Is Unseen Unsold?

The Impact of Product Visibility on Store Sales

Adithya Suresh, Darby Brown, Nicole (Man Wei) Kan

Lab 2 | Datasci 203-08 | Spring 2023

Introduction

Throughout the Covid-19 pandemic, grocery stores across the US struggled to manage surging demand for products ranging from toilet paper to oat milk.[1] While this extreme situation has since subsided, the problem of fluctuating demand is evergreen. The recent proliferation of online grocery sales, where Acme has been successful, has resulted in the need to manage demand variability stemming from multiple sales channels.[2] We seek to understand the levers that Acme can use to manage this demand variability and maximize sales in normal operations. A robust understanding of these mechanisms in normal operations can also benefit Acme during more extreme surges. This research explores store layout as one 'lever.' The store layout is a highly complex problem, so our scope is limited to a single aspect of layout: product visibility.

Research question: ***How does a product's visibility in grocery stores affect sales?***

Understanding the impact of in-store product visibility is highly beneficial in that:

1. Acme can more effectively adjust its grocery store shelves to manage demand in relation to inventory fluctuations
2. Acme can maximize revenue through optimal visibility of high margin products.

Data

The BigMart retail sales dataset was collected by data scientists at BigMart in 2013 to understand the effects that products and stores play on sales. The dataset contains observational data for 1,559 products across ten stores in different cities for 14,204 rows and 12 columns, each row representing a product sold in-store and each column representing an attribute of the product. The attributes of a product can be divided into two areas. The first area is related to the item-specific attributes, including the item's identifier, weight, fat content, and visibility in store, which is measured as the % of the total display area of all products in a store allocated to the particular product, item type in 16 different categories, the maximum retail price and the total sales for the item. The second area of attributes is related to the stores where the item is sold, such as outlet's identifier, establishment year, outlet size in ordinal form from small to high, outlet type in categorical form (grocery store, supermarket type1, supermarket type2, and supermarket type3) and outlet location type in categorical form (tier1, tier2, and tier3). We saw some cleanups needed with the initial exploration of the dataset. First, there were 2,439 missing values in the item weight column. We filled in the missing item weight based on the item identifier, as the same item should have the same weight. Second, we observed 879 rows with 0 item visibility that can be removed as they practically do not make sense given that an item in-store should at least physically occupy more than 0% of the display area. Lastly, there were some inconsistent labeling on the categories in the fat content column and missing values in the outlet size column. We did not modify the two columns since they were not used within the different models.

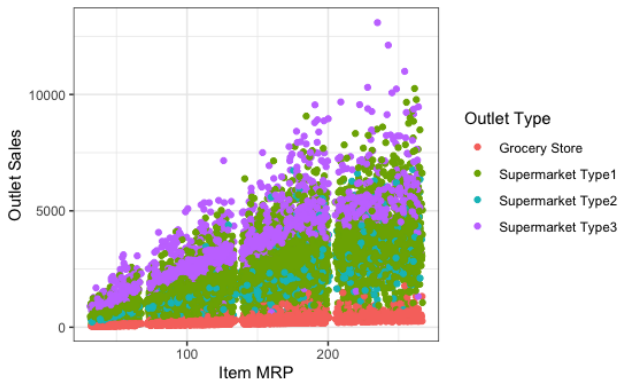
Key Concepts and Operationalization

The key concept for our research question is to understand the impact of product visibility in stores on sales, thus allowing store managers to maximize sales by strategically allocating display areas for different products. For the operationalization of the research, we determined the sales of a product in a particular store as the dependent variable (Y) and the product's visibility as the independent variable (X). Other key explanatory variables include the product price, item type, and store type.

Here are the key and supporting variables in BigMart's dataset used for operationalization:

- **Item_Outlet_Sales:** Sales of the product in the outlet
- **Item_Visibility**: The visibility of the product in store as a percentage of available space
- **Item_MRP**: Maximum retail price of the product

- **Item_Type:** Categorical variable representing 16 different product types (Dairy, Fruit and Vegetables, Household, Health and Hygiene, etc.)
- **Outet_Type:** Categorical variable representing three Outlet Types (Supermarket 1, Supermarket 2, Supermarket 3, Grocery Store). The BigMart dataset did not provide descriptors for the categories, however, based on the distribution of price vs. sales and the mean sales for each Outlet Type, we can deduce the size of the Outlet Types from largest to smallest on average follows the order:
  Supermarket Type 3 > Supermarket Type 1 > Supermarket Type 2 > Grocery Store.



