

Aryan Sharma
Problem Statement

Download the file "[SnackChain.xlsx](#)". This file contains real sales and promotions data from a large retail chain (of 79 stores) on 58 products belong to four product categories (bagged snacks, cold cereal, frozen pizza, and oral hygiene products) from multiple manufacturers like Frito Lay, Kellogg, and General Mills, over a period of 156 weeks. The data comes from three database tables: stores, products, and transactions, as shown in different tabs in the spreadsheet; the different tables and fields are shown in the Glossary tab. It has over 500,000 transactions and the file is 33 MB in size.

Note that this is exactly how data is structured in a corporate database. You will have to join data from the different tables (using foreign keys) to run the analytics. This is easy to do in R and R handles big data just fine. We are not interested in oral hygiene products; so you can drop this product category and all associated transactions.

The transaction table has weekly information on price and promotions of products (e.g., whether a product was assigned a special store display, or was in an in-store circular, or had a temporary price reduction). We want to examine the effects of these pricing and promotion strategies on multiple dependent variables. You have to control for a variety of different things (to be determined by you), perhaps do some feature engineering, and run different models (or sets of models). At the end of your analysis, we are interested in answers to the following questions:

1. What are the effects of product display, being featured on in-store circular, and temporary price reduction on product sales (spend), unit sales, and number of household purchasers?
2. How do the effects of display, feature, and TPR on SPEND vary by product categories (cold cereals, frozen pizza, bag snacks) and store segments (mainstream, upscale, value)?
3. What are the five most price elastic and five least price elastic products? Price elasticity is the change in units sold for change in product price.
4. As the retailer, which products would you lower the price to maximize (a) product sales and (b) unit sales, and why?