Revised: November 22, 2016

Solution

The solutions provided below are only for reference which helps in evaluation and by no means should these be treated as the only way answer the questions. A fair evaluation should be done if the student comes out with a different and perhaps better explanation for any question

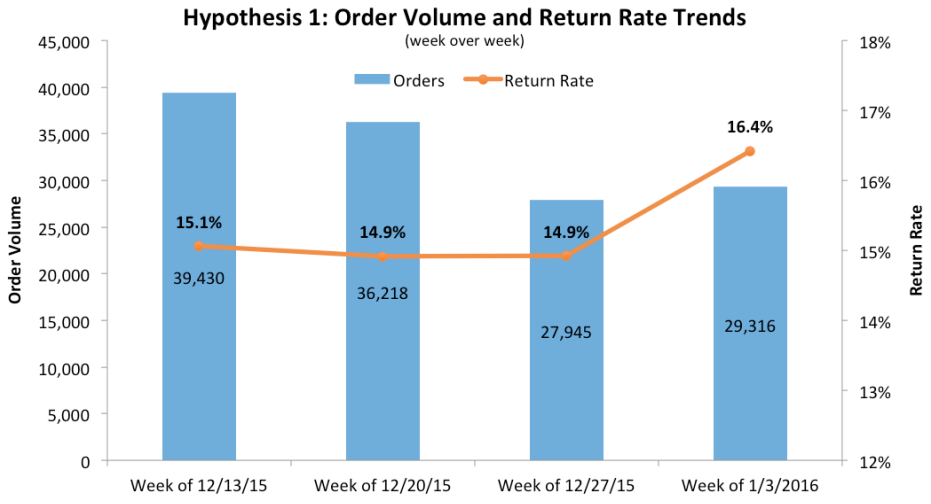
Step 1: Data Collection and Cleaning

- Merge both the transactional and return data into a single dataset using VLOOKUPs. Data should be merged on Order ID, the most unique column in both tables
  - Recommended that two columns are added to sales data: 1) "Returned" column recording using 1 and 0 whether an order was returned or not, 2) "ReturnCode" column recording return code if order was returned

STEP 2: Identifying *the Root Cause*

1. ***Hypothesis 1: The increase in return rates coincides with a surge in the number of orders after the holidays***

   To evaluate this hypothesis, we want to compare order volume and return rates week over week. After merging the transaction data, we'll want to use a pivot table to summarize total orders and return rates for each week, which we will then plot onto a bar and line chart as follows:

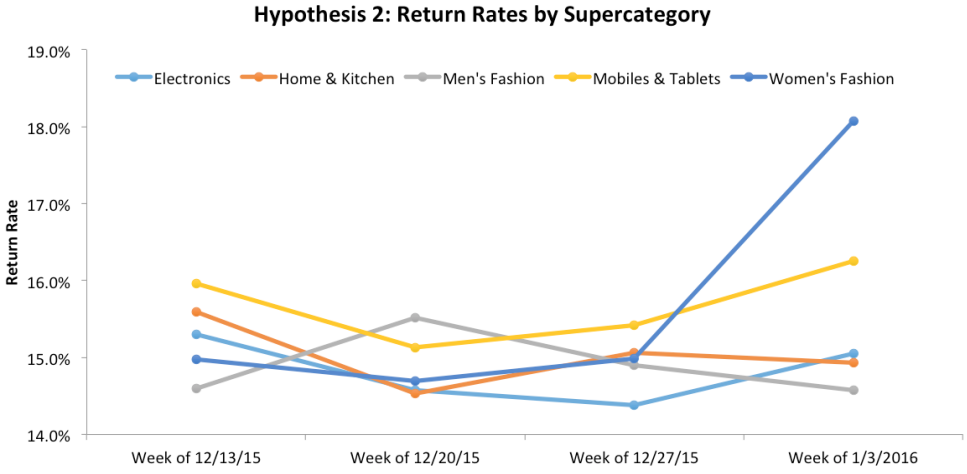


   As expected, we see a 10% increase in return rates from 14.9% (4,163 orders returned) to 16.4% (4,807 orders returned) between the week of 12/27/15 and 1/3/16. This is clearly unusual as the three weeks prior show a steady state of ~15% return rate. Absolute returns increased by 15.5% while total orders only increased by 4.9% between those two weeks. Even though orders increased in the week of 1/3/2016, it is still substantial less than the order volume during weeks 12/13/15 and 12/20/15, where return rates were consistently 15%.

