MSIN0041

DEC 2021

GROCERY & DATA ANALYTICS

MARKETING SCIENCE

NINON LAVOLLE
GIULIA ZHANG
TANIA TURDEAN
GAURI MUNJAL
THAIS PARISOT
MIHNEA MACOVEI
JACK LIMZUN

Word count:1990

Table of contents:

1. INDUSTRY BACKGROUND AND RESEARCH QUESTIONS

1.1 U.S. GROCERY RETAIL INDUSTRY

In the US, many grocery retailers are experiencing falling gross margins due to three major forces: consumers' changing preferences, intensifying competition, and new technologies. (1) Personalization has also gained massive traction, evolving from a predominantly mass promotion–based approach to segmented, customized, and real-time dynamic offers. In 2020, 60% of grocery retailers indicated they had made investments to enhance capabilities to better personalize promotions and pricing. (2)

This report analyses Dunnhumby' grocery data and proposes a set of informed-marketing strategies to help US grocery retailers navigate industry change and achieve profitable growth.

1.2 RESEARCH QUESTIONS

1.2.1 PRICING

Despite wide variations in consumer demographics, most US food merchandise chains charge nearly uniform prices across stores. (3) This report investigates if the same applies to Dunnhumby's grocery data and if there is an opportunity to capture consumer surplus through price discrimination. Our research questions are as follows:

- Do US retailers price discriminate based on geographical region (States and Cities)?
- Is this explained by price elasticity?

1.2.2 PROMOTION

US grocery retailing is a dynamic and highly competitive industry, and it's becoming more so. To stay ahead of the competition, industry leaders have to fully understand their clients and best serve their needs. As consumer needs vary based on their location, geo-targeting is crucial to driving customer acquisition and sales.

Dunnhumby' data presents 3 promotion techniques: *Display Promotion*, *Feature Promotion* and *Temporary Price Reduction* (TPR).

Display	Feature	TPR (only)
• Product was part of in-store promotional display	• Product was in-store circular	• Product was reduced in price but not on display or in an advertisment

Selecting the right promotion technique is important to deliver effective and successful promotional campaigns. Thus, this report research:

- Out of the 3 promotion techniques, what is the most effective one on sales and revenue applied individually?

Then, based on this, this report further investigates:

- How can we target customers based on the store price tiers, states and cities in the US?
- Is there a correlation between geographical position and treatment effect?
- How can we target based on the interaction between those?
- Which targeting strategy (if any) is the best? (compared to non-targeting)

2. DUNNHUMBY DATA - EXPLAIN THE DATA

2.1 DATASETS

DUNNHUMBY DATA AND OUR FOCUS ON COLD CEREAL

The Dunnhumby dataset, "The Breakfast At the Frat," consists of sales and promotional information from a sample of stores, starting from January 2009 to January 2012. It presents five products from each of the three major brands in four selected categories: *frozen pizza*, *pretzels*, *cold cereal* and *mouthwash*. (3) Data was collected from stores in 4 states: *Kentucky, Ohio, Texas* and *Indiana*.
To narrow down the scope of the report, it has been decided to focus only on *cold cereal* as it has the highest number of observations (169,686).

UPC	DESCRIPTION	MANUFACTURER	CATEGORY	SUB_CATEGORY	PRODUCT_SIZE
1111085319	PL HONEY NUT TOASTD OATS	PRIVATE LABEL	COLD CEREAL	ALL FAMILY CEREAL	12.25 OZ
1111085345	PL RAISIN BRAN	PRIVATE LABEL	COLD CEREAL	ADULT CEREAL	20 OZ
1111085350	PL BT SZ FRSTD SHRD WHT	PRIVATE LABEL	COLD CEREAL	ALL FAMILY CEREAL	18 OZ
1600027527	GM HONEY NUT CHEERIOS	GENERAL MI	COLD CEREAL	ALL FAMILY CEREAL	12.25 OZ
1600027528	GM CHEERIOS	GENERAL MI	COLD CEREAL	ALL FAMILY CEREAL	18 OZ
1600027564	GM CHEERIOS	GENERAL MI	COLD CEREAL	ALL FAMILY CEREAL	12 OZ
3000006340	QKER LIFE ORIGINAL	QUAKER	COLD CEREAL	ALL FAMILY CEREAL	13 OZ
3000006560	QKER CAP N CRUNCH BERRIES	QUAKER	COLD CEREAL	KIDS CEREAL	13 OZ
3000006610	QKER CAP N CRUNