UVA DATA SCIENCE

RETAIL STORE SALES PREDICTION

UVA School of Data Science Capstone Project

05/03/2023

UVA School of Data Science year-long capstone project focused on real-world application of methods learned throughout the 2-year dual degree program

The semester will be structured around the following goals:

- Re-establish **communication plan** with your client if you have been out of communication.
- Establish **a meeting schedule** for your team based on the semester's course schedule, etc.
- Summarize what you learned last semester by creating a **brief project report** including **a scope statement**, a list of **requirements**, and **budget** if necessary.
- Complete the process of **establishing your data and preparing it** for the modeling and/or deep exploration phase. This will include various forms of data reduction, feature engineering, and data shaping to meet the computational requirements of your models and tools.
- **Develop and test one or more models, or other data products**, that meet your project requirements and expectations; revise these expectations if necessary.
- **Produce a six-page publishable paper**, along with an oral presentation, on your work summarizing your team's research and findings. This will be preceded by an **abstract** due after spring break.
- **Package and share any data products and research artifacts** for clients and other stakeholders.

Source: UVA School of Data Science

Shreyas Adiyodi
MBA/ MSDS '23

Nidhi Shah
MBA/ MSDS '23

Ben Wilson
MBA/ MSDS '23

The Cross-Industry Standard Process for Data Mining (CRISP-DM) framework was leveraged to guide our modeling approach throughout the capstone project

Define & Understand Business Problem	▸ Understand business objectives and associated processes ▸ Translate business problems into analytical approach	**Problem Definition**
Collect & Understand Data	▸ Understand data availability, data quality, data insights ▸ Form hypothesis of expectations	**Data Feasibility**
Explore & Analyze Data	▸ Explore to data to identify associations, anomalies, trends, patterns, and relationships	**Selected & Cleaned Data**
Build & Validate Model(s)	▸ Build & iterate model to refine performance ▸ Conduct sensitivities to evaluate validity of results	**Analytical Results**
Integrate Qualitative Insights	▸ Translate analytical results into business terms ▸ Integrate qualitative insights for well-rounded insights	**Actionable Business Insights**

Source: UVA School of Data Science

Table of Contents

01.

Problem Statement

Predicting sales across 3k+ stores within different countries requires great care in understanding cultural, regulatory, and seasonal differences

- Fluctuations in customer demand can greatly impact the financial performance of a business. Therefore, accurately estimating future sales and customer demand is essential for business growth. Sales forecasting involves predicting the sales or demand for a specific product during a certain period. Our paper demonstrates we are able to use modern machine learning principles to predict sales for a retail drug store chain with 3k stores in 7 countries. In our analysis, we assess how sales are impacted by factors such as promotions, competition, holidays, seasonality, and location.

- Currently, store managers are tasked with predicting their daily sales up to six weeks in advance, and this process can lead to inconsistent results due to individual factors. As such, accurate predictions, especially for perishable or time sensitive items, is incredibly important so increasing sales and growing the business.

Source: Retail store chain

Retail store chain provided 3 datasets with macro level insights into each unique store, although the data requires merging to extract necessary understanding

Dataset Description

- You are provided with historical sales data for 1,115 stores. The task is to forecast the "Sales" column for the test set. Note that some stores in the dataset were temporarily closed for refurbishment.

Data Files

- **train.csv** - historical data including Sales
- **test.csv** - historical data excluding Sales
- **store.csv** - supplemental information about the stores

Data Fields

- **Id** - an Id that represents a (Store, Date) duple within the test set
- **Store** - a unique Id for each store
- **Sales** - the turnover for any given day (this is what you are predicting)
- **Customers** - the number of customers on a given day
- **Open** - an indicator for whether the store was open: 0 = closed, 1 = open
- **State Holiday** - indicates a state holiday. a = public holiday, b = Easter holiday, c = Christmas, 0 = None
- **School Holiday** - indicates if the (Store, Date) was affected by the closure of public schools
- **Store Type** - differentiates between 4 different store models: a, b, c, d
- **Assortment** - describes an assortment level: a = basic, b = extra, c = extended
- **Competition Distance** - distance in meters to the nearest competitor store
- **Competition Open Since** - approximate year and month of the time the nearest competitor was opened
- **Promo** - indicates whether a store is running a promo on that day
- **Promo2** - Promo2 is a continuing and consecutive promotion for some stores: 0 = store is not participating, 1 = store is participating
- **Promo2 Since** - describes when the store participated in Promo2
- **Promo Interval** - describes the consecutive intervals Promo2 is started

Source: Retail store chain

02.

Data Preparation

A light data cleansing process was undertaken to prepare 2.34k null values, merge files, and encode categorical features for regression purposes

Summarization identifies null values within store data

```
#summarize train data
summarize_dataframe(df_train)
```

	Data Type	Missing Values	Unique Values	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Store	int64	0	1115	1017209.0				558.429727	321.908651	1.0	280.0	558.0	838.0	1115.0
DayOfWeek	int64	0	7	1017209.0				3.998341	1.997391	1.0	2.0	4.0	6.0	7.0
Date	object	0	942	1017209.0	942	2015-07-31	1115							
Sales	int64	0	21734	1017209.0				5773.818972	3849.926175	0.0	3727.0	5744.0	7856.0	41551.0
Customers	int64	0	4086	1017209.0				633.145946	464.411734	0.0	405.0	609.0	837.0	7388.0
Open	int64	0	2	1017209.0				0.830107	0.375539	0.0	1.0	1.0	1.0	1.0
Promo	int64	0	2	1017209.0				0.381515	0.485759	0.0	0.0	0.0	1.0	1.0
StateHoliday	object	0	5	1017209.0	5	0	855087							
SchoolHoliday	int64	0	2	1017209.0				0.178647	0.383056	0.0	0.0	0.0	0.0	1.0

```
#summarize test data
summarize_dataframe(df_test)
```

	Data Type	Missing Values	Unique Values	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Id	int64	0	41088	41088.0				20544.5	11861.228267	1.0	10272.75	20544.5	30816.25	41088.0
Store	int64	0	856	41088.0				555.899533	320.274496	1.0	279.75	553.5	832.25	1115.0
DayOfWeek	int64	0	7	41088.0				3.979167	2.015481	1.0	2.0	4.0	6.0	7.0
Date	object	0	48	41088.0	48	2015-09-17	856							
Open	float64	11	2	41077.0				0.854322	0.352787	0.0	1.0	1.0	1.0	1.0
Promo	int64	0	2	41088.0				0.395833	0.489035	0.0	0.0	0.0	1.0	1.0
StateHoliday	object	0	2	41088.0	2	0	40908							
SchoolHoliday	int64	0	2	41088.0				0.443487	0.496802	0.0	0.0	0.0	1.0	1.0

```
#summarize store data
summarize_dataframe(df_store)
```