Outlet_Type <fctr>	Mean_Outlet_Sales <dbl>
Grocery Store	321.9807
Supermarket Type1	2220.4602
Supermarket Type2	1920.1703
Supermarket Type3	3596.0209

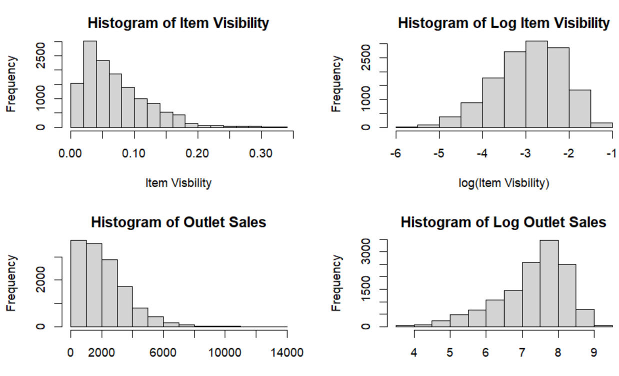
These variables effectively represent the overall effectiveness of visibility in-store. A critical gap between the concept and operationalization is that item visibility measured as the % of display area occupied is not representative of the visibility of a product in real life as it does not include factors such as location, positioning, and design of products. We also wanted to explore potential nuances for product types. A separate regression can be fit to review the visibility impact in different product types. We included only one of the outlet-related variables to avoid issues with multicollinearity that can lead to unreliable coefficient estimates. Item weight and fat content were also not included as they did fit our conceptual causal model.

Modeling Approach

After filtering the initial BigMart dataset, 879 observations were removed from the dataset due to having 0 visibility leaving 13,325 rows remaining. Then, 3998 rows (~30%) were used for the exploration set, while 9327 rows (~70%) were used for the test dataset.

	GVIF	Df	GVIF^(1/(2*Df))
Item_Visibility	1.119097	1	1.057874
Item_MRP	1.010528	1	1.005250
Item_Type	1.022522	15	1.000743
Outlet_Type	23.499824	3	1.692430
Outlet_Size	32.405988	3	1.785545
Outlet_Location_Type	26.696520	2	2.273074

Outlet Location and outlet Size covariates were intentionally left out of the model due to high collinearity among Outlet Type, size, and location type per the high VIF values. Since Outlet Type did not have missing values and provided the best data for classifying an outlet, we prioritized that variable over location and size.



Log transformations were applied to item visibility and outlet sales variables since they were both right skewed, and applying the log transformation to the variables normalized the data. We also decided not to include item types as we wanted to see the impact of visibility on different item types. To do so, we created models that used data filtered for the two largest item types for comparison.

Regression Table

We created three different models starting with only Log(Visibility) and adding new covariates as we enhanced the model specifications. As shown in model2 and 3, we decided to include an interaction term for Log(Visibility) and MRP to capture a combined effect of the two variables on sales.

Table 1: Regression Model Results Comparing Models

	Dependent variable:		
	Log(Outlet Sales)		
	(1)	(2)	(3)
Log(Visibility)	-0.196^{***}	-0.156^{***}	0.028^{*}
	(0.012)	(0.026)	(0.015)
Maximum Retail Price (MRP)		0.007^{***}	0.008^{***}
		(0.001)	(0.0003)
Type 1 Supermarket			1.954^{***}
			(0.014)
Type 2 Supermarket			1.774^{***}
			(0.019)
Type 3 Supermarket			2.476^{***}
			(0.019)
Log(Visibility) - MRP Interaction		-0.0003^{*}	-0.0002^{**}
		(0.0002)	(0.0001)
Constant	6.727^{***}	5.686^{***}	4.470^{***}
	(0.038)	(0.079)	(0.045)
Observations	9,327	9,327	9,327
R^2	0.026	0.308	0.789
Adjusted R^2	0.026	0.308	0.789
Residual Std. Error	0.945 (df = 9325)	0.797 (df = 9323)	0.440 (df = 9320)
F Statistic	247.287^{***} (df = 1; 9325)	$1,381.434^{***}$ (df = 3; 9323)	$5,807.752^{***}$ (df = 6; 9320)
Note:			*p<0.1; **p<0.05; ***p<0.01

Results and Discussion

The results of the three models show that Log(Visibility), MRP, and Outlet Type where an item is sold and the interaction between Log(Visibility) and MRP are all statistically significant to predict Log(Outlet Sales).

Comparing the difference in the coefficient and significance of Log(Visibility) across the three models, we see that adding Outlet Type decreases its significance drastically, and the coefficient has changed from a negative to a positive term. The change indicates that visibility was partially absorbing the impact of Outlet Type, and this could be because the amount of visibility an outlet can afford to give an item is dependent on the type of outlet, as that determines the amount of available space and number of goods that are stored in the outlet.

The interaction term with negative coefficient and the positive coefficients for Log(Visibility) and MRP in the third model suggests that the effect of Log(Visibility) on Log(Outlet Sales) is not the same at all levels of MRP, which indicates that the impact of visibility on sales changes with price. This highlights that different prices may result in differences in the impact of visibility on sales, where lower MRP may lead to more impact of product's visibility on sales and higher MRP may lead to weakened or reversed the effect of visibility on sales.

Also, when analyzing the models, we see that MRP and Outlet Type have a much more practical impact on the Log(Outlet Sales) prediction as the R-squared value drastically increases from 0.026 to 0.789 on the test dataset after the two variables are added. This massive increase indicates that the first model had a high omitted variable bias. Including MRP and Outlet Type has improved the model fit and provided a more accurate estimate of the true relationship between the explanatory variables and sales.

Table 2: Regression Model Results Comparing Item Data

	Dependent variable:		
	Log(Outlet Sales)		
	All Products	Fruits/Veg	Snack Foods
Log(Visibility)	0.028^{*}	0.030	0.033
	(0.015)	(0.044)	(0.042)
Maximum Retail Price (MRP)	0.008^{***}	0.008^{***}	0.008^{***}
	(0.0003)	(0.001)	(0.001)
Type 1 Supermarket	1.954^{***}	1.966^{***}	2.003^{***}
	(0.014)	(0.038)	(0.040)
Type 2 Supermarket	1.774^{***}	1.797^{***}	1.856^{***}
	(0.019)	(0.050)	(0.053)
Type 3 Supermarket	2.476^{***}	2.593^{***}	2.526^{***}
	(0.019)	(0.050)	(0.053)
Log(Visibility) - MRP Interaction	-0.0002^{**}	-0.0001	-0.0001
	(0.0001)	(0.0003)	(0.0003)
Constant	4.470^{***}	4.472^{***}	4.485^{***}
	(0.045)	(0.131)	(0.131)
Observations	9,327	1,327	1,315
R^2	0.789	0.796	0.768
Adjusted R^2	0.789	0.795	0.767
Residual Std. Error	0.440 (df = 9320)	0.442 (df = 1320)	0.450 (df = 1308)
F Statistic	$5,807.752^{***}$ (df = 6; 9320)	857.568^{***} (df = 6; 1320)	720.034^{***} (df = 6; 1308)
Note:			*p<0.1; **p<0.05; ***p<0.01

When determining the impact of the model, we also wanted to look at how the model performed on the test dataset and how it worked on individual item categories. Interestingly, we found that the Log(Visibility) and its interaction with MRP were no longer significant when the dataset was filtered for a single item. This result indicates that visibility impacts sales but is not consistently significant across different item types.

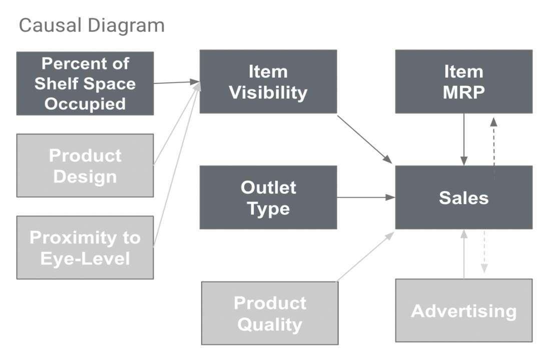
Furthermore, the coefficient change between the models does not look like Item type affects the impact of MRP, Outlet Type, or Log(Visibility) on Log(Outlet Sales). Thus, the model to explain Log(Outlet Sales) is not significantly influenced by the product type sold in an outlet.

