Medtronic Report

Linda Amaduzzi, Stephen McCarthy, Tran To Anh

University College Maastricht

Tran To Anh	i6140544
Linda Amaduzzi	i6152941
Stephen McCarthy	i6129703
Faculty:	University College Maastricht
Course:	Data Analytics
Tutor:	Peeters, Marianne
Tutorial Group:	01
Date:	May 9, 2018
Academic Year:	2017/18
Place:	Maastricht
Word Count:	2462 (excluding Appendix)

Table of Contents

Medtronic

Executive Summary

Our initial goal:

       To identify groups of interest among the customers, find interesting information through segmentation modelling, and to look for opportunities for upselling. Due to complications (see limitations section), we repositioned our goal.
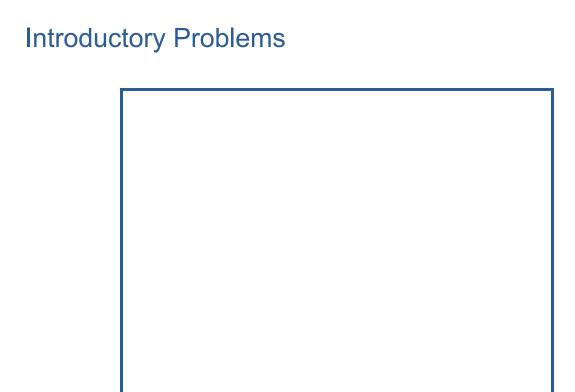
Our re-positioned goal:

       To identify groups of interest and if possible overcome limitations to achieve left out parts from initial goal.

Summary of Results:

       We identified three groups and a sub-group; Loyal Customers, Active Customers, 'Whales', and 'Islands'. Loyal Customers make purchases consistently. Active Customers made purchases within the last two fiscal quarters. Whales are customers that spend the most relative to other customers. Islands are easily identifiable Whales in low-sales countries. Top Customers within the groups are shown in some figures, and our process in identifying the groups can be found in the Approach, Overview, and Appendix.

Tentative Results-based Suggestions:

- Loyal Customers are kept track of (i.e positively flagged in some way), to see what they are buying, and make note if any of them drop out of the Loyal Customers group.
- Promote heavily to Active Customers with targeted Marketing, if possible find out why they have recently chosen or have continued to choose Medtronic as their provider of Medical equipment.
- Investigate what 'Whales' are buying, why they make such huge purchases in comparison to other customers, and if a steady relationship can be fostered to keep them cashing in.
- Similarly to 'Whales', investigate why Islands make large purchases when other customers of the region don't. Capitalize on any holes in marketing or availability there may be in that region.

Introductory Problems

Our preliminary introduction to the Medtronic case was through a lecture given by Marcel Krouweel (2018). This slide, taken from his lecture, indicates the four areas of interest to potentially address in this report. Our tutor advised that it was better to have a limited scope and drill deeper into it rather than answer all of them superficially. Therefore, our scope will largely focus on **Interest groups** and some **Segmentation** based on sales quantity.

Approach

Key steps in our approach are as follows:
1. Data Cleaning and Preprocessing
2. Data Modeling and Interpretation

1. Data Cleaning and Preprocessing

In this step, we examined the variables in the Medtronic data, looking out for typos in field names as well as renaming those that are not intuitive to us at first glance. For example, we renamed the field 'Mpg', which indicates the ID code of a certain Medtronic product, as 'Product Code'. In addition, we also filtered out redundant information in the dataset; specifically, we did not consider data from the fiscal year 2016, as well as that from May of the fiscal year 2017, as they have been clarified by Medtronic representatives not to add value to our analysis.

An important step in this preliminary step in our approach is the proper parsing of datetime variables in the Medtronic dataset. As a matter of fact, we were given information on fiscal years, quarters as well as months. However, in order to show, for example, sales trends over time, or conduct time series analyses, a proper datetime variable that captures the time aspect of the data is necessary. Particularly, converting the given fiscal time to calendar time is of great interest. Taking into account the fact that, at Medtronic, the fiscal year starts in May, we performed the appropriate datetime formatting and conversion in Tableau, to derive the Calendar Date variable which we could use for further analyses.

Lastly, we examined the sole numeric variable in the Medtronic data, namely 'Sales Qty', which denotes the quantity of sales for each customer. Specifically, we considered the skewness of this variable; the larger the skewness, i.e the more significantly the variable departs from normality, the less accurate our predictive analysis may be. Therefore, we attempted to correct the skewness, which was originally equal to 103.622 (Figure 1), by applying a logarithmic transformation. This linear transformation helped to alleviate the degree of skewness, which became 0.420 (Figure 2). This indicates a great improvement in the skewness, and that this continuous variable now has a distribution that is more similar to the ideal normal distribution.