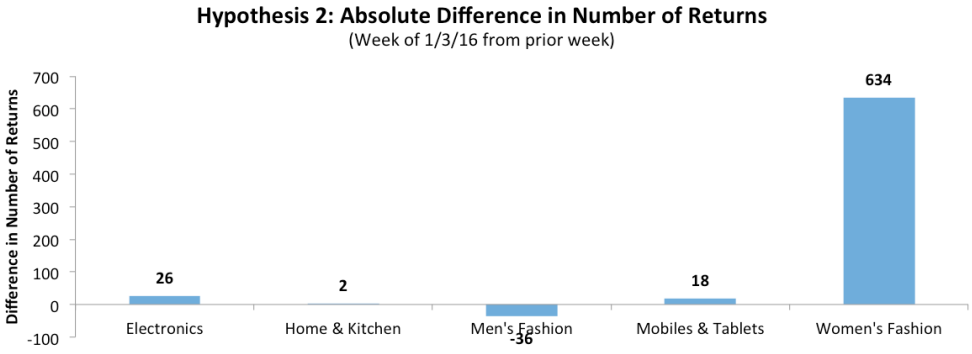
   Thus, hypothesis 1 is FALSE.

2. ***Hypothesis 2: The increase in returns can be attributed to products in the Women's Fashion supercategory***

We'll want to break out the return rate trends by supercategory before digging into specific product categories or product IDs. This gives us a systematic approach to drilling down into the data so as to avoid missing key insights that affect higher levels of the product categorization.



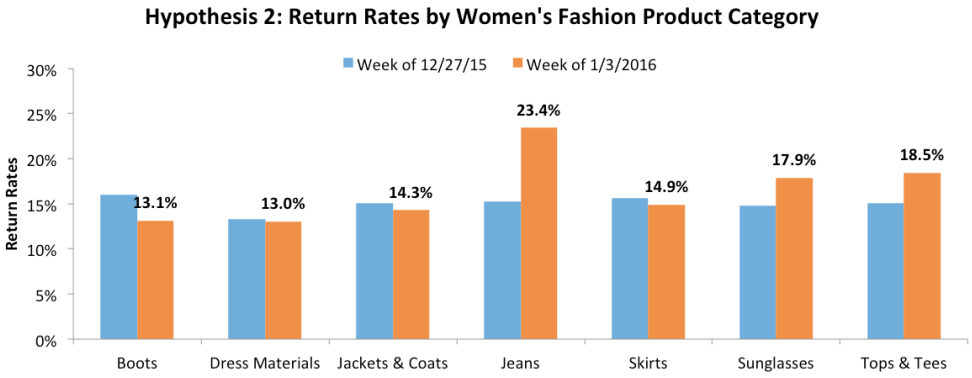
Over this period, there's a 20.5% surge in return rates in the Women's Fashion category (from 15% to 18%) and slight increases in Mobile & Tablets (5.4% increase) and Electronics (4.7%). To determine how significant the contribution of these increases are, let's evaluate the absolute difference in number of returns between week of 1/3/16 and week of 12/27/15:



Even though there are slight increases in return rates for Electronics and Mobile & Tablets, the surge in returns within Women's Fashion accounts for nearly 94% of the absolute increase in returns.

Thus, hypothesis 2 is TRUE.

While we have proved hypothesis 2, before moving further to hypothesis 2a, we will want to see if we can isolate the cause to specific product categories or product IDs. Let's start with product categories.



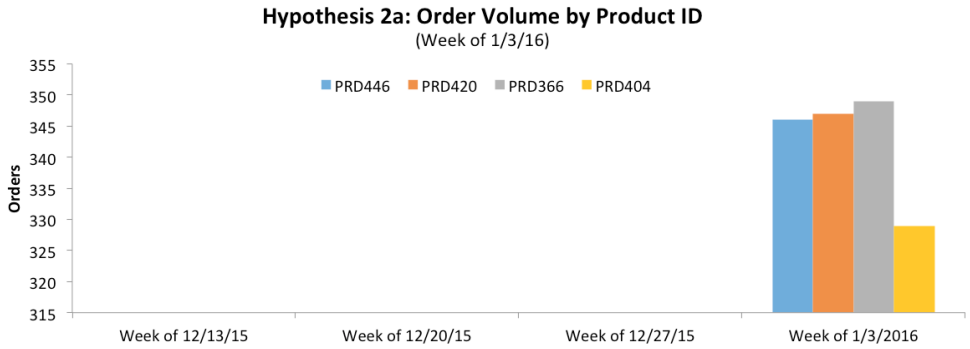
At the product category level, it appears that there are increases in return rates for Jeans, Sunglasses, and Tops & Tees. Let's dig into return rates in the week of 1/3/2016 for the individual product IDs in these three categories. Let's sort them in a pivot table similar to this:

ProductID	Category	Return Rate (Week of 1/3/16)	Absolute Returns (Week of 1/3/16)
PRD446	Sunglasses	50%	174
PRD420	Tops & Tees	49%	170
PRD366	Jeans	46%	162
PRD404	Jeans	40%	132
PRD436	Jeans	27%	19
PRD453	Jeans	26%	19
PRD490	Tops & Tees	23%	15
PRD390	Tops & Tees	23%	16
PRD457	Tops & Tees	23%	16
PRD486	Tops & Tees	22%	15
PRD421	Sunglasses	22%	13

While there are 61 product IDs that have a return rate greater than 15% that week, there are four outliers that have 40%+ return rates and more than 100 absolute returns. We can now isolate the root cause to these four product IDs.

3. ***Hypothesis 2a: The increase in returns within the Women's Fashion supercategory is specific to new products added in the first week of January***

Since we have isolated the problem to four specific products, this hypothesis can only be proven if all four products are new products added in the first week of January. Let's evaluate this by seeing order volume over the four week period.

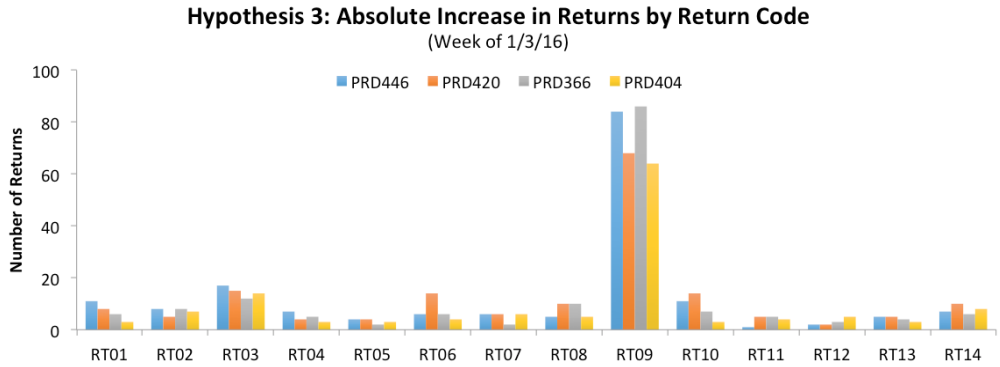


From this, it is clear that all four products with especially high return rates are new products added in the first week of January. From looking at the pivot table of all products in this supercategory, there are no other new products added in the first week of January, meaning that the increase in returns within the Women's Fashion supercategory is specific to new products

Thus, hypothesis 2a is TRUE.

4. ***Hypothesis 3: The increase in returns can be attributed to a surge in items arriving late***

Now that we have isolated the affected product IDs, let's figure out the primary return reason(s) for this increase. We again only care about the increase in returns for these product IDs in the week of 1/3/16. Since there is no return data prior to January since these are new products, this is the same as absolute return numbers.



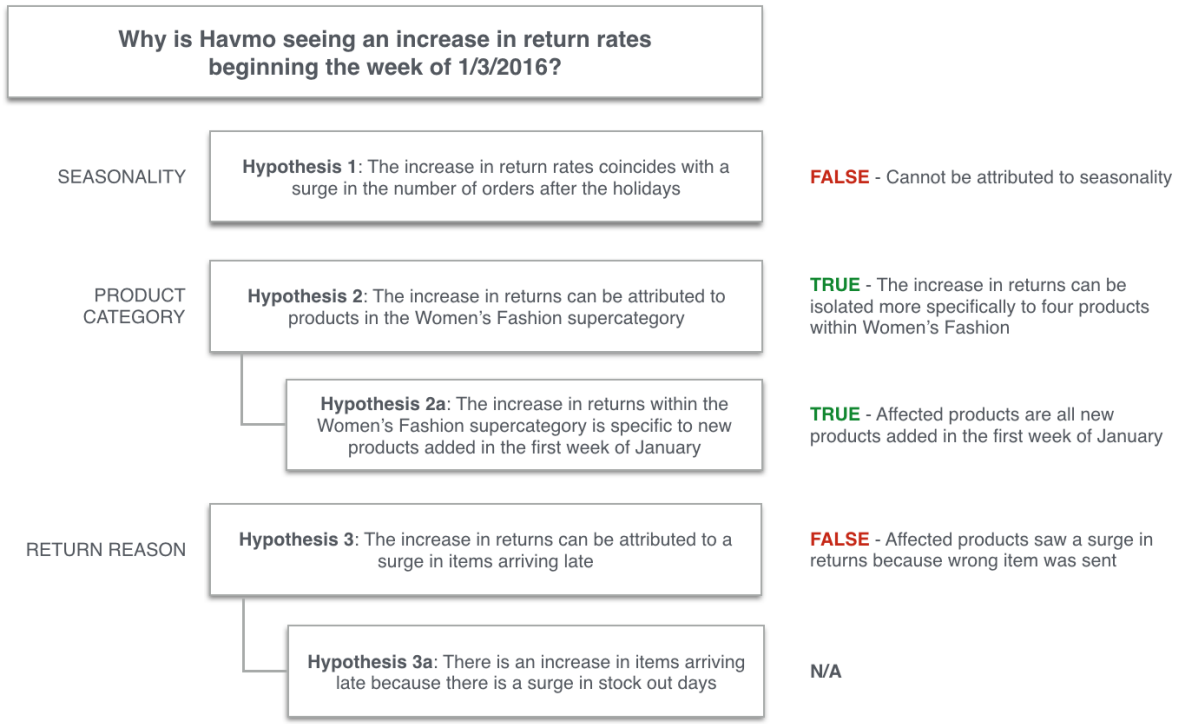
There is a clear surge in returns for return code RT09, "Wrong item was sent." The suspected RT06 return code, "Item arrived too late," is not a significant return reason.

Thus, hypothesis 3 is FALSE. This means that we also do not need to dig into hypothesis 3a (stock outs), which is contingent on 3 being TRUE.

Note that while hypothesis 3 turned out to be false, we did discover something else while trying to validate it: that "Wrong item was sent" seems to be a significant issue, and as a recommended next step (beyond this analysis), we should look further into what's causing that.

STEP 3: Summarize Results of the Analysis

Revised: November 22, 2016



STEP 4: Recommendations, next steps, and final comments

From the analysis, we can conclude that the increase in return rates is a result of wrong items being sent when customers ordered new products within the Women's Fashion supercategory that were added in the beginning of January. The root cause is then a quality assurance (QA) issue within the Women's Fashion logistics process.

Havmo may need to further investigate why these new products are not being retrieved correctly. Perhaps these products were mislabeled or the inventory received in the warehouse were not correct to begin with.

To resolve this issue completely, the recommended next step would be to conduct a full investigation, and further root cause analysis into these logistics issues. e.g. Is this issue specific to the Women's Fashion vertical, or is something broken with the way we handle all new products? That will let us isolate the points in the process that are mishandled and correct these issues with better QA throughout the storage, retrieval and packaging process. Havmo may also want to apply similar improvements to other warehouses to prevent this issue from arising.