategories of products for

```
#element (itemset/transaction) length distribution:
  #sizes (sub-category)
#    1     2     3     4     5     6     7     8     9    10    11
#12800  6469  3193  1484   626   304   101    42     9     5     2
```

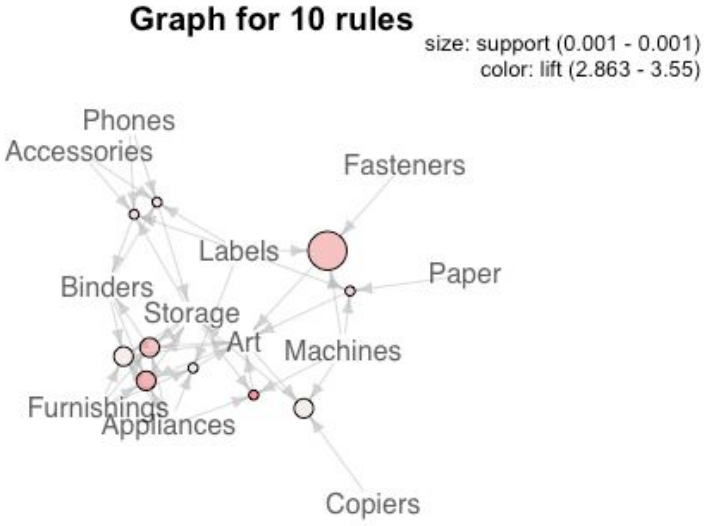
each transaction.

A final discovery made from our initial insights of our basket analysis is that our item matrix has a density of approximately 0.1113805. In this model, density, which can also be labeled as support, is interpreted as the fraction of transactions that contains two items. So in our case, a set of items is purchased on average approximately .111 of the time. Our low average support level can actually be explained by the previous insight that customers mostly buy 1 or 2 sub-categories per transaction. We only have 17 items, so many transactions and baskets occur, which doesn't bode well for our analysis, as this indicates that on average the fraction of transactions that contains two products is low. Having said that, we are still very confident in our future recommendation system due to the insights found in the second part of our model.

Insight 5

After exploring the results of our item matrix and conducting a frequent itemset generation, which

is the process of finding all frequent item-sets with a support greater than or equal to a predetermined minimum support count, we then generated our rules. Rule generation lists all rules from our frequent item sets based on our predetermined support and confidence levels. The model eliminates rules that fail to meet our minimum thresholds. We use the density found in the previous insight to determine the support threshold we use for mining. After experimenting through trial and error, we decided to use a support of .0005. Because our support level is so low, we choose a confidence level of .50. Confidence can be interpreted as how often one item appears in transactions with another. By setting our confidence at .50, we eliminate or correct the issue of having a low support level as best as possible. Our model generated 22 rules, which we then mapped out according to rules with the 10 highest lifts. For reference, lift is the amount our confidence increases that one item will be purchased given that another was purchased; it is the most important metric in market basket analysis, and will play a key role in the business implications we gather from our findings.



Larger circles indicate higher support levels and darker color indicates higher lift levels. Looking at our plot, we established that our top baskets showed that when appliances, binders and furnishings are bought together, storage will be purchased; when appliances, art, and furnishings are purchased together,

binders will be purchased; when fasteners, machines, and labels are bought together, art will be purchased; and when copiers, machines, and art are bought together, storage will be purchased. These baskets are all within lift range of 2.863 - 3.55, so we are very confident we can make these recommendations when customers have these left handed products in their baskets.

Managerial Implications

Implication of Insight 1

A manager can leverage the information we gained from block 2 in order to increase their profitability as well. Firstly, they can identify the difference between the best selling items and the most expensive products. Technology products sell for the highest price which opens them up to larger profit margins but office supplies sell in much higher quantities. Therefore, a manager could acknowledge two distinct approaches to increasing their revenue and profit numbers. One solution would be to attempt to sell technology products in higher numbers. This could be achieved by either running promotional sales on the products or by increasing advertising of the products. Looking even deeper into the technology category, you will find that phones are the most profitable category. Perhaps a manager would like to pursue greater market share in the office phone's business. If you wanted to use a quantity of sales approach you could focus on moving more office supplies quickly. For instance, we find that the best selling product in office supplies is binders and the slowest selling product is fasteners. Now we can offer a bundle deal that includes binders, pens, and fasteners. This deal will facilitate an increase in the number of fastener sales as people buy the bundle because they wanted the binders and pens. A final section we can analyze in regard to block 2 is how profit varies by year. 2013 earned higher profits than its successive year despite having lower sales. A manager would be wise to discover what drove such high profits in 2013 and attempt to emulate the promotions or pricing strategies that were implemented in 2013. These frequency tables contain valuable information that can inform a manager on how to best allocate resources. These are just one of the many sources we can draw on from the report.

