Sarah Pritzker, Rabindra Tambyraja

Paul Mcgie, Christopher Stevenson &

Rob Dirks

Business Intelligence

IDSC 6444

3.9.2015

Data Mining in The Electronics Business: Searching for Answers; Sometimes Finding Questions

**Introduction**

In order to better understand how the skills learned in Business Intelligence apply to our professional lives, our group wanted to explore data we encounter every day at our jobs. While we come from a variety of industries and backgrounds, data pulled from a Fortune 500 electronics vendor, proved not only the most accessible with regards to acquiring comprehensive data without privacy limitations, but also the most interesting to our group collectively. But to understand our data comprehensively and to successfully employ our business analytics toolset, we thought it necessary to specify our data further. We chose to hone in on data surrounding the online resale electronics website operated by this larger electronics retailer. Despite the selectivity in the dataset chosen, we still encountered some limitations in our data collection. Because we chose to pursue a dataset generated and supervised by a business, some of the data we may have liked to see was limited by confidentiality. Nonetheless, we feel that we were able to assemble a data set that provided ample space for exploration and ultimately demonstrated our mastery of business intelligence subject matter.

**Context**

As mentioned above, our group chose to explore sales data gathered from an electronic re-sales website. The site is operated by its parent company, a large electronics retailer that generated over 45 million dollars in revenues in 2014[[1]](#footnote-1). This retail website, operated as subsidiary brand, provides a reputable sales channel for new, pre-owned and refurbished electronic goods directly to customers.

**Goals**

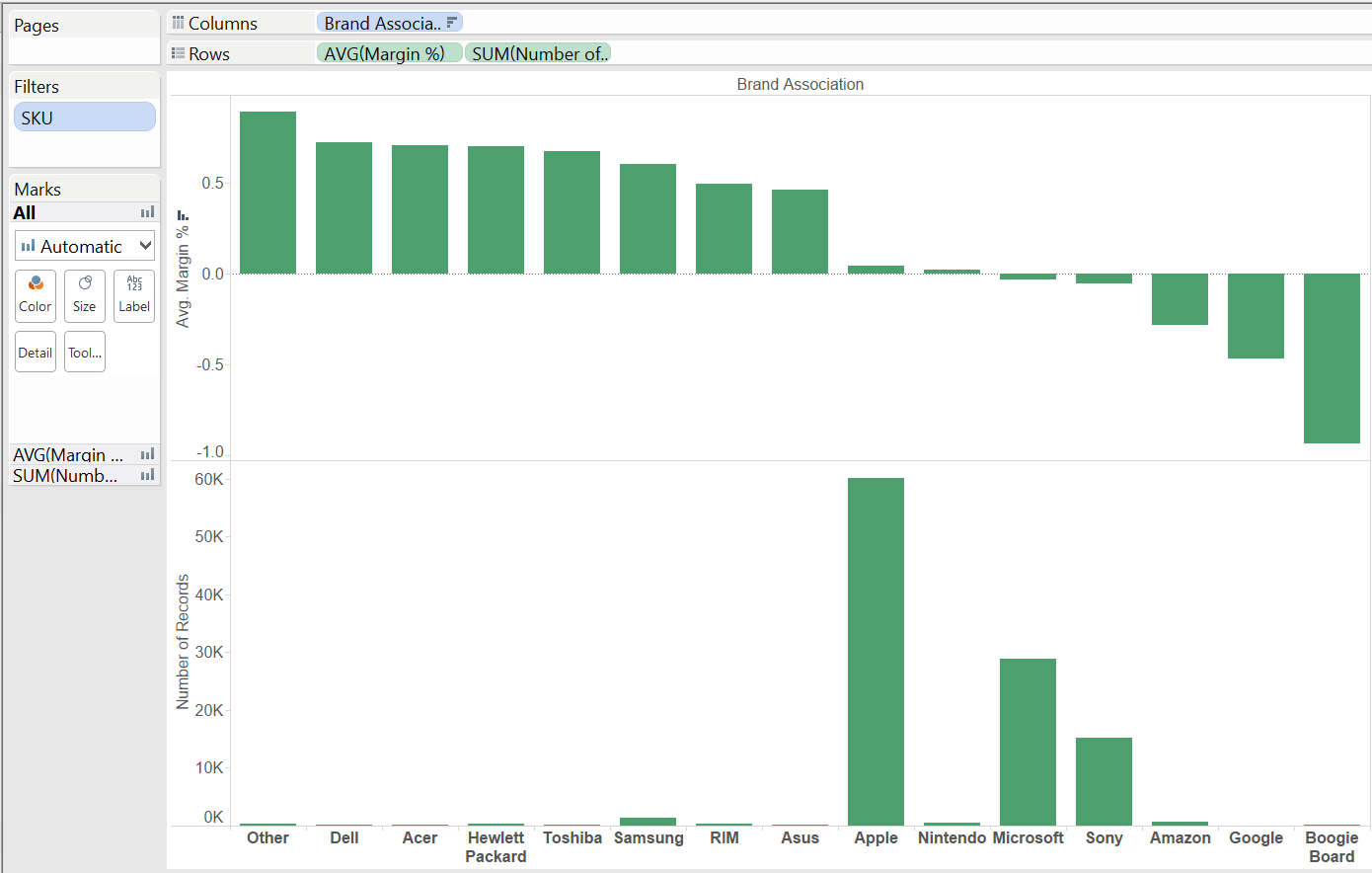
Ultimately, our group wanted to understand what factors best predicted a successful resale. After an initial analysis of our data set, we determined that profit margin per product served as the best indicator for successful resale. With that in mind we defined several questions that would help us begin to define what factors served as the most salient predictors of profit margin.

