Joshua Nichols

Capstone 1 proposal

1. Problem: Is spending money on promotional displays/prominent placement in a store worthwhile for brand name distribution companies, or does shopper behavior show that products would be purchased regardless?

2. Client: The client would be food distribution companies. The data gathered would allow them to decide whether paying additional fees for displays and placement in stores is a worthwhile investment with regards to ROI.

3. Data: The dataset is downloadable here: https://www.dunnhumby.com/sourcefiles

4. After cleaning the data, I would like to parse through many of the variables to determine if there is a correlation between the two. In doing so, I should be able to find the variables that have the strongest correlation. From this I should be able to create a story using graphs which depicts whether or not prominent placement in stores is beneficial for pasta, sauce, pancake mix, and syrup.

5. Deliverables: I will be able to provide my cleaning/scrubbing code, as well as any code that was used to parse through the data and create any graphs. Additionally, I will be able to provide a summary of the findings in paper format, as well as additional slides that summarize the findings.

Data Dictionary

UPC - Standard 10 digit UPC barcode.

Dollar_sales - Amount of dollars spent by the consumer.

Units - # of products purchased by the consumer.

Time_of_transaction - time of transaction expressed in military time.

Geography - Distinguishes between two large geographical regions. Each region contains portions of several states, and is expressed as a 1 or 2.

Week - Week of the transaction, 1-104.

Household - Identifies unique households.

Store - Identifies unique stores.

Basket - Identifies unique baskets/store trips.

Day - The day of the transaction, 1-728.

Coupon - Coupon usage - 1 if used, 0 if not.

Store_zip_code - 5 digit zip of the referenced store.

Product_description - Description of product.

Commodity - Pasta, Pasta Sauce, Pancake Mix, or Syrup.

Brand - Brand of item.

Product_size - Package size of product.

Feature_desc - Where the product is located in the weekly mailer.

Display_desc - Where the temporary in-store display is located.