An analytics-driven strategy for Store Managers to assess the effectiveness of promotions

ADSP 32024: Data Science for Algorithmic Marketing

Deepak Vanjani, Devdutt Sharma, Marc Edwards, & Peter Fuentes Rosa

Agenda

- Business objective
- Data description and preparation
- Exploratory data analysis
- Model selection and execution
- Model results
- Insights and recommendations
- Future work

Business objective

Design & implement an analytics-driven strategy for Store Managers to determine the most effective elements of promotions between:

1. **Displayed in-store**

2. **Featured in in-store circular**

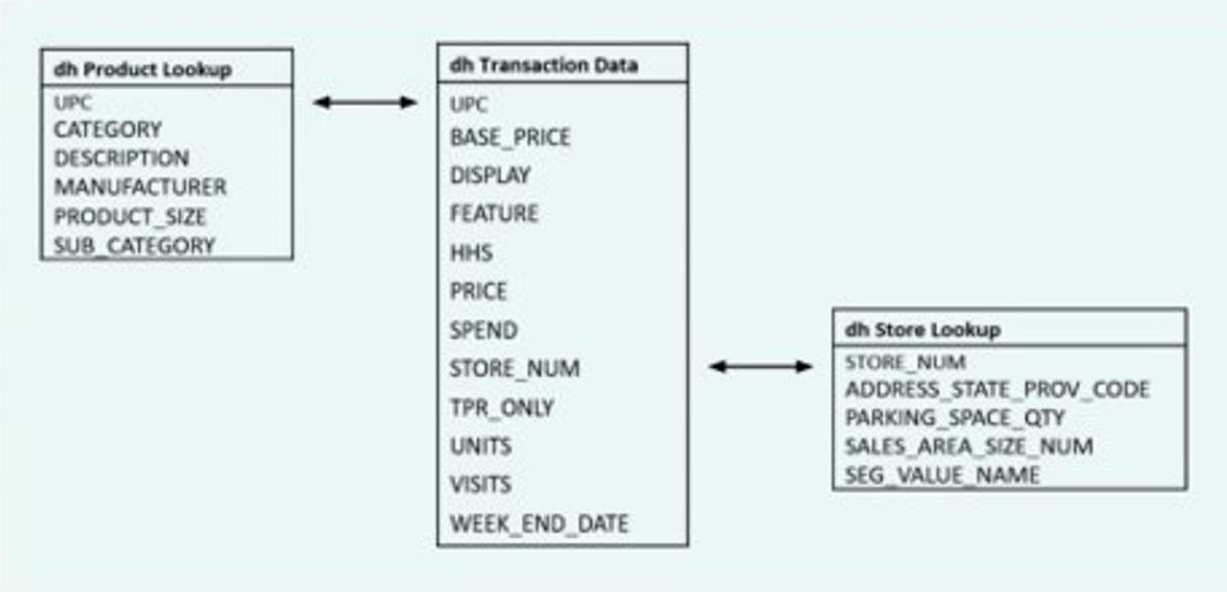
3. **Temporary Price Reduction Only** (TPR) (i.e., product was only reduced in price, not on display or feature)