These questions, defined as our Phase 1 investigation asked the following:

1. How does brand affect product margin?
2. What is the item with the highest profit margin? With the lowest profit margin?
3. How does processing speed contribute to product margin?
4. How does product condition affect our margins? Is their correlation between product condition and product margins?
5. How does processing speed affect our margins? Do products with more computing power generate better margins?
6. What effect does turn-around time have on profit margin? What about sales volume?
7. What time of day is our sales volume highest? What day of the week is our sales volume highest? Which items are selling then?
8. Are their products that we shouldn’t be reselling at all, i.e. where we are losing money on the resale?

**Answering Our Questions**

*Understanding Product Margins*

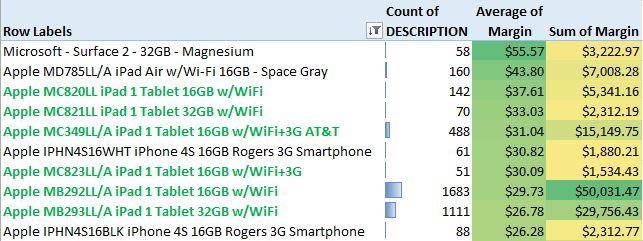
To get a better understanding of what was occurring in our data set, we first broke down the data to look at how brand affected sales volume and profit margin percent. We used profit margin as a percent in place of actual profit margin in an attempt to compare the data more equally. As price and margin both vary widely between products, the percentage profit margin served as a more understandable explanation of our data:

The graphic above shows volume of sales and profit margin percentage by brand association. It is notable that the highest margin brands occur in very low volumes. The highest margin percentage brand is actually the “Other” category; a hodgepodge of cell phones and low-end laptops. Volume on the other hand is clearly driven by three brands: Apple, Microsoft, and Sony. Furthering this, the data shown *exclude* those with fewer than 100 sales. Lastly, the low margin on Boogie Boards suggests that perhaps these products should not be accepted for resale at all. Ultimately if a manager were to look at this data and this data alone, one would think that this operation should reduce the number of products they accept for resale.

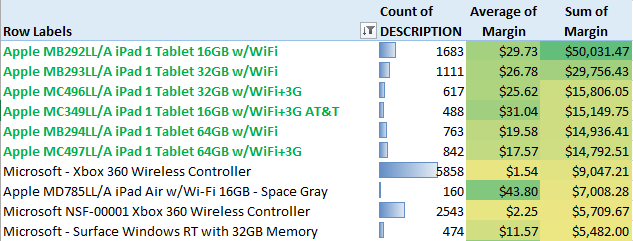
In order to distill this information even further, we wanted to explore how the type of product affected the product margin of a sale. In order to this, we next asked the question: on which products are our margins highest? Which products have the lowest margins?