Figure 1. Descriptive statistics of 'Sales Qty' in the Medtronic data

Figure 2. Descriptive statistics of 'Sales Qty' after the logarithmic transformation. The value of the skewness is displayed in the colored box.

2. Data Modeling and Interpretation

To create our interest groups, we first thought, "Medtronic is a medical equipment supplier, therefore the customers are mostly likely hospitals, clinics, and universities (for research purposes)". However, customers have been anonymised through their IDs and location data has been limited to country of purchase. Since we cannot tell which purchases are from hospitals, which are from clinics, which are from nursing homes or hospices, and which are from universities, we decided to take a look at the variables available. What immediately came to mind was sales data. So we shifted from our initial idea to segment by priori to instead segment by value ('Sales Qty'). While we cannot identify the customer types based on use/function, we can identify based on sales values, sales quantities, consistent purchases, and so on. Through Tableau and SPSS Modeler, we investigated and created different interest groups and segmentations.

Groups of Interest (with some segmentation)

Loyal Customers

Loyal Customers are customers that consistently make purchases throughout the fiscal year.

Loyal Customers are found through association between the customer id, sales qty, and date variables.

Active Customers

Customers that have made purchases within the last 6 months (two fiscal quarters)

Association between Customer ID, Sales qty, and filtered date to last 6 months.

Whales

There is a demographic in the mobile games industry that can independently hold a whole company afloat - whales. Whales are customers that spend more than regular customers, and do so fairly consistently.

Whales are the top customers overall. Another interpretation is above-average spenders.

Islands

A subset of whales in a country that does not spend a lot.

Top Customers in Countries where the sales are lowest.

Overview

Figure 2.5 Overview Dashboard

As can be seen in the Cumulative Sales graph, the orange line indicates the running sum of sales quantity, which grows steadily, while the blue line is the percentage difference in sales quantity between different calendar months: we notice a spike in the sales around October 2016 and another one in July 2017. The months of January and April also constitute an upward trend; therefore, it could be useful to look into what happened then more in depth. Beside the Cumulative Sales we have the Top 5 Customers, Products, and Countries measured by total volume of sales. At the top banner we see an overview of the sales performance of the company over the two years of the data: we can easily assess the growth of the sales as well as important sales statistics, such as average sales, minimum sales and maximum sales. There is not much difference between the two years of data, in terms of these statistics.

LOYAL CUSTOMERS

Figure 3. Top 20 loyal customers

Figure 3 shows the top 20 loyal customers of Medtronic medical equipment. Here, a metric for customer loyalty is frequency of purchase; we define loyal customers as those who purchase a Medtronic product at least every two months. The number ones in the table denotes the number of months that have passed since the customer's last purchase. In other words, from the table alone, we can already recognize the customers who purchase Medtronic products every month.

ACTIVE CUSTOMERS

Figure 4. Overview of active customers in the past 6 months

Figure 4 is a snippet of the list of customers who have been active in the past 6 months, i.e those who have purchase something from Medtronic within this time frame. These customers have been sorted in descending order of their total sales volume.

In total, there are 49214 active customers within the past 6 months (i.e up to the the latest date in the data), excluding those with negative or zero sales.

Figure 5. The five customers shown above are the ones that have bought the most overall.

The top 5 customers from each country can be found in the appendix and by looking there, one can also clearly see the whales standing out from the pack.

Figure 6. Top 20 'islands'

Figure 6 shows the top 'islands', i.e customers who are biggest spenders in countries who rank the lowest in terms of total sales volume.

Discussion

Limitations and Confounding Factors

There is a shortcoming of the figure 2. transformation: it is applicable only to nonnegative values. Since the original sales data has negative values (e.g the minimum sales qty value is -$1072800.00), this transformation might distort the original data and gives rise to misinterpretation. An alternative transformation technique we could opt for is Box-Cox Transformation (Box et al, 1964), which incorporates a shift variable and is a common way to transform non-normal dependent variables into a normal shape.

With the interest groups and segmentation,

Loyal Customers: as a sales-over-time group, it was limited to the scope of the two fiscal years given to us in the data. Possible confounding factors include; the durability of the equipment and how often it needs to be replaced or upgraded. We also have no way of knowing the reasons for whether a customer that was previously a loyal customer (prior to the 2016 fiscal year for example) but is no longer one. We have no data on whether they switched to a competitor, if all of their equipment doesn't need replacing so there is no need for new purchases, if the customer no longer exists (the hospital closed for example), or other reasons unknown.

Active Customers: As with the Loyal Customers group, this group's available data is limited to the scope of the last two fiscal years. With less data over time available, we have had to limit the definition of an active customer to one that has made purchases in the last two fiscal quarters.

Whales: One possible confounding factor is the price of the products. It is hard to decide whether a customer that bought one very expensive product vs customers that consistently bought expensive products can both be whales. If the value/qty of the output is all that matters, then it's safe to say both are whales.

Islands: Whales are big spenders and islands are supposed to be big-spenders relative to the average of their country, but not all islands are whales when relative to all customers. We still think it would be interesting to know who the islands are in order to better understand the market in their region, but they should not influence europe-wide decisions.

Segmentation: With SPSS, we identified some association among different products and business lines.

Figure 7. Association Rules

In the top 10 most interesting rules by confidence, we can identify certain products that are closely associated with certain Medtronic business lines, and also in certain countries (e.g Germany). Although there is association with certain timeframes, since the data has only 2 years under consideration, this aspect might not be meaningful to be looked into. Nonetheless, the relationship between products and business lines is interesting, and warrants further scrutiny.

With the modelling,
For some reason, SPSS Modeler would not read the product code variable in our clustering model, which was unfortunate. We had issues with producing an accurate regression model that predicts the sales quantity because there were negative values in the dataset; coupled with significant skewness (i.e departing from normality), our attempts at OLS regression were less than meaningful. In sales data, negative sales volumes can be explained by loss-leaders or other factors, but we tackled this issue by using linear transformations that work with even negative data, such as the Box-Cox transformation. We also worked with the original data which was not normalized especially for classification and clustering models.

What would have been useful…

A couple things to improve the accuracy of the Loyal Customers group would be to include data from earlier fiscal years and if there was a way to know whether the customer has switched to other suppliers or not. It would be possible if if Medtronic was a 'as a service' company or if there was a retailer/online service that sold Medtronic equipment along with its competitors, then it would be possible to know if a customer showed interest in Medtronic or not and did not make a purchase, or did the same with competitors and bought from Medtronic. We also do not know how durable the equipment sold is and how often it needs to be replaced on average. Given data from a larger span of time, we could potentially track how often the same product has been bought from a customer, and if we knew which products were similar, we could know if new purchases were upgrades or not.

It would have been useful to know which products were similar to each-other in function (different models of a type of equipment for example). Since we don't know how durable the products are, and which ones are designed to be sold en-masse repeatedly, it makes interpreting certain sales trends on specific product-IDs too abstract so we didn't include it.

It also would have been useful to be able to drill deeper than countries into the geographical data (whether provinces, cities, whatever), especially with the larger countries such as Germany and France.

Medtronic's sales quotas, targets, as well as profits for the Fiscal Quarters and Years could have also been interesting information to have.

Suggestions

1. Look at the countries that have 'Islands' customers: why are these buyers isolated? The surroundings of them should be analysed to understand what makes them a special case: if their reasons to be with Medtronic are valid ones, maybe other customers can be found with a metric founded upon their reasoning within the same country and similar ones.
2. Organize sales personnel taking into account the Loyal Customers: these should be handled by the same person over time, so that the relationship with Medtronic gets stronger and so that the salesperson knows the necessities of the client and their special needs.
3. Loyal Customers are kept track of (i.e positively flagged in some way), to see what they are buying, and make note if any of them drop out of the Loyal Customers group.
4. Promote heavily to Active Customers with targeted Marketing, if possible find out why they have recently chosen or have continued to choose Medtronic as their provider of Medical equipment.
5. Investigate what 'Whales' are buying, why they make such huge purchases in comparison to other customers, and if a steady relationship can be fostered to keep them cashing in.

Citations

Krouweel, M. (2018). MEDTRONIC.

Box, George E. P.; Cox, D. R. (1964). "An analysis of transformations". Journal of the Royal Statistical Society, Series B. 26 (2): 211–252. JSTOR 2984418. MR 0192611.

Appendix

Our step-by-step process:

Medtronic provided us with four sales_dump files which were quickly stitched together in a union through Tableau. The sales_dump files contained some useful variables for our report such as Customer (anonymized, just a code), Product Code, Sales Quantity, Fiscal Dates, and more that will be elaborated upon later. There were some variables we would have liked to see for the purpose of customer segmentation and a couple limitations with specific variables (such as the geographic information only going as deep as country), but we will cover that in the *Discussion*.

Figure ___.
Here are the Top 5
Customers from each
country.