Implication of Insight 2

To continue, we will dive deeper into how we can leverage our 2nd insight to inform business decisions. As we discussed previously, we have identified some key regions for our business. These regions include Europe and North Asia. These areas are important to us because we have strong footholds in both of these markets. We have a consistent revenue stream and loyal customer base. Furthermore, these regions have great potential for growth in our most expensive sector which is technology. Given their populations and affinity for technology, we see an opportunity to increase the number of technology products sold in China and Japan. In Canada we'd like to focus on storage and appliances as they are the most profitable sectors. This information can allow us to focus marketing campaigns about technology to these regions or embark on long-term promotional strategies to gather a greater market share in these regions. Understanding where we perform well is essential to long-term growth and success. We need to maintain our competitive edge by emphasizing on areas where we are currently proficient. Eventually, we hope to become dominant in these sectors in each region. Another highlight from this section is our discussion of countries that we deem to have a high potential for future sales. In similar fashion to our discussion of China, Japan, and Canada. The countries we have highlighted are investments we deem to be smart. We believe the upfront cost could pay for itself if we manage to become more established in these areas. We should target these areas with future plans for expansion and begin to develop marketing strategies for these regions.

Implication of Insight 3

The findings from block 1 can be leveraged by managers to further increase Customer Lifetime Value and sales profit by analyzing the results. These insights can resolve existing problems as well as prevent future issues that may occur to improve the business' financial performance. As we determined the regions that have the Customer Lifetime Value, we can also point out the regions with the lowest overall Customer Lifetime Values. The most significant ones are the Northern, Central, and Eastern regions, as well as Oceania. While regions such as Oceania may have lower Customer Lifetime Values

due to accessibility and logistical issues, the business may increase promotional efforts in other regions. One recommendation is to offer paid memberships in exchange for free shipping, higher order priorities, and special discounts to increase sales per customer and thus Customer Lifetime Values. We suggest that the membership should be paid annually to lock in customers over a longer period of time, instead of collecting monthly payments as it more likely that they may cancel. Another recommendation is to allocate a greater portion of the company's marketing budget to these regions in order to increase exposure and target potential new customers. Lastly, the business may offer products that are especially appealing to customers in the above regions to strategically increase sales revenues.

In terms of increased sales revenue and profits, our first recommendation is to obtain rates on shopping cart abandonment to determine if higher shipping costs contribute to more forgone purchases. Managers should find strategies to offset profit losses from offering discounts, for example by implementing results found in the market basket analysis. For instance, managers may offer bundles, which are sold at a higher price than single items and thus generate higher profits. Second, managers may sell select products at a higher markup than other similar products, and periodically offer discounts on them by reducing their price to the same as other comparable products to make them appear more attractive. This will offer an incentive for customers to purchase them before they return to their regular price due to their fear of missing out on a deal. In some instances, customers may even purchase multiple quantities to take advantage of the apparent deal.

Implication of Insights 4 & 5

The findings in block 4 can be leveraged and implemented in many aspects of the business. Most e-commerce based companies use recommendation systems. As we learned in the Customer Analytics at Bigbasket case, "Amazon earned 35% of its revenue through its product recommendations." Outside of improving the profitability of the company, we can also use a recommendation system to improve customer service and shopping experience, which is especially important during the Covid-19 pandemic. Basket analysis can also be used to further improve pricing strategies, as we can use the rules based plot

in the creation of smarter bundle discounts, which will make the customer feel as though they are saving more and getting a higher value for their purchase. And in terms of retaining a higher profit margin for the business, we specifically recommend discounting the lowest priced item associated with each rule.

Looking forward in the long-term, once stores around the world are able to safely operate at full capacity, we can continue to utilize the results from our basket analysis through smarter product placement. Putting products from left hand baskets near or in the same location as right hand basket products will further improve the customers shopping experience. For example, placing binders, appliances, furnishings, and storage near one another may not have made sense at first, but our data mining helped us determine this to actually be winning a combination. Thus, data mining through basket analysis has also helped us improve the layout of our in-person stores as well as our e-commerce platform.

References

Abraham, Paul, Manaranjan Pradhan, Lakshimarayanan, Ganesh Iyer, and U Dinesh Kumar. "Customer Analytics At Bigbasket - Product Recommendations" *CRM Analytics 2020*. *Harvard Business Publishing*, https://hbsp.harvard.edu/download?url=%2Fcourses%2F754168%2Fitems%2FIMB573-PDF-ENG%2Fcontent&metadata=e30%3D. Accessed 12/16/2020.

Chen, C. and Donald N. (2018). Shipping Fees and Product Assortment in Online Retail. *Harvard Business School Working Paper 19-034*. https://www.hbs.edu/faculty/Publication%20Files/19-034_b2382177-a462-447e-86f8-690d1ea7af18.pdf