

Data Explanation

predictor	effect	rationale
Target (y)		
sales	dv	
outlet_type	+/-	Some stores could do better than others I am not sure.
city_type	+/-	I am not sure on the effect as the items could be need based.
Outlet_ID	+/-	I will need this for item #3 + upper-level.
Predictors		
Item_Weight	+/-	Meat is heavy, but spices are light both are pricey items.
Item_Fat_Content	+/-	I am not sure I think it depends on the buyer.
Item_Visibility	+/-	An item may have favorable visibility because it heavy and large. Or it may be a seasonal item, which may of already purchase items for the occasion.
Item_Type	+/-	Once again I think it depends on the buyer's needs.
Item_MRP	+	low MRP may have more sales.
Outlet_Year	-	An old outlet may have less appeal.
Outlet_Size	+	A large outlet will have more sales than a small one.
Excluded		
Item_ID	NA	This level of detail is not needed.

I used fixed and random effects models because the data had levels (outlet_id) and multi-level data works with these models. It seems overall the levels were not as pronounced as we would like. I also used log for sales to alter to a somewhat normal distribution and scaled item mrp to log as well. I dropped item_weight (1463) and outlet_size (2410) because of NAs. I did include item_type in my R code models to see which items were were best performing. I used outlet_type and city_type to help recommend which area and supermarket was best.

A5. Big Mart Model Output

Dependent variable:		

log_item_sales		
OLS	linear	
	mixed-effects	
(1)	(2)	(3)

item_visibility	0.011	0.007	0.009
	(0.113)	(0.113)	(0.113)
log_item_mrp	1.022***	1.022***	1.022***
	(0.011)	(0.011)	(0.011)
outlet_year	0.002	0.020*	0.002
	(0.001)	(0.012)	(0.002)
city_typeTier 2	-0.017	-0.135***	-0.015
	(0.020)	(0.052)	(0.027)
city_typeTier 3	-0.027	-0.284*	-0.030
	(0.018)	(0.159)	(0.024)
outlet_typeSupermarket Type1	1.938***	1.689***	1.938***
	(0.020)	(0.158)	(0.025)
outlet_typeSupermarket Type2	1.752***	1.567***	1.756***
	(0.038)	(0.135)	(0.048)
outlet_typeSupermarket Type3	2.512***	2.756***	2.511***
	(0.026)	(0.158)	(0.034)
outlet_idOUT013		0.464	
		(0.290)	
outlet_idOUT017		-0.025	
		(0.065)	
outlet_idOUT018			
outlet_idOUT019			

outlet_idOUT027

outlet_idOUT035 0.054
(0.034)

outlet_idOUT045

outlet_idOUT046

outlet_idOUT049

1.) What type of outlet will return him the best sales: Grocery store or Supermarket Type 1, 2, or 3.

I would recommend my client to invest their money into Supermarket type 3 tier 1 city. if cost is not a factor Supermarket type 3 would **increase sales by 251%** when compared to the grocery store. When Supermarket 2 is compared against Supermarket 3, **Supermarket 1 gets fewer sales by 75%**. When Supermarket 1 is compared against Supermarket 3, **Supermarket 2 gets fewer sales by 57%**

2.) What type of city will return him the best sales: Tier 1, 2 or 3.

The **best city to build the supermarket type 3 is in city tier 1** as there is a decrease in sales in cities tier 2 and 3. **Tier 2 cities see a sales decrease of 1.6% while tier 3 cities see a decrease of 3%.**

3.) What are the top 3 highest performing and lowest performing stores in the sample.

Top: outlet_idOUT013 0.464 , outlet_idOUT035 0.054, outlet_idOUT049

Bottom: outlet_idOUT017 -0.025, outlet_idOUT046, outlet_idOUT027

The spread is small because theses stores may sell need-based items.