Limitations

The results of our study are contrasted by affirmative studies in recent years.3, 4 This contrast and significant error in the models lead us to believe that omitted variables and reverse causality impacted our results. These are visualized in Figure 2.

Considering the operationalization of our key concept, product visibility, the variable used only represented how much physical space an item type occupies on grocery store shelves. It did not consider the effect of the product's visibility at eye level or the appeal of the product's design. An item tucked away on the bottom shelf or one with underwhelming packaging would conceivably have fewer sales than a flashy, front-and-center one. Omission of *product design* and *proximity to eye level* both introduce a bias away from zero for visibility and price. We suspect that price absorbs some of the causal effects of these omitted variables. Other omitted variables could include the product's advertising (i.e., catalogs or sales) and the product quality. We suspect advertising would vary by Outlet Type, and therefore, Outlet Type is absorbing some of the effects of advertising. The same could be true of product quality. Omitting product quality also likely introduces a bias away from zero for the item price.

Figure 2. Causal diagram of grocery sales.



We also expect minor reverse causality between Item MRP and sales which would introduce a bias away from zero, artificially heightening the effect of price on sales in our model. While the model needs only to meet two assumptions, iid and no perfect collinearity, we question whether the assumption of iid is met. The data may only be partially independent because it is only gathered at BigMart stores, and the percentage visibility of one product directly depends on the percentage visibility of all others within that store. This problem would be complicated to remedy while maintaining a large sample, so we accept it as a necessary limitation of the experimental design. Based on these limitations, we place less confidence in our results and would hesitate to conclude that Outlet Type and price are the only primary contributors to product sales.

Conclusion

When examining the impact of visibility, we see that its practical significance to outlet sales is at least a degree of magnitude smaller than the impact of MRP and Outlet Type. However, we did see a correlation between the impact of visibility and item MRP, where for extreme prices, visibility mattered more, while for moderate prices, it mattered less. Based on this result, our recommendations to the store would be to provide low-priced items with high visibility and extremely high-priced items with lower visibility. This recommendation intuitively also makes sense as for much higher priced items, a customer is only likely to buy them if they intend to, making visibility much less important to outlet sales. In contrast, cheap items can be bought compulsively, potentially increasing outlet sales.

While our results showed that Outlet Type and MRP greatly impact sales, the impact of all variables in this model is likely being overestimated due to omitted variables. Visibility's impact should be better defined as the combined impact of all variables that affect an item's visibility to customers in an outlet. Further research specific to an individual store and product visibility concept is recommended. Even for the same chain, there are vastly different results, so it would make sense for each store to look at data specific to their location to make better decisions on whether visibility is vital for different item types.

References

1. 'There Is Plenty of Food in the Country'. Michael Corkery, David Yaffe-Bellany, Amelia Nierenberg and Quoctrung Bui. March 15, 2020. https://www.nytimes.com/2020/03/15/business/coronavirus-food-shortages.html
2. Grocery omnichannel perishable inventories: performance measures and influencing factors. Soroosh Saghiri, Emel Aktas, Maryam Mohammadipour. International Journal of Operations & Production Management. 7 March 2023. https://www.emerald.com/insight/content/doi/10.1108/IJOPM-06-2022-0397/full/html
3. Retail Store Layout Optimization for Maximum Product Visibility. Evren gul, Alvin Lim, Jiefeng Xu. May 2021. https://www.researchgate.net/publication/351706444_Retail_Store_Layout_Optimization_for_Maximum_Product_Visibility
4. Predictive model of attention in viewing selected grocery products. Dr. Rosemary R. Seva, Kimberly Go, Kathleen Garcia, Winnie Grindulo. 2011. https://www.researchgate.net/profile/Rosemary-Seva/publication/279603642_Predictive_model_of_attention_in_viewing_selected_grocery_products/links/60a1b944458515c265994486/Predictive-model-of-attention-in-viewing-selected-grocery-products.pdf?_sg%5B0%5D=started_experiment_milestone&origin=journalDetail
5. **Data**: https://www.kaggle.com/datasets/thedevastator/bigmart-product-sales-factors