	Data Type	Missing Values	Unique Values	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Store	int64	0	1115	1115.0				558.0	322.01708	1.0	279.5	558.0	836.5	1115.0
StoreType	object	0	4	1115.0	4	a	602							
Assortment	object	0	3	1115.0	3	a	593							
CompetitionDistance	float64	3	654	1112.0				5404.901079	7663.17472	20.0	717.5	2325.0	6882.5	75860.0
CompetitionOpenSinceMonth	float64	354	12	761.0				7.224704	3.212348	1.0	4.0	8.0	10.0	12.0
CompetitionOpenSinceYear	float64	354	23	761.0				2008.668857	6.195983	1900.0	2006.0	2010.0	2013.0	2015.0
Promo2	int64	0	2	1115.0				0.512108	0.500078	0.0	0.0	1.0	1.0	1.0
Promo2SinceWeek	float64	544	24	571.0				23.595447	14.141984	1.0	13.0	22.0	37.0	50.0
Promo2SinceYear	float64	544	7	571.0				2011.763573	1.674935	2009.0	2011.0	2012.0	2013.0	2015.0
PromoInterval	object	544	3	571.0	3	Jan,Apr,Jul,Oct	335							

4 transformation steps taken to prepare data

Preparing the dataset for modelling requires key steps to verify the optimal performance is achievable. The following steps were taken to prepare the data files for the modelling step:

- **Replace Nan Inputs** – The NaN input reviles a null value within the cell. Such values can be found for the below fields are were replaced with a '0' in the updated data, indicating that there was for example no promotion.
  - Open, Competition Distance, Competition Open Since Month, Competition Open Since Year, Promo 2 Since Week, Promo 2 Since Year, Promo Interval
- **Deploy Ordinal Encoder** – The categorical features such as store type require transformation into a numerical value for regression purposes, so an ordinal encoder was utilized for the conversion.
- **Merge Files** – The 'Store' file was merged with both the 'Train' and 'Test' files to verify that all useful information was incorporated into the modelling process.
- **Remove Date** – For the final modelling step, the 'Date' field was removed as it would cause errors when running a regression based model. To verify the information is incorporated into the modelling process, it was divided into day, month, and year.

Source: Team analysis

03.

Exploratory Analysis

Exploratory analysis covered internal store decisions and external market influences to quantify trends and impacts to sales

Internal Decisions

Correlation of Attributes
Quantify degree of association between variables and sales to assess relationship

Store Type
Compare four store types to determine size and performance

Store Comparison
Compare individual stores to quantify differences and outliers

Promotion Type
Evaluate success of different promotions performed to generate sales growth

Exploratory Analysis

External Influences

Seasonality
Verify seasonality trends in sales by year, month, and week across stores

Holiday
Gauge impact to sales from store and state holiday trends

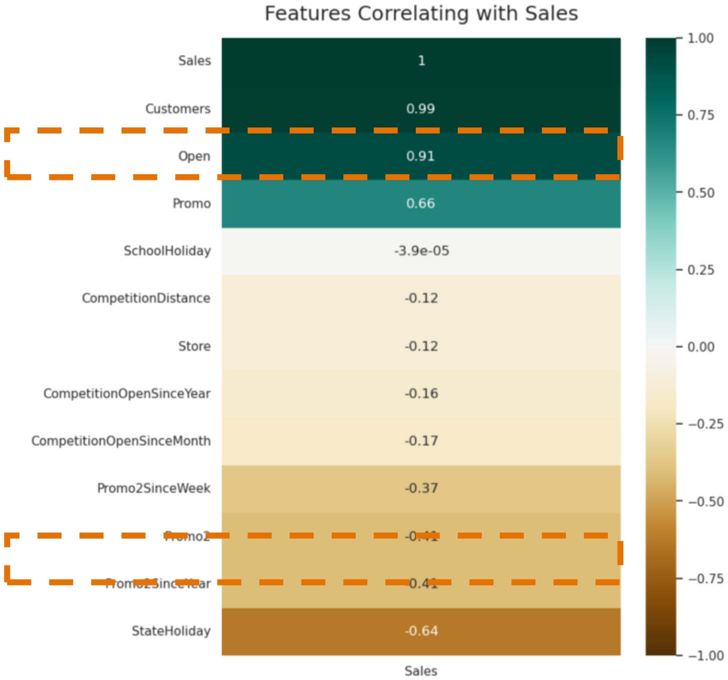
Competition
Sales elasticity impact based upon competition distance and age

Day of the Week
Determine store closures and day of the week sales trends

Source: Team analysis

Strong correlations between 7 data attributes and Sales identifies a clear path for focus when developing models to enhance Sales predictions

Correlation reveals key numerical attributes



7 attributes with strong correlations to guide team

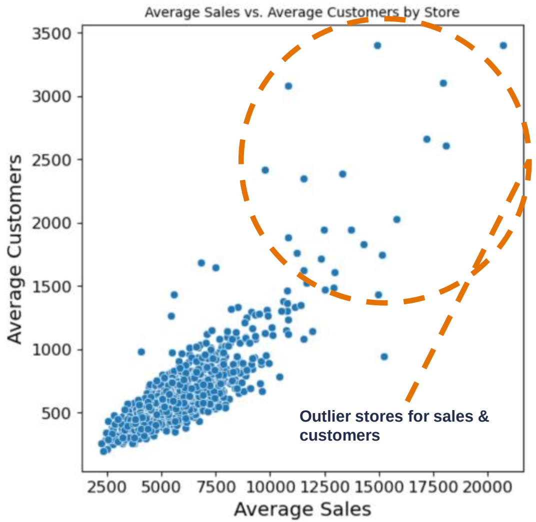
- Customers have unsurprisingly a nearly perfect positive correlation with Sales, indicating that this is a powerful prediction tool for future sales although will be unavailable for predicting in the test data.
- Similarly, StateHoliday has a unsurprisingly strong correlation to sales as stores are unable to sell items when closed.
- Promo 2 indicates a fairly strong negative correlation with Sales, identifying a takeaway that the second promotion offered by stores was largely ineffective at increasing sales. This is quite different from Promo where the first promotion has a very strong correlation with a sales increase.
- An interesting takeaway is the low correlation of the competition features to sales, as one may believe that the closer a competitor is or the longer said competitor has been in business, the lower sales would be. Although this is the case, the correlation is fairly weak.

UVA DATA SCIENCE

Source: Team analysis

Store type B, although fewer stores, consistently outperforms others with $4k greater sales and 1k+ greater customers on average

Outlier stores create comparison issues

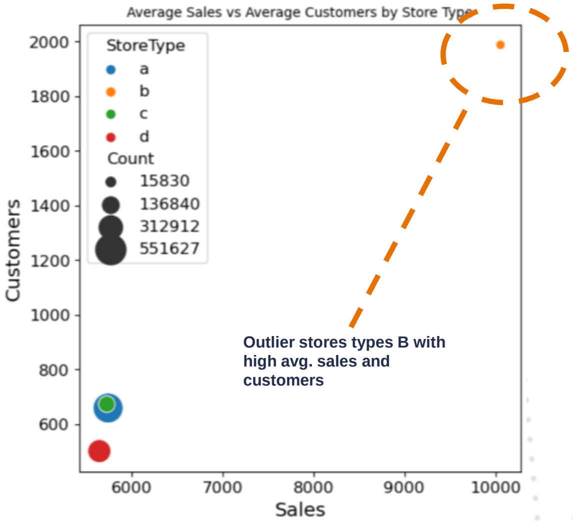


- 20+ stores see above average sales and customers, driving major growth of sales predictions

D stores consistently perform lowest

- Store type D performs below average compared to the other 3 store types in customers and sales

B stores outperform by a wide margin

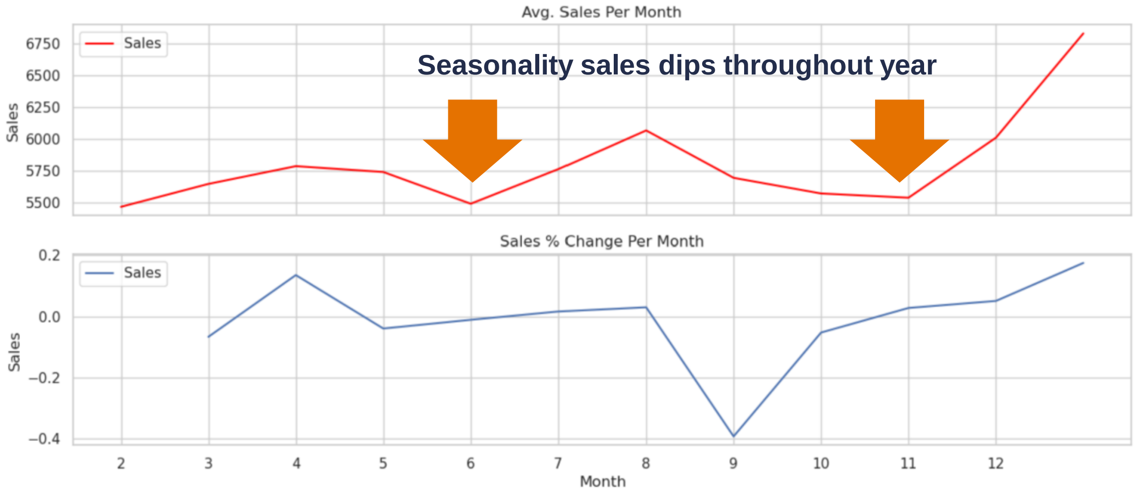


- Store type B, although limited in number of stores, has a high # of customers per sales, indicating low basket size per customer

Source: Team analysis

Monthly sales seasonality is present with 2 significant swings down and up spaces consistently throughout the year

Seasonality negatively impacting sales in June and November



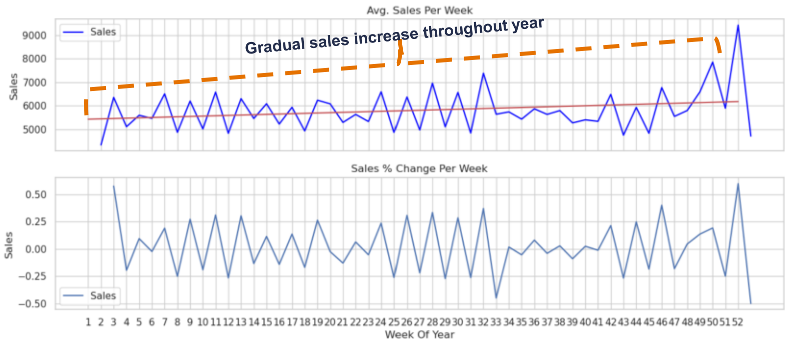
Commentary

- Multiple sales drops and precipitous increases can be seen throughout the monthly sales cycle
- Decrease of November sales of -40% followed by ~10% growth in December shows holiday seasons materially influence sales cycles
- A ~15% sales increase in April should be factored into further analysis to understand cause for future predictions

Source: Team analysis

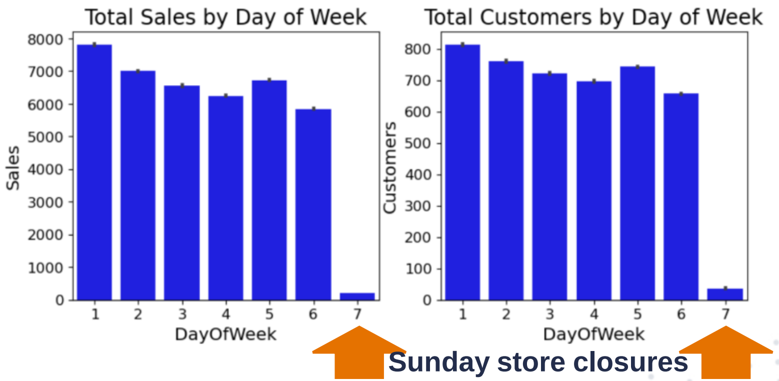
Sunday store closures significantly influence weekly sales cycle for stores predictions

Weekly sales changes correlate to promotions



- Slight sales increase throughout the year can be seen from regression line, indicating sales predictions at the end of the year should be progressively greater
- Large sales swings from spikes in chart influenced materially by promotion schedule and store closures

Sunday store closures leads to Monday spikes



- Gradual decrease in store sales throughout week dependent upon day with most stores closed on Sunday
- Minor sales increase on Fridays (day 5 on chart), indicating end of week purchases by customers

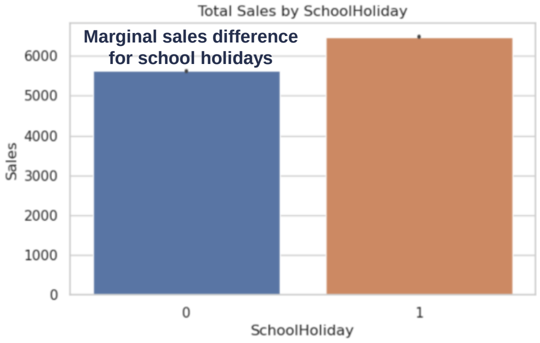
Source: Team analysis

Store closures for state holiday's yet not for school holiday's creates effective prediction point in assessing number of store days open for sales week

State holiday closures factor considerably into prediction

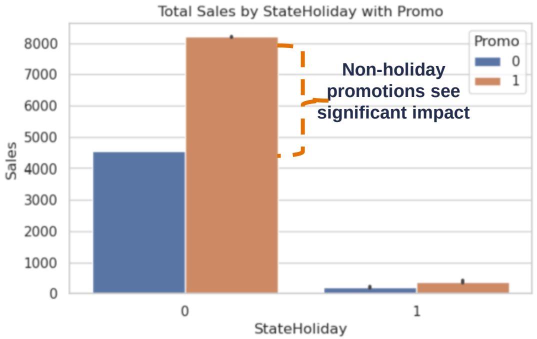
- Similar to Sunday store closures, few stores choose to open on State holiday's, likely causing significant sales bump days before such closures occur
- Predicting store sales x days prior to store closure for holiday purposes could create strong prediction variable

School holiday less applicable in sales prediction



- School holidays show little to no difference in sales between stores, indicating that it is a less powerful attribute to perform predictions with throughout modeling process
- Similar variable importance seen in correlation test

Non-holiday with promotion creates positive interaction effect



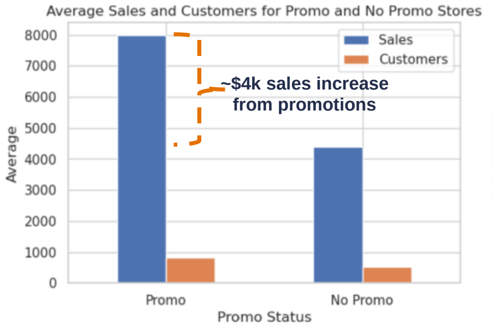
- Promotions are at initial glance an effective driver of sales and become a strong predictor of future store sales
- Few promotions are run on or during holidays given that most stores choose to close

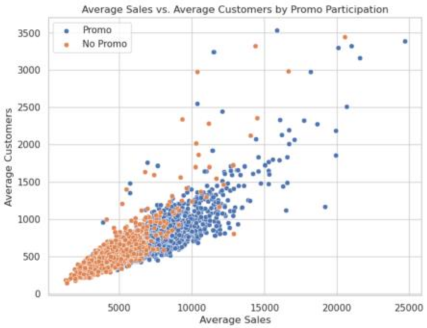
UVA DATA SCIENCE

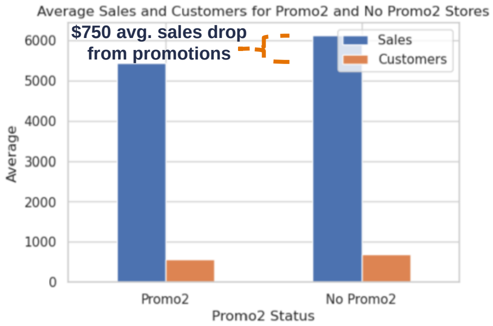
Source: Team analysis

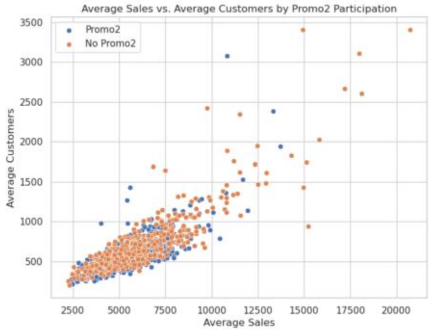
A $4k average sales increase from promo 1 versus $750 average sales decrease from promo 2 initially indicates promo 2 should be removed while promo 1 continues

Promotion type (1 vs. 2) can significantly impacts sales with an immaterial impact to customer count









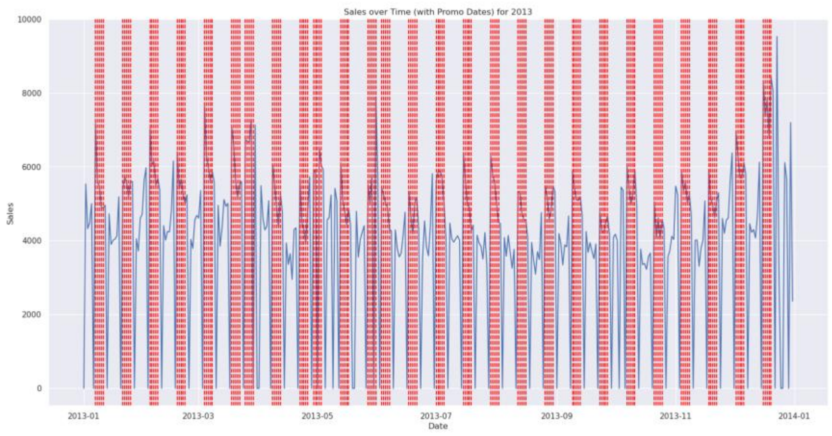
27 more stores engage in consecutive store promotions (promo 2)

- A ~$4k average sales increase is achieved from participating in the single promotion (promo 1)
- The single promotion appears to be a success in driving sales growth although cannot say unless further analysis is conducted on the promotion level (i.e. discount on items)
- 571 stores participated (compared to 544 that did not participate) in consecutive promotions (promo 2) yet see a decline of ~$750 on average sales from doing so
- The consecutive promotion appears to be unsuccessful as a discount was provided to items yet drove fewer average sales than without the promotion

Source: Team analysis

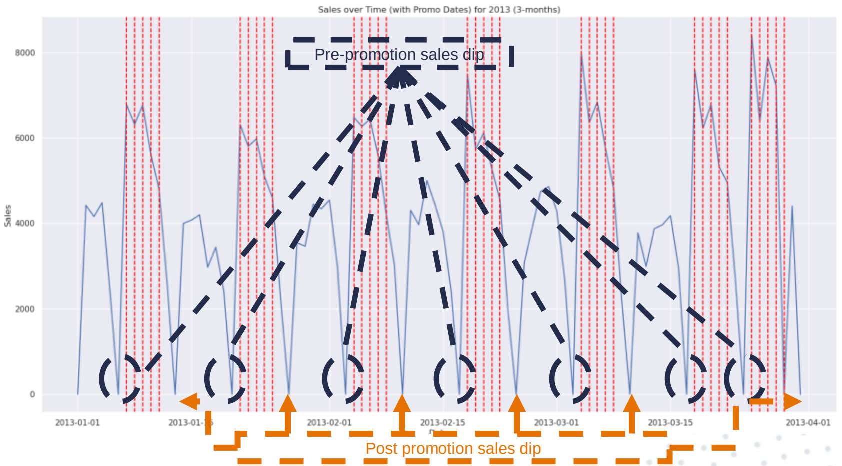
Sales increases from promotion periods may be ineffective due to customers predicting promotions will occur and buying items in bulk until future promotion occurs

Consistently scheduling store promotions…



- Consistently planned promotions and consecutive promotions can be seen across the 2013 year with slight sales dips from seasonality
- Clear trends appear pre and post promotion periods depicting clear customer buying patterns

…leads customers to successfully predict discounts

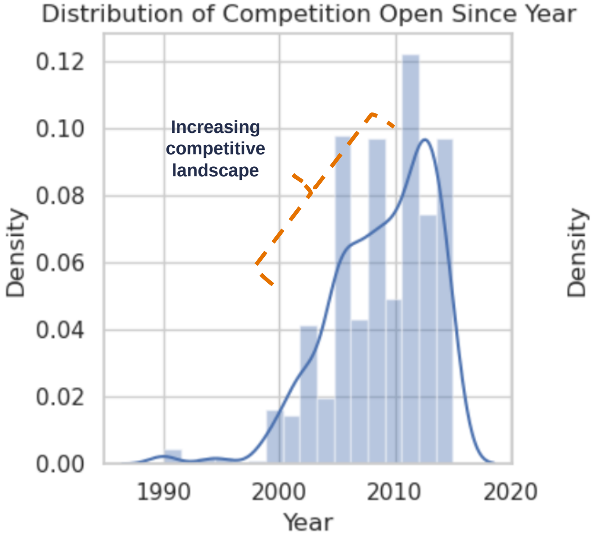


- An apparent pre promotion sales dip indicates customers predict the promotion will occur and choose to not buy until the promotion takes effect
- Similarly, a post promotion sales dip occurs from customers buying items in bulk to hold over until the next promotion occurs

Source: Team analysis

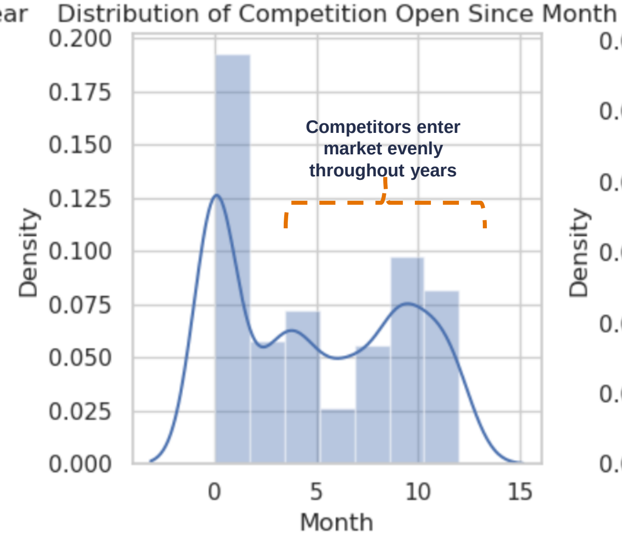
As more competitors enter the market in recent years and locate geographically close to retail stores, understanding competitor dynamics will be crucial to predict sales

Increasing competitive landscape in 15-year period



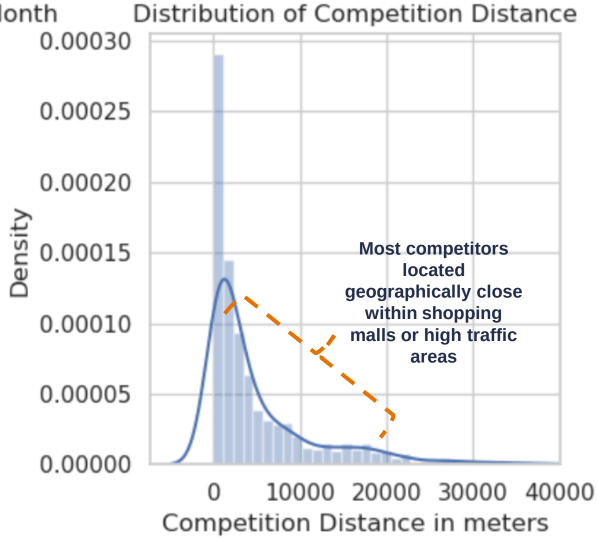
- Peak competitors entering market between 2007 – 2013, with competition increasing aggressively until data cutoff

Competitors not entering the market in seasonal approach during years



- Customers do not consistently choose a specific month throughout the year to enter the market

Geographically close competitors create larger competitive environment



- As competitors enter the market, the desired choice is to enter closer to the drug stores which only increases competitive market

Source: Team analysis

As we build the prediction model, 5 noticeable takeaways from EDA will guide our approach to informing enhanced sales forecasts

Promotion Prediction	▸ Customers ability to predict sales promotions, leading to a steep drop pre and post promotion, shows clear patterns to use in future forecasts.
Ineffective Promotions	▸ Promotion 2 should be considered, on average, a net negative to future sales for predicting impact.
Geographically Close Competition	▸ Competition is growing aggressively in the retail store market, especially as competitors choose to locate themselves in close proximity to the retail stores.
Closed Operations in Week/ Year	▸ Store closures on Sunday's and due to state and/ or store holiday's negatively impact sales and should be understood when forecasting future sales.
Gradual Sales Increase In Year	▸ Although quite minor, the gradual sales increase throughout the year should be integrated into the model to more accurately assess sales.

Source: Team analysis

04.

Modelling Performance

Our modeling approach integrates five models into a single ensemble for enhanced predictive performance overtime to limit model bias and increase robustness

Cross-Validation

- Performed cross-validation utilizing a 75/25 train/ test split
- Training data separated into smaller train and validate for purposes of training individual models

Individual Performance

- Individual model performance based upon five model types
- Feature importance per model assessed to determine optimal features

Ensemble Performance

- Optimal models weighted based on linear regression
- Weights utilized to determine final ensemble model for Kaggle competition in sales prediction

Evaluation Metrics

- RMSE
- MAE
- R2
- RMSLE

Source: Team analysis

Each of the five models brings a unique strength to the ensemble that balances to an improved long-term prediction tool for the drug stores

Generate unique models with varying strengths to consolidate into a single ensemble that emphasizes the strengths of each type

Key Modelling Infrastructure Components

Linear Regression

Simple, interpretable model that assists in identifying linear relationships present

Random Forest

Ensemble of numerous prediction trees to overcome decision tree limits and outliers/ noise

Light GB Model

Optimized approach for the numerous features within the semi high dimensional data

Decision Tree

Capture complex interactions between model features, especially with categorical features

XGB Model

Take advantage of regularization component to reduce risk of overfit and assess feature importance

Ensemble Model

Combine strengths across optimal models to enhance accuracy and interpretability

Out a robust model that weights each input model based upon a linear regression assessment

Source: Team analysis

Prediction scores greatly enhanced as more advanced models were trained, providing a top R^2 score of 96% from the Light GB model

	Linear Regression	Decision Tree	Random Forest	XGB	Light GB
Training Accuracy – RMSE	2,543	1,644	1,024	1,704	743
Training Accuracy – MAE	1,771	1,013	629	1,180	503
Training Accuracy – RMSLE	1.75	0.23	0.16	1.52	1.45
Training Accuracy – R^2	0.56	0.81	0.92	0.80	0.96
Public Score	11.65645	0.26695	0.22736	0.3014	0.13938
Private Score	11.93165	0.27615	0.22954	0.30909	0.16726

Feature Importance

Linear Regression

Feature	Importance
Promo2	56846.077699
Open	5569.212953
Promo	2080.024105
StateHoliday	1091.166099
Assortment	344.645348
Season	174.225658
Year	163.870786
DayOfWeek	140.172291
PromoInterval	105.817995
Month	62.485700
StoreType	51.430653
CompetitionOpenSinceMonth	38.323945
SchoolHoliday	30.155207
Promo2SinceYear	28.686405
Promo2SinceWeek	16.327543
WeekOfYear	6.501328
Day	2.798275
CompetitionOpenSinceYear	0.173394
CompetitionDistance	0.024131
Store	0.023870

Decision Tree

Open	0.549097
Promo	0.086883
Store	0.066344
CompetitionDistance	0.060514
CompetitionOpenSinceMonth	0.045516
DayOfWeek	0.035516
StoreType	0.032657
CompetitionOpenSinceYear	0.031030
Promo2SinceYear	0.018447
Assortment	0.015627
Promo2SinceWeek	0.012269
WeekOfYear	0.009878
Day	0.008620
PromoInterval	0.006902
Month	0.006506
Promo2	0.005976
Year	0.003644
SchoolHoliday	0.001963
Season	0.001877
StateHoliday	0.000735

Random Forest

	Importance
Open	0.489435
CompetitionDistance	0.102646
Store	0.090986
Promo	0.077423
CompetitionOpenSinceYear	0.039730
DayOfWeek	0.037324
CompetitionOpenSinceMonth	0.032414
StoreType	0.019408
Promo2SinceYear	0.017431
Day	0.017136
Assortment	0.016915
WeekOfYear	0.016292
Promo2SinceWeek	0.015351
PromoInterval	0.008265
Month	0.006802
Year	0.006174
SchoolHoliday	0.002160
Promo2	0.001906
Season	0.001309
StateHoliday	0.000894

XGB

Open	0.447405
Promo	0.127730
StateHoliday	0.079144
DayOfWeek	0.064982
Promo2	0.033898
Promo2SinceYear	0.025658
Store	0.024803
StoreType	0.023501
Assortment	0.022142
CompetitionDistance	0.021107
Promo2SinceWeek	0.019721
CompetitionOpenSinceMonth	0.017656
CompetitionOpenSinceYear	0.017616
Day	0.017330
PromoInterval	0.016267
Month	0.013060
SchoolHoliday	0.010945
WeekOfYear	0.010844
Season	0.003185
Year	0.003007

Light GB

Store	3440
CompetitionDistance	3240
DayOfWeek	2188
CompetitionOpenSinceMonth	1245
Day	1187
WeekOfYear	1185
CompetitionOpenSinceYear	1177
Open	1106
Promo	897
StoreType	843
Promo2SinceWeek	834
Year	764
Month	653
Promo2SinceYear	581
Assortment	573
PromoInterval	322
Promo2	211
SchoolHoliday	197
Season	189
StateHoliday	168

Source: Team analysis

Performing a linear regression to appropriately weight each model type, a final ensemble model was produced that, although not optimal, balances bias and weaknesses from inputs

Linear Regression	0.01715
Decision Tree	-0.01199
Random Forest	0.34142
XG Boost	-0.14861
Light GB	0.80145
Intercept	4.29812

Ensemble Model

RMSE	MAE
679	459
RMSLE	R^2
1.38	0.96
Public Score	Private Score
0.2236	0.23747

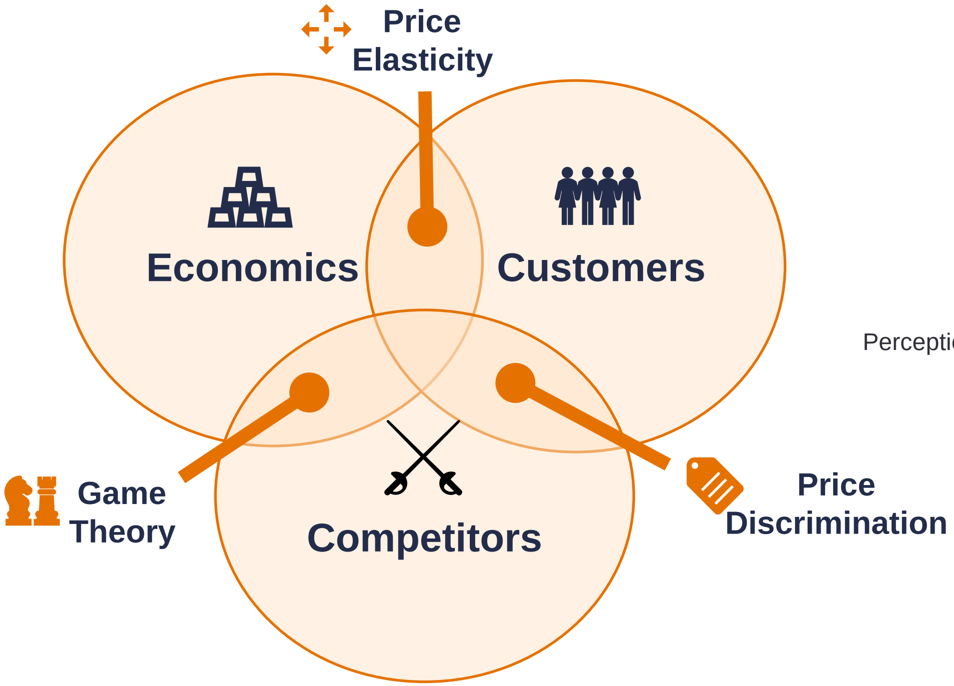
Source: Team analysis

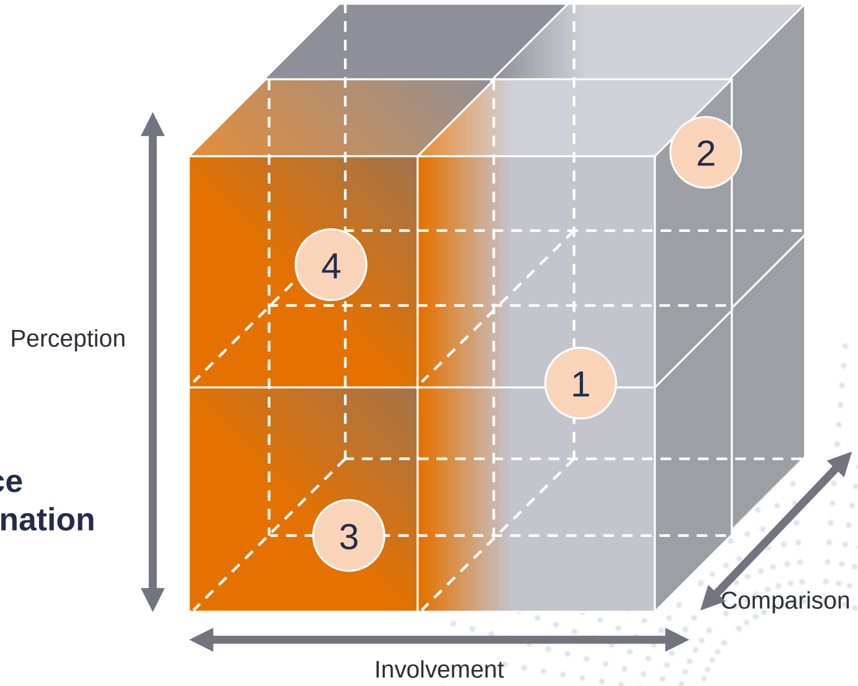
05.

Qualitative Insights

Optimized pricing and promotion behavior for retail stores requires balancing the 3 lenses to pricing with human natures underlying psychological responses

Three Lenses to Pricing			Psychology to Pricing	
Economics	Customer	Competitor	Involvement Level	Human Nature



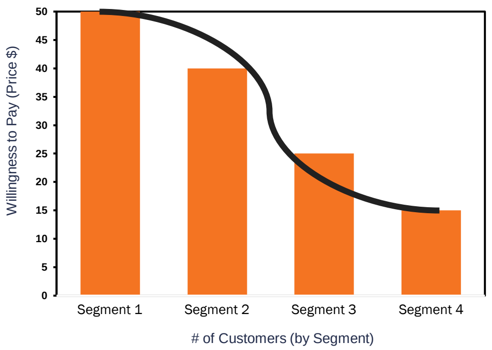


Source: GBUS 8047 - Darden Pricing

Pricing economics are built on 3 core pillars of the segments demand flexibility, stores local market share, and the size of the local channel

Three Lenses to Pricing			Psychology to Pricing	
Economics	Customer	Competitor	Involvement Level	Human Nature

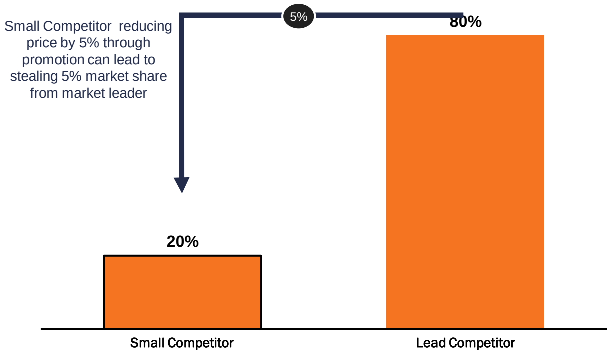
1 Shape of customer demand influences segment elasticity



The drug stores should determine segment and elasticity of segment of customers

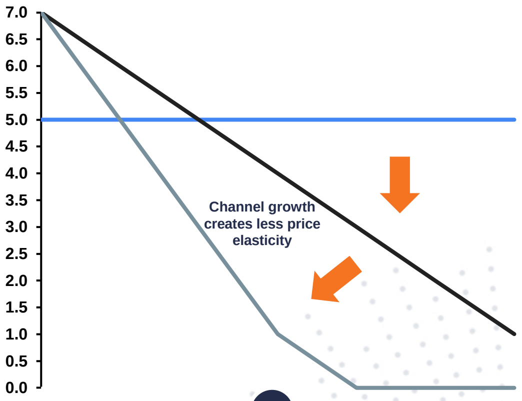
E<0 = Inferior Good | E=0 = Sticky Good | 0<E<1 = Necessity | E>1 = Luxury Good

2 Market share is a strong price elasticity predictor



The retail stores require an enhanced understanding of each locations market share compared to competitors to understand promotion effectiveness

3 Channel size effects elasticity



Smaller channels create more elasticity with prices and promotions which the stores should use in determining how to run future promotions

Source: GBUS 8047 - Darden Pricing

Retail customers are complex, requiring a balance of value drivers per customer segment with the propensity of each to buy

Three Lenses to Pricing | **Psychology to Pricing**

Economics | **Customer** | **Competitor** | **Involvement Level** | **Human Nature**

1 Utilize customer value drivers to enhance experience

Technical Drivers

- Stores features that provide a technical benefit
- Connection to functional and emotional benefits
- KPI: Item availability

Functional Drivers

- How customer experiences the product or service during the purchase process
- KPI: Experiences

Emotional Drivers

- How customer feels about the buying process based upon the brand perception, reputation, personal experience
- KPI: Satisfaction

2 Identify various customer segments to target

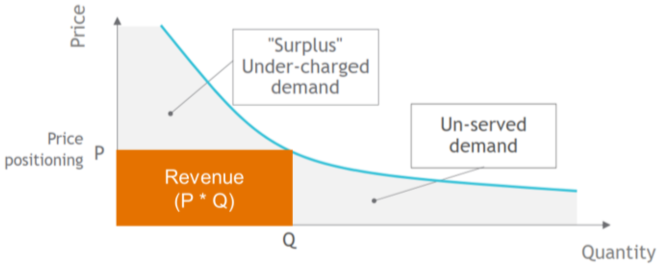
Deep Customer Understanding
- Differentiate customer value
- ID segments
- Articulate winning proposition

Capabilities & Assets
- Technology
- Capacity
- Brand
- Recognition

Customer Access
- Reach each segment effectively
- Market efficiently
- Sell through right channels

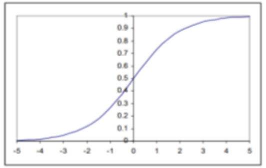
Enabling Economics
- Favorable cost structure
- Critical mass in each segment for scale
- Sufficient pricing power available

4 Assess customer demand curve for opportunities



3 Calculate customer propensity to purchase for segments

$$S_i = \frac{e^{U_i}}{\sum_{j=1}^{n} e^{U_j}}$$



- Leverage conjoint analysis to identify value of attributes to customers
- Determine price applicability on buying behavior and set promotions based upon analysis
- Assess all outputs on utility of product to an individual customer

UVA DATA SCIENCE

Source: GBUS 8047 - Darden Pricing

Competitors have rapidly crept into the retail stores geographic markets, requiring a balance of leading and reacting to price promotions

Three Lenses to Pricing			Psychology to Pricing	
Economics	Customer	**Competitor**	Involvement Level	Human Nature

Three common competitor price models

Price Setting

Setting prices based upon value of next best alternative (i.e. substitute)

Price Moves

Active price management through war-gaming exercises

Price War

Avoid or break the vicious cycle between yourself and competitor

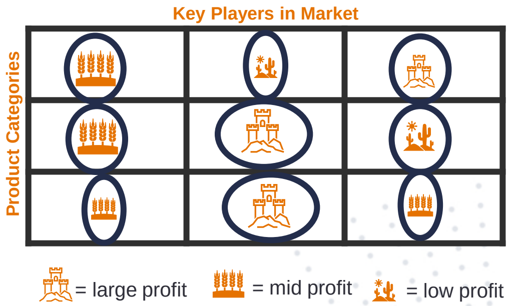
Chart competitive response

	Resulting degree of competitor response: Low	High
Competitor pricing strategy	Conservative	Advanced
Cross-elasticity w/ competitors	Low	High
Competitor's strategy goals	Not in focus	Important
Competitor's margin room	Low	High

As Rossman competitors align promotions:

- Do not initiate aggression to price war
- Make response clear so that competitor does not misunderstand
- Practice reciprocity if competitor is aggressive

War-game competitive landscape



Rules of thumb to approach wargame results:

- Protect the castle by lower prices or running promotions
- Attack competitor castles to gain market share/ increase sales
- Avoid price reduction in grassland (capture adjacent grassland)
- Do not respond to price wars in deserts

Source: GBUS 8047 - Darden Pricing

Customers have low visibility and involvement with drug store items, requiring competitive prices and promotions to retain the base against competitors

Three Lenses to Pricing			Psychology to Pricing	
Economics	Customer	Competitor	Involvement Level	Human Nature

Customers make 4 types of purchase decisions…

…determined by external visibility & internal mental involvement…

Involvement

- Mental energy and time expended by customers in considering product purchase
- Unconscious Decision (Low) vs. Complex Mental Thought (High)

Visibility

- Level of visibility people have when consuming or using the product
- Inconspicuous (Low) vs. Conspicuous (High)

…that the client should assess to understand their customer patterns.

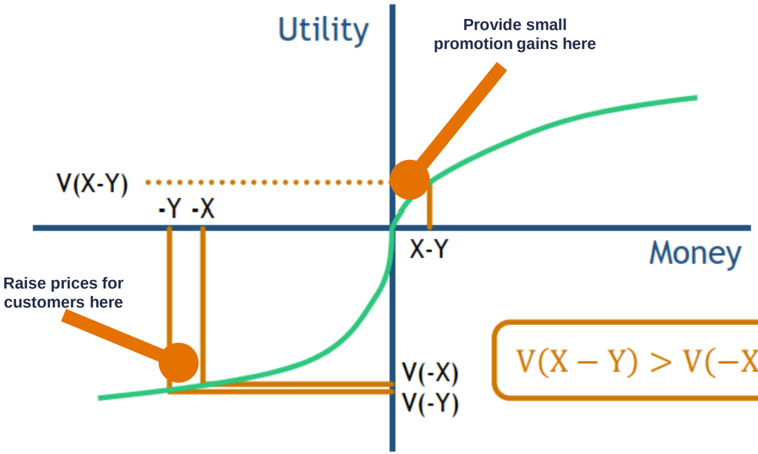
- Customers have low involvement in the decision to buy with primarily low visibility for drug store items
- Customers have little time or mental energy expended on the decision to make store purchases
- Price thresholds and price points significantly influence customer purchases
- Price is a crucial factor in determining whether customers make repeat purchases
- Customers actively seek deals to influence buying patterns in future

Source: GBUS 8047 - Darden Pricing

Drug stores should spread smaller, varied promotions for minor utility gains while integrating price raises at once to limit negative utility losses to the end customer

Three Lenses to Pricing			Psychology to Pricing	
Economics	Customer	Competitor	Involvement Level	Human Nature

Utility curves emphasize negative > positive experiences



$$V(X - Y) > V(-X) - V(-Y)$$

Commentary

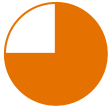
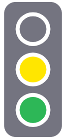
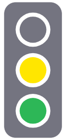
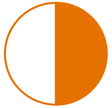
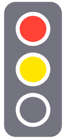
- Our brains see negatives more clearly than positives due to human evolution to survive
- A negative impact to a customer is felt more than an equivalent positive impact due to how we value negative vs. positive utility
- Ex. Price raise of $5 creates more anger for a customer than a $5 promotion creates happiness
- Promotions should be small, spread out, and varied for customers to receive consistently small utility increases (receive greater utility from multiple positive results than all at once)
- Price raises should come all at once toward end of negative utility function where customer experiences diminishing rate (integrate losses)

Source: GBUS 8047 - Darden Pricing

06.

Final Recommendations

Qualitative and quantitative insights have led to four key recommendations surrounding promotions, customers, and competition to enhance future store sales

Key recommendations		Assessment	Expected Impact
Enhance Promotion Effectiveness	▸ Remove consecutive promotions (promo 2) due to ineffectiveness in increasing sales or profit ▸ Randomize future promotions to limit customers predicting promotions and stockpiling goods ▸ Run frequent, smaller promotions to incentivize recurring customer shopping patterns ▸ Bundle all price increases for one time to limit negative utility impact to customers		
Utilize Elasticity of Demand	▸ Segment customer base and evaluate each segments level of demand ▸ Calculate market share and channel size of stores across markets for reviewing pricing power ▸ Utilize inputs to analyze elasticity of demand for customers and set promotions based on result		
Understand Competitive Mindset and Tactics	▸ Assess competitive pricing and promotion strategy in terms of impact to future sales ▸ Determine competitive response to historical store promotions ▸ War-game pricing and promotion landscape to optimize strategy for future response		
Target Customer Value Drivers	▸ Weight three customer value drivers to determine customer preferences ▸ Align store experience and promotions to weighted drivers to increase customer surplus ▸ Enhance customer propensity to buy to increase customer basket size		

Very limited ··· Substantial

Realistic / favorable Challenging Adverse / potentially critical

UVA DATA SCIENCE

Source: GBUS 8047 - Darden Pricing


Thank You.
UNIVERSITY of VIRGINIA
DARDEN SCHOOL of BUSINESS