We first approached this problem in the most simplistic way: we sorted by margin and discovered that our most profitable item was an **Apple MacBook Pro 15.4in Laptop i7 2GHz 8GB 500GB DVDRW WiFi** that was sold for $825 more than the purchasing price. But this single item (amongst 122,000+) was an outlier. What we needed was information that was helpful and actionable for this business; something an outlier cannot provide.

To get a more explanatory result, we began by filtering out all items that were bought and sold less than 50 times. This is still a relatively low threshold, but large enough to potentially have meaningful data. With this threshold in place, the highest average margin item is the **Microsoft – Surface 2 – 32GB – Magnesium**. This item’s average margin was $55.57. However this item still sells in relatively low volume and when we look at the Sum of Margin column, its total impact is still fairly minimal.

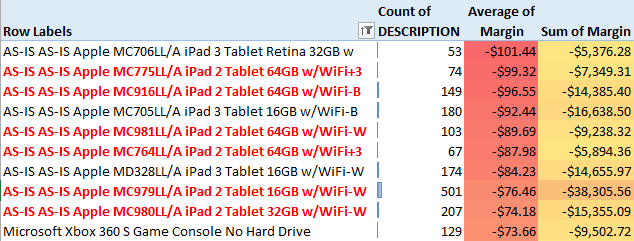


The list above shows the ten most profitable items (based on Average Margin) that were bought and sold – Apple’s iPad 1 shows up six times in various different models/colors, 60% of the list. This is interesting and potentially useful information.

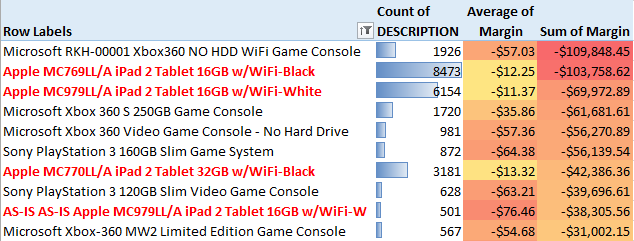


The list above shows the 10 most profitable items, but is based on the total margin (sum) rather than the average. Again we see iPad 1 dominating the list, which indicates there is definitely a market for the older model iPads. Sellers are willing to part with them for an attractive price and they’re able to be profitably re-sold to buyers.

We next applied the same processes to the lowest margin items. Using the same threshold of 50x or more buy/sell exchanges, the least profitable item (by average margin), was the **AS-IS AS-IS Apple MC706LL/A iPad 3 Tablet Retina 32GB**.

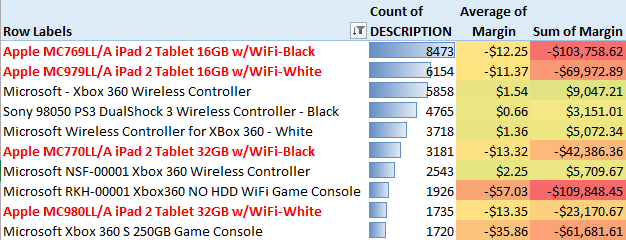


Ironically, this is almost the exact same situation happening on the highest margin end of this list, but with iPad 2. In the top 10 items, 60% of them are iPad 2[[2]](#footnote-2). Again, none of these counts are too large, thus they don’t sum to a major impact to margin. Sorting again by total margin, but for the least profitable:



This process yields similar results regarding iPad 2’s, but now we’re seeing some major numbers that impact the total margin and bottom line. iPad 2’s, Xbox 360, and Sony PlayStation 3 are all major contributors to the negative margin. With this in mind, it may be worth banning re-sales of these specific items (iPad 2, Xbox 260, PlayStation 3). Or possibly set a non-negotiable trade-in price that would ensure higher profitability as these 10 items alone accounted for –$609,000 in margin.

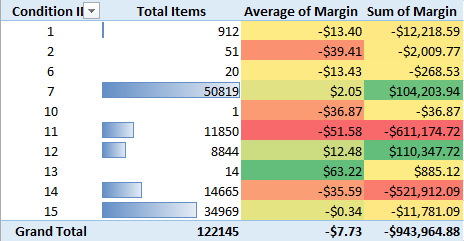
Given these results, we are curious as to which items are bought and re-sold the most, profitable or not. Below are our findings:



As you can see, the items being bought back and re-sold most are some of the most unprofitable in the entire assortment. This trend needs to be remedied for overall company health.

*Understanding Product Conditions*

As noted earlier, when evaluating average margin, the least profitable items were consistently identified as “AS-IS”. Per the Product Condition tab, this is defined to be “This item has some damage or defect as noted here, and is sold AS IS”. Based on that list of 10 items, it’s worth exploring overall profitability by Product Condition.



“AS-IS’ accounts for about 12% of the overall items bought and sold. Yet, the total margin tied to this condition is –$522,000, a significant loss for the company.

In that same vein, “NOT WORKING – Liquidation” (Condition 11) is even worse. Defined as “This item is NOT WORKING, and is being sold for parts”, items in this condition account for less than 10% of the total bought and sold, yet add up to –$611,000 in total margin.

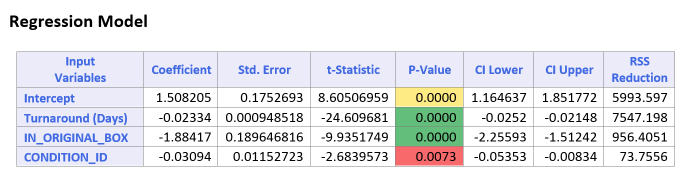
Eliminating these two conditions by either not accepting them for trade-in or offering the customer an extremely low buyback price could solve the margin issues for this business. It would only remove about 205 of the trade-in items, but it would make the entire operation profitable.

*What Are Consumers Actually Looking For?*

To examine this phenomenon further, we took a random data set of 15,000 records.  We partitioned data to 60% training and 40% validation.  We limited variables to “Margin Percentage” (which was our outcome), “In Original Box”, and “Product Condition”.

From this result, we saw all three predictors were statistically significant, although average turnaround time and IN\_ORIGINAL\_BOX variables had a larger influence on profit margin.  This was interesting, as we had theorized that the CONDITION\_ID of an item would influence the profit margin more than the turnaround days, or whether or not the item was in its original packaging.

It is also worth mentioning that all variables have a negative correlation to Margin Percentage, so the longer an item stays in inventory, the higher the product condition ID (meaning, a worse grade of inventory, and whether or not the item is in the original box) all negatively affect profit margin.



As we delved deeper into understanding the factors at play with respect to our product margins, we recognized that we were seeing some interesting patterns with respect to the processing speed of the electronics up for resale. After running a regression to examine more precisely how processing speed affects margins, we saw that, oddly enough, the more processing speed the device has, the smaller the margin we get on resale.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Input Variables** | **Coefficient** | **Std. Error** | **t-Statistic** | **P-Value** | **CI Lower** | **CI Upper** |
| **Proc Speed (Ghz)** | (0.8308) | 0.2626 | (3.1644) | 0.0016 | (1.3456) | (0.3160) |

After searching for an explanation of this strange phenomenon, we came to a conclusion as to why this was taking place. Our primary customers are people looking to buy valuable components. Furthermore, the process of extracting the component and conserving processing speed is very difficult. Thus, what we have here is the company putting too high of a value on processing speed when the product’s buyers are focused on the parts and attaining the cheapest price for them. Therefore, we could avoid this effect on product margins if we emphasized products that were easy to dissemble rather than those traits that a consumer of a new product might desire from their electronics.

To generalize, the turnaround time indicates that the consumer for the most part wants something that has not been on the market for a long time, is in good condition, and is in the original box. We also know that processing speed is not as desirable of a component as originally anticipated.

*How Does Time Change Outcomes*

We also wanted to better understand how time contributed to the resale of the product and its margins. Using scatter plots and regression analysis, we explored how sales volume was affected by turnaround time.

We defined turnaround time as the difference between the date the product was purchased by the site for resale and the date the product sold to a customer. In order to be considered, each turnaround day unit had to contain 10 or more data points. After eliminating some of the data which did not meet this threshold, we then took this data and placed it on two scatter plots: (1) margin % versus turnaround days and (2) volume of sales versus turnaround days.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Turnaround | Count of Product Type | Average of BOUGHT\_FOR | Average of SOLD\_FOR | Margin |
| 165 | 41 | 109.37 | 82.27 | -32.9% |

From the above data, we concluded that within the first 50 days for approximately 10% of our products, the margin is decreasing at a rate of .6% per day. This means within 50 days a product decreases in margin by 15%.

But while our examination lead to some conclusions, it is also produced more questions.

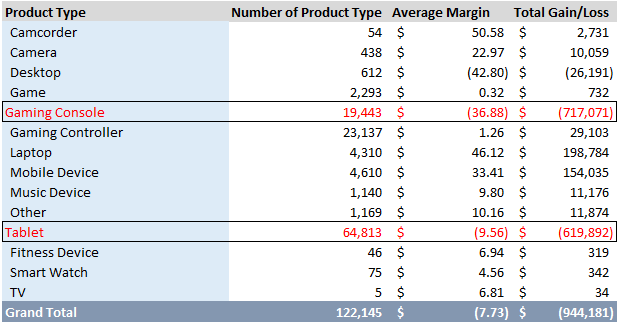
Thereafter, we see more variability and a deceleration in our margin drop. Is this due to our different product types? Knowing the different product types and the volume of sales, if we separate the product types are we going to come up with more conclusive results that can be verified by P-values as significant?

We also noticed that from 51 days to 60 days, we have another 10% of our dataset being sold. We were able to conclude the following hypotheses:

* The company is noticing this drop in margin and unloading the product.
* New inventory is coming in; therefore it is necessary to get rid of old inventory in order to make room for the new inventory. This becomes a deeper dive into operations then and asks the questions: how are they currently tracking inventory?

Based on the trends in our data and corresponding graphs, we can generalize that at about 30.5 days, our margin goes into the red for most of our products.

Moreover, on the table below, we see, per product type, our average margins, volume of product sales, and the net gain/loss from those product types from August 26, 2014 to March 6, 2015. While tablets and gaming consoles are generating the highest volume of sales, they are also creating substantial losses.



Gaming Consoles and Tablets

Given the sheer volume of sales and the financial impact of gaming consoles and tablets, we thought it was best to see at what rate they are impacted when it comes to turnaround days and margin.

With gaming consoles, we generated somewhat perplexing results after running a regression analysis: our coefficient is positive. Based on our initial data however, we know that that is not the case.

Looking at the graph below, we see sudden declines in margin, but we also see where margin rebounds. This is due to the timing of events, the introduction of new products, and the memory of the console.

For our data set we pulled only old systems, that because it reflects a story. As we can see below, the higher memory for our device, the more likely we were able to keep steadier margins on the product.

This is true, until new products with better technology and more memory are introduced (in this case the PS4 and the Xbox One). That is where we see an inflection point and where statistically, we see a p-value that is significant. In conclusion, if new products are introduced into the site, when that occurs, our margin is going to drop at a rate of 1.5% per day. Furthermore, these two products are eating away at other products during the holiday season, meaning that these two products are probably the “hit” toy for the holiday.

PS4’s were selling on the site in December for approximately 120-150 days after our other products, meaning that this coincides with our erosion of margin for our older consoles. We can take this example and we could run further analysis on other product types and most likely field similar results.

|  |  |  |
| --- | --- | --- |
| **Product** | **Volume** | **Average of DATE\_SOLD** |
| Sony Playstation PS4 500GB Gaming Console - Black | 300 | 12/14/2014 |
| Microsoft - Xbox One 500GB Console with Kinect Sensor | 44 | 12/4/2014 |
| Microsoft - Xbox One 500GB Console without Kinect Senso | 23 | 12/24/2014 |
| Microsoft - Xbox One 500GB Console without Kinect Sensor | 25 | 12/14/2014 |
| Microsoft - Xbox One Console | 88 | 12/26/2014 |

From our regression analysis:

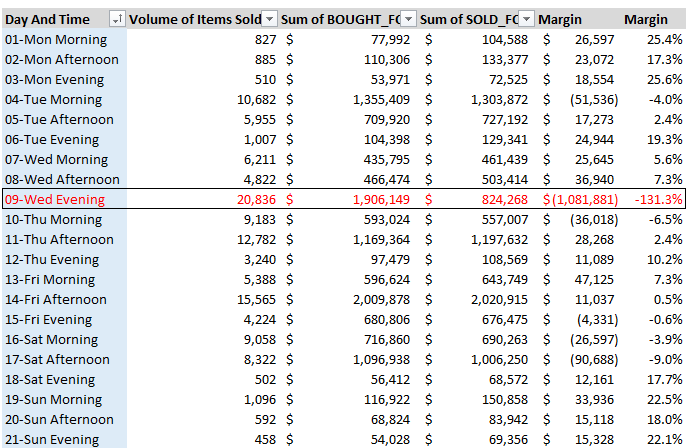
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Input Variables** | **Coefficient** | **Std. Error** | **t-Statistic** | **P-Value** | **CI Lower** | **CI Upper** | **RSS Reduction** |
| **Intercept** | 1.548 | 0.250 | 6.181 | 0.000 | 1.047 | 2.049 | 29.92 |
| **Turnaround** | (0.015) | 0.002 | (9.002) | 0.000 | (0.018) | (0.011) | 4.92 |

For tablets, our regression statistics make more sense. From this data, we can state that for every increase in turnaround days our margin is eroding by .1% for tablets. Moreover, a similar analysis such as that of the gaming consoles can be done, and most likely would concede similar results.

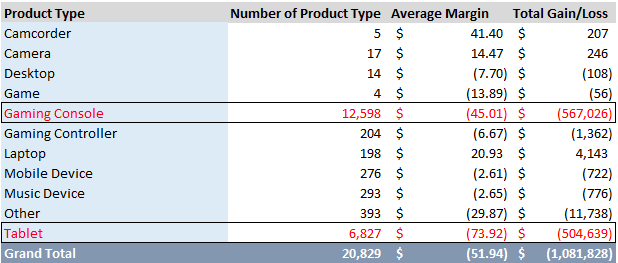
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Input Variables** | **Coefficient** | **Std. Error** | **t-Statistic** | **P-Value** | **CI Lower** | **CI Upper** | **RSS Reduction** |
| **Intercept** | 0.091 | 0.015 | 6.223 | 0.000 | 0.062 | 0.120 | 0.19 |
| **Turnaround (Days)** | (0.001) | 0.000 | (6.964) | 0.000 | (0.002) | (0.001) | 12.86 |

*How Does Time of Purchase Effect Volume and Margins?*

Wednesday evening is our worst day and time of the week for sales. If we exclude Wednesday evenings we see an overall positive margin of $137,916. Since Wednesday is also our highest volume day, we hypothesize that the losses incurred on Wednesday night are linked to =poor inventory management. We guess is that we are stocking up on too many items and are then trying to liquidate the inventory to make room for the new inventory, which arrives on either Wednesday or Thursday.



Below are the items sold on Wednesday and their respective margins. This begs the question: can we package products? In other words can we use products with better margins to pass on the items with worse margins. Given the net losses, it is very unlikely. Therefore, the site must focus more on inventory management to prevent the huge losses incurred on Wednesday evenings.



**Overall Conclusions**

If we run a regression with all variables included, we can come up with an all-encompassing model for best predicting product margin. There are risks to doing so however, as our other analyses indicate that seasonality and new product introductions may influence the model and ultimately make it less applicable.

In conclusion, we can’t paint everything with a single paintbrush. That’s where segmentation of products becomes very important. Here are some takeaways:

1. Turnover matters per brand type and product type. Some products retain their value while others are not able to retain their value for as long.
2. New introductions of products matter as it will affect margins of older products.
3. Erosion can be stopped based on better inventory management as we have found out that 30 days is the optimal time for turnover. We also can come to the conclusion based on time of day sold.
4. Better selections of individual products are needed.
5. Different times of year bring about different markets. So, when it’s not Christmas your buyer is more likely to buy more for the parts instead of the latest and greatest device.
6. As-is items are a horrible buy as indicated by the significant loss of dollars.
7. Apple is a popular brand and sells high volumes, but then again, it revolves around buying too much and not being able to turnaround inventory fast enough.
8. Memory helps retain value…to a point.
9. Overall, the company needs better inventory management.
10. As with any data analysis, we need further digging into the data. This is limited though to the confidentiality of the company of where we are retrieving the information.

Ultimately, delving into this data set gave us the opportunity to explore how business intelligence can be applied in a real business experience. It also demonstrated that oftentimes, though a comprehensive solution may seem possible, it does not actually best predict the events or circumstances behind the data. With that in mind, it is important to acknowledge the human component that must have ultimate say in the decision-making process.

1. http://fortune.com/fortune500/best-buy-co-inc-60/ [↑](#footnote-ref-1)
2. *Also worth noting is that almost every item on this list is classified as “AS-IS” – more on this in a bit.* [↑](#footnote-ref-2)