Data description

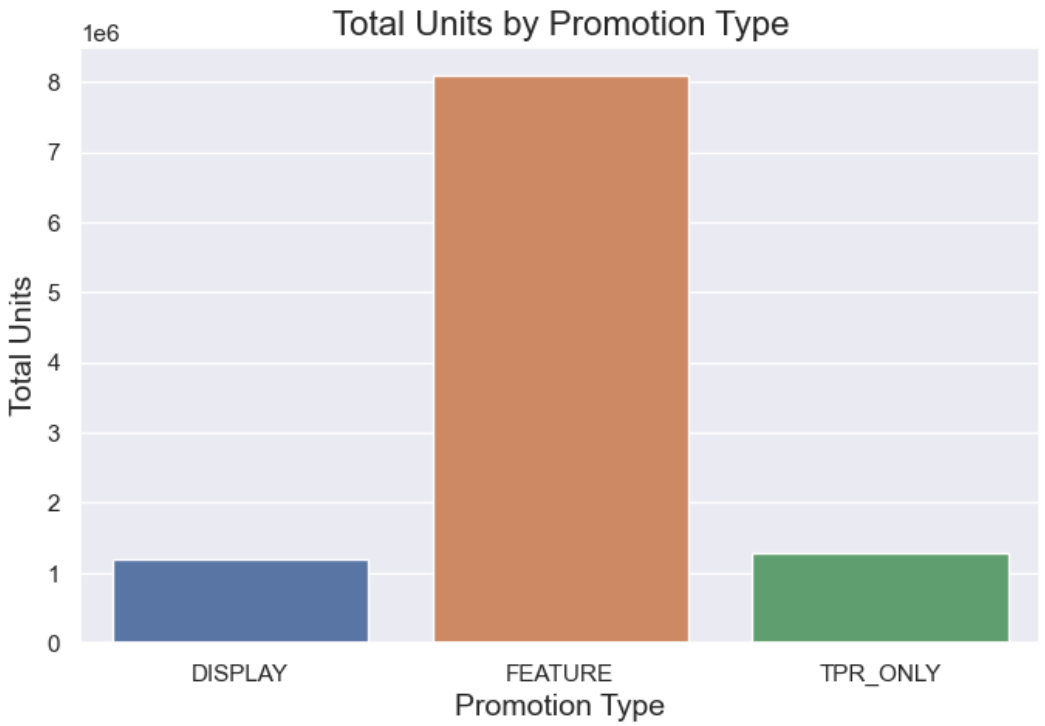
- **Breakfast at Frat dataset by dunnhumby**, containing a representation of sales and promotion information over 156 weeks.
- **Three interlinked datasets:** Store Lookup, Product Lookup and Transaction Data.
- **Time from sample**: 2009-2011
- **Sales metrics of products in 4 categories**: bag snacks, oral hygiene products, frozen pizza, and cold cereal, the top 5 products from each of the top 3 brands in each category
- **In addition, dataset focusing on marketing features that impact**

Data preparation

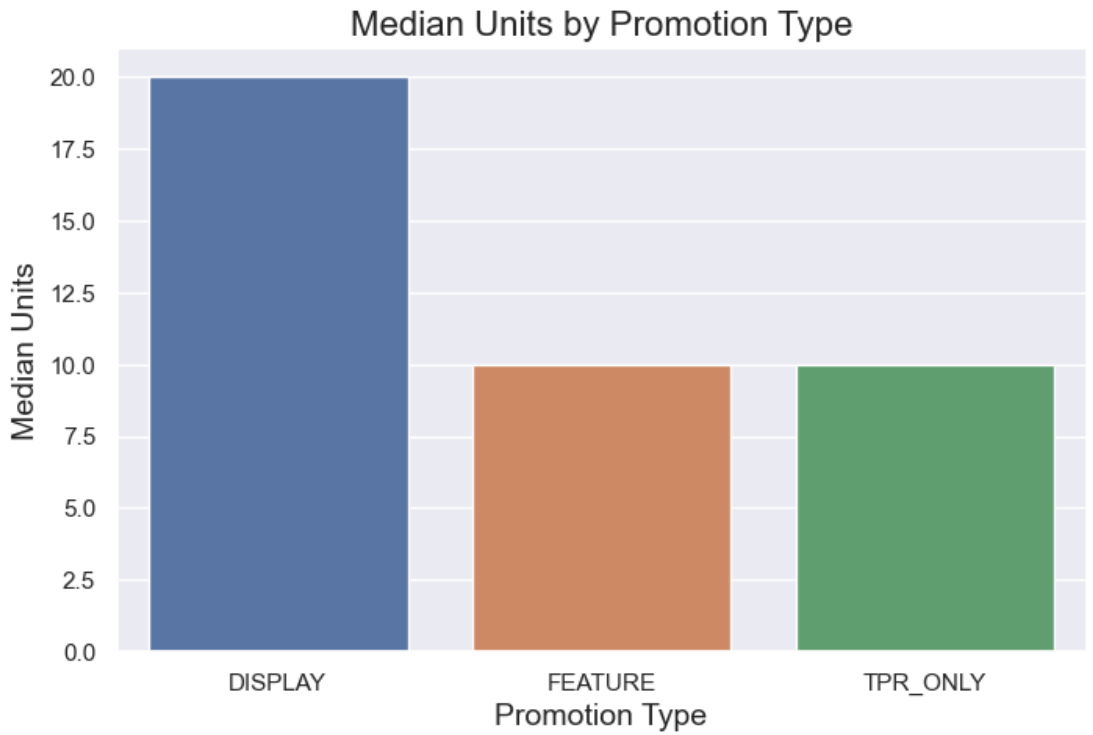
- **Combined three datasets** into one global dataset using unique identifiers (i.e., UPC and Store Number)
- **Addressed missing values** by dropping records for features with <0.1% missing values and converting Nulls to 0 for Parking Space Qty
- **Identified and kept outliers in dataset** due to low materiality (<0.5% of records identified as outlier candidates)
- **Total transactions in final dataset: 538,435**



Insights from exploratory data analysis (1/4)

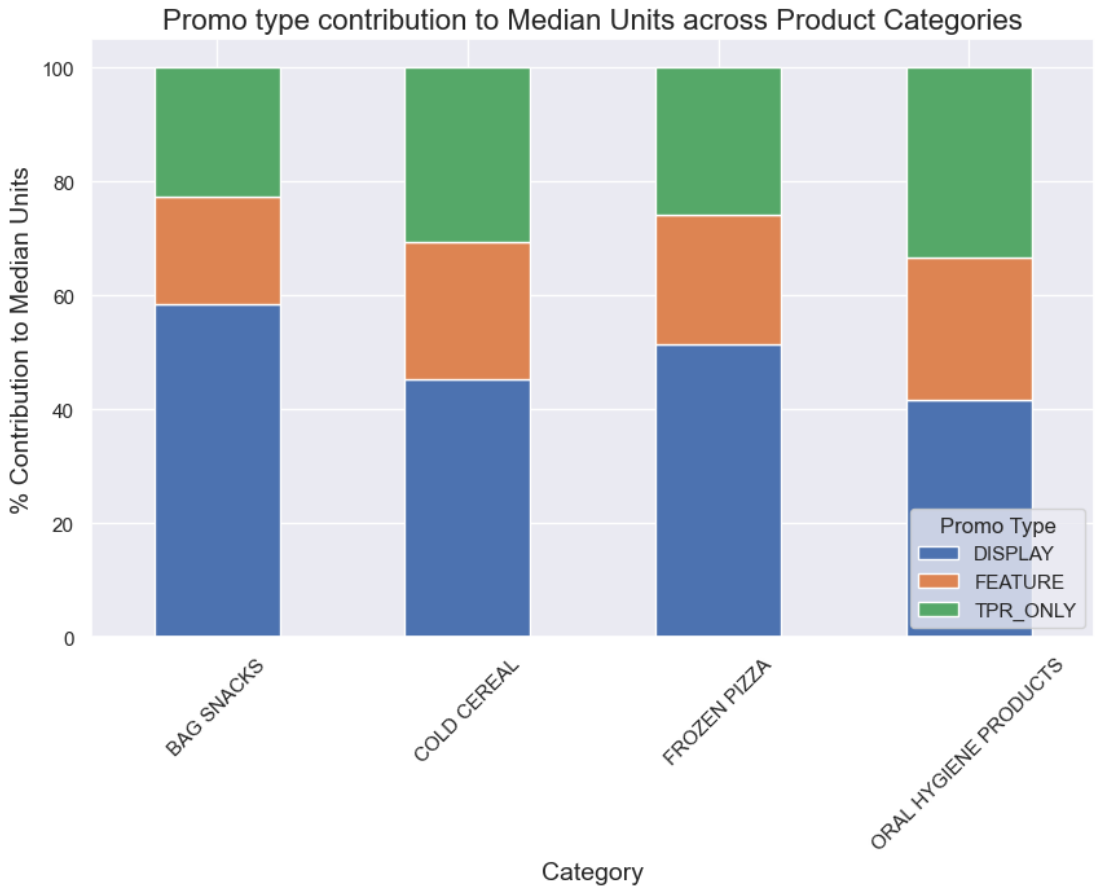


- Feature promotions far outperform Display and TPR promotions based on Total Units, however this result is skewed by the fact 80% of the transactions use Feature promotions
- As such, Median Units can give us a better sense of impact *(see chart on right)*

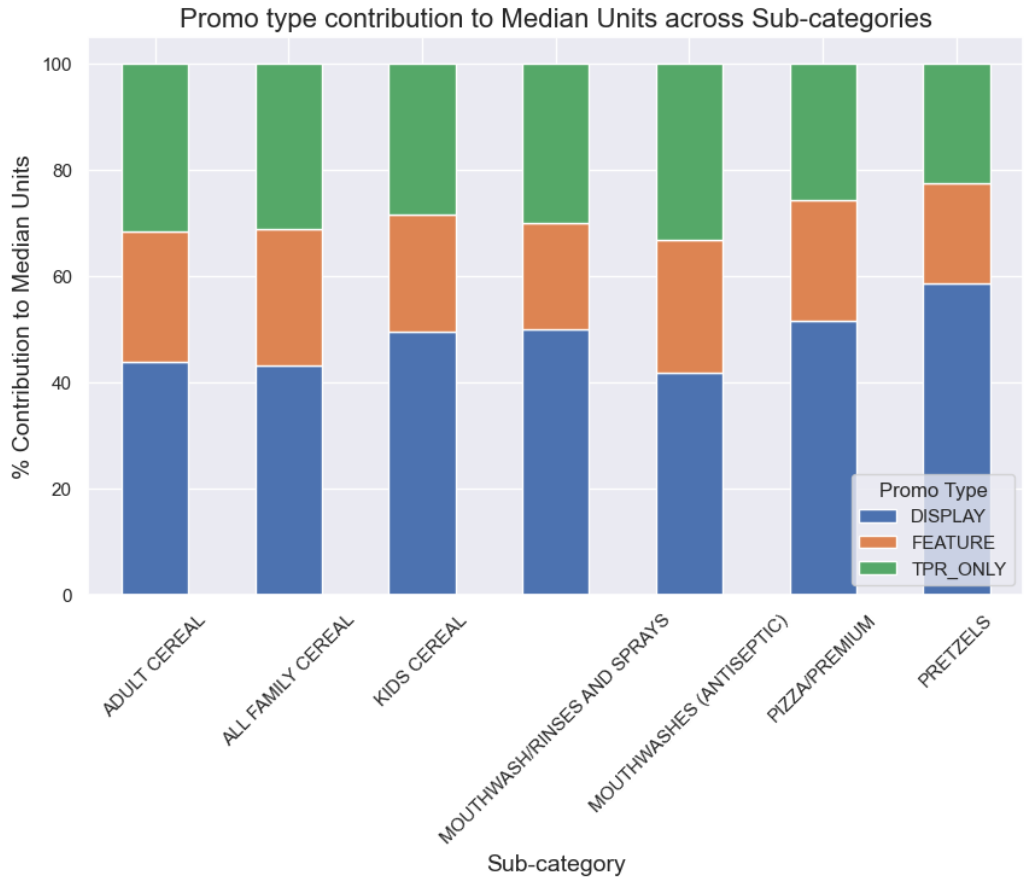


- Display ads result in the highest median units, while Feature and TPR are relatively equal

Insights from exploratory data analysis (2/4)

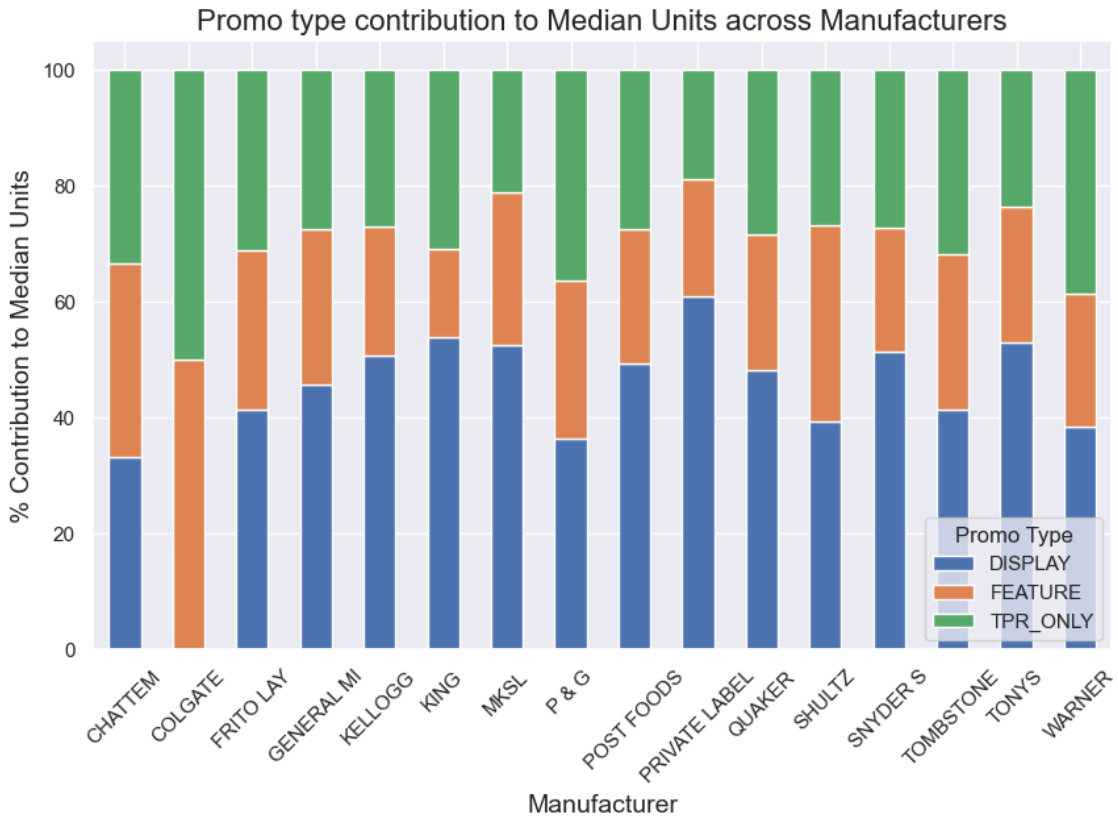


- Differing dynamics across product categories
- While display ads have the highest unit contribution across all product categories, they tend to be the most influential for bag snacks

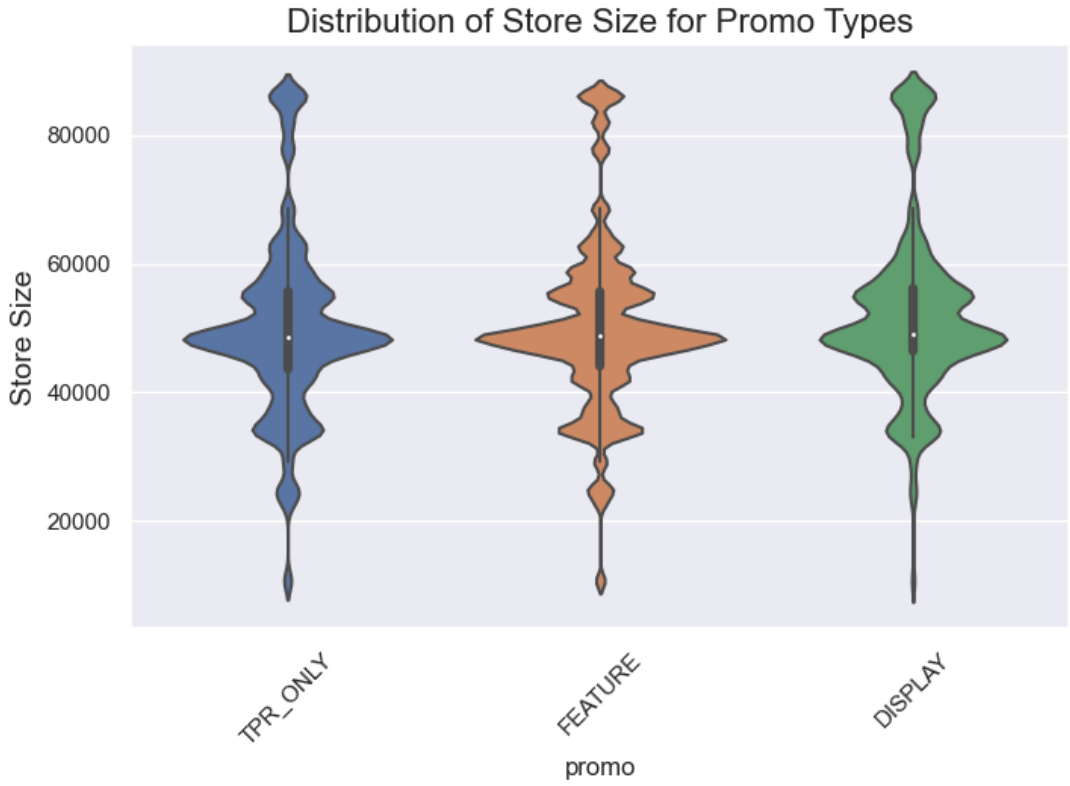


- While display ads have the highest unit contribution across all product sub-categories, they tend to be the most influential for pretzels, and least influential for Mouthwash

Insights from exploratory data analysis (3/4)

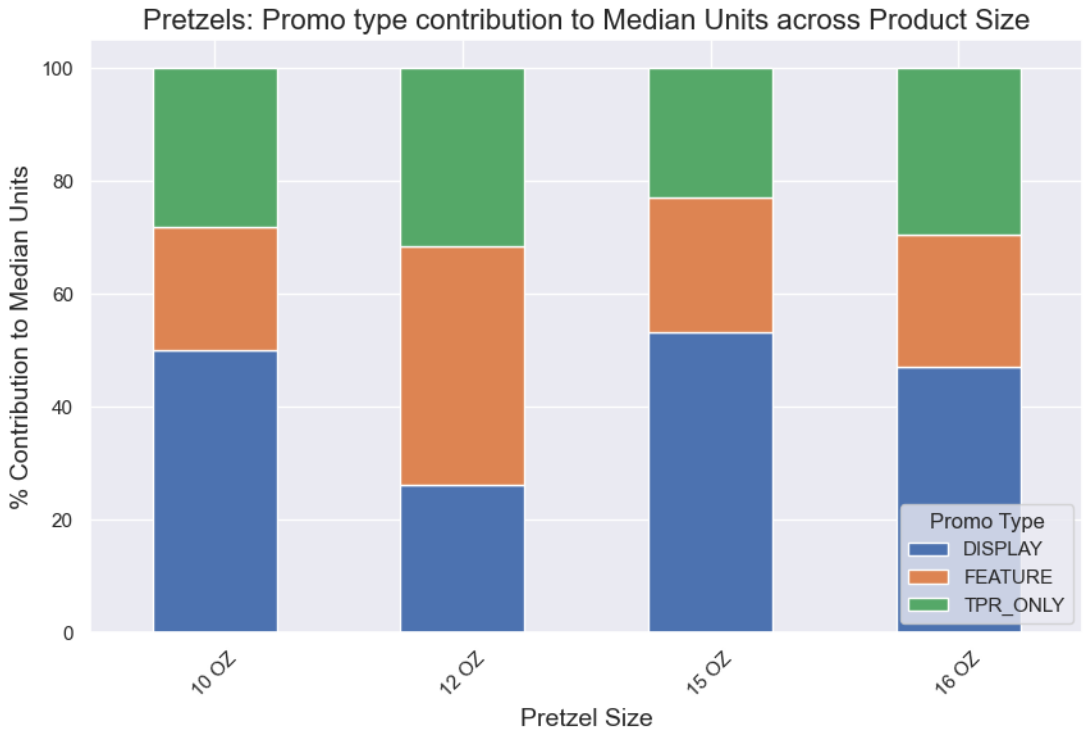


- Private Label products show the highest contribution to units from Display ads
- Colgate products tend to never utilize Display ads

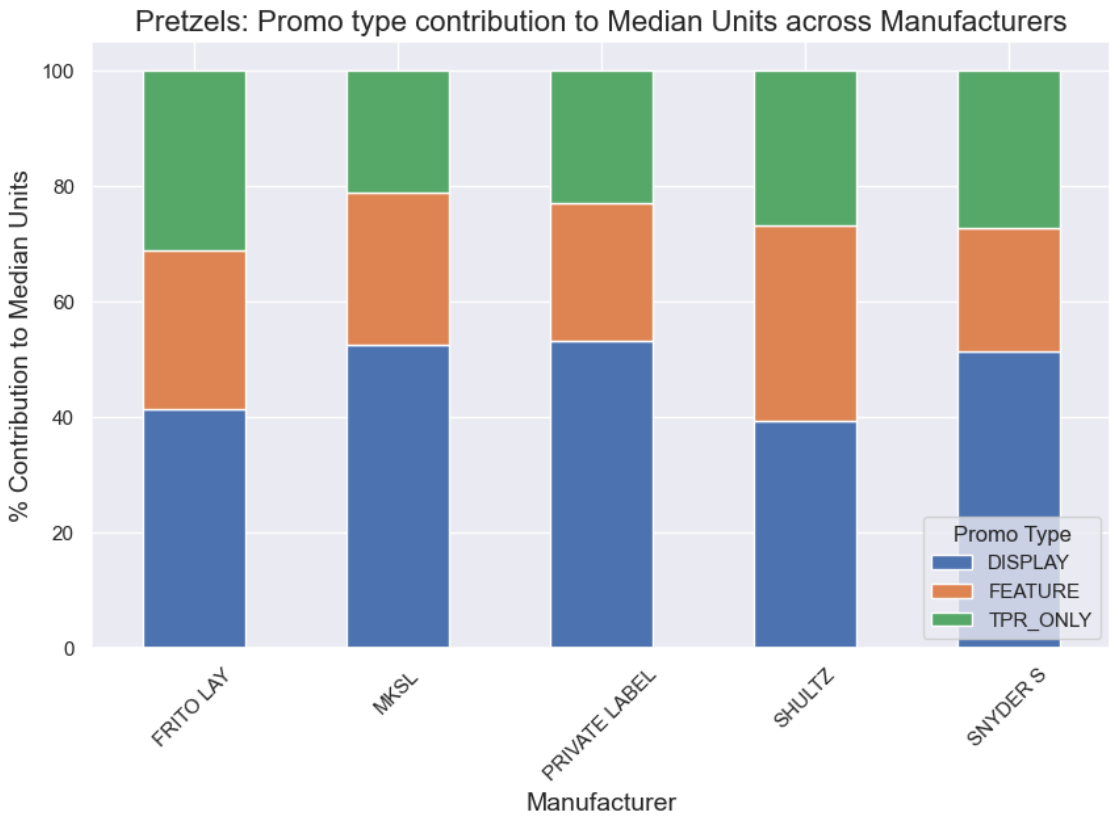


- There is not a noticeable difference for stores sizes across promotion types, indicating that store size may not be an important feature

Insights from exploratory data analysis (4/4)



- Display ads have the most sizable impact on units for the 15 oz model
- For the 12oz model, Display ads are relatively less effective compared to Feature and TPR Only



- Private label and MKSL show the highest relative contribution of Display ads to units, while Shultz shows a more even relative contribution between Display Ads and Feature

Overview of modeling framework

- Model selection process for optimizing marketing promotions.
- Goal: determine the most effective methods – Feature, Display, or TPR (temporary price reduction) for each category (bag snacks, oral hygiene products, frozen pizza, and cold cereal) during the observed period.
- Focus on the Pretzel sub-category, 16 oz. package size for analysis and methodology for each of the 3 brands (Frito-Lay, Shultz, Snyder)

Modeling methodology

Data preprocessing

- Prioritized analysis to be focused on the ‘Bag Snacks’ category, ‘Pretzels’ sub-category, and ’16 OZ’ product size due to being the largest sub-category in terms of total transactions

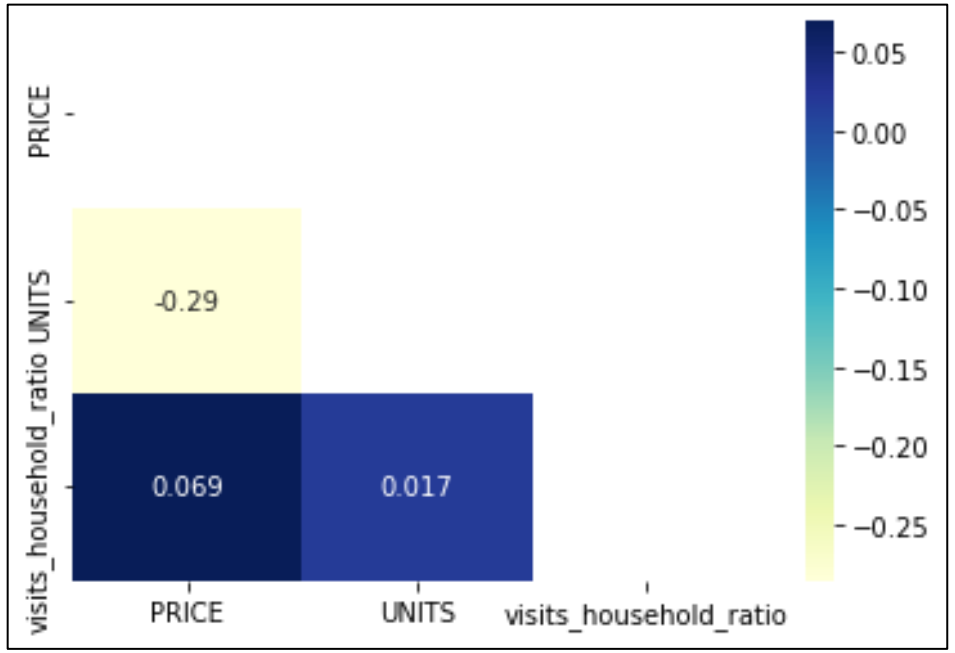
Key features considered

- Month
- Price
- Feature (binary indicator)
- Display (binary indicator)
- TPR only (binary indicator)
- Manufacturer
- Visits to Household Ratio (engineered feature)

Target variable: Units

Modeling methodology

- Log-log model
- Applied log transformation to the numeric variables
- Correlation and variable analysis



Modeling results & recommendations

Simplify the model across the brands using backward elimination by retaining the most relevant predictors for each brand

Brand	Model R^2	Model Adj. R^2	MSE	Feature	Display	TPR Only
Total (All Brands)	0.694	0.692	0.3223	N.S	0.5174	-0.1157
MKSL	0.455	0.441	0.6511	N.S	1.0001	-0.6764
Frito-Lay	0.293	0.276	0.7935	0.7826	0.8627	0.9600
Shultz	0.181	0.161	0.9512	N.S	N.S	-0.3484
Snyder	0.578	0.565	0.4822	N.S	0.7947	0.5469

- Aggregating log-log model across all brands leads to highest predictive power vs. brand-specific models
- Comparing across brand-specific models, the model for Snyder outperforms
- Display tends consistently outperform vs. The other promotion strategies, and therefore should be prioritized in future marketing initiatives
- Brand-specific models show that TPR Only can be a highly effective strategy (e.g., Frito-Lay), but can also be detrimental (e.g., MKSL and Shultz)

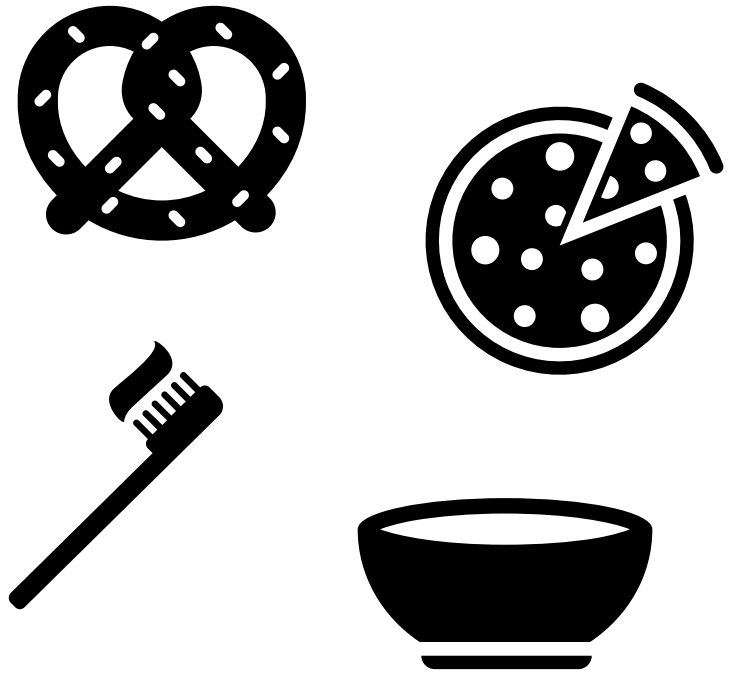
Future work

1. Apply our methodology to find best model and strategy for every combination of product category & sub-category

2. Refine model to incorporate aggregation where possible to avoid having an insurmountable number of models driving promotional strategy

3. Source more recent data

4. Test additional models such as random forest, to search for additional insights and provide better fit for log-log modelling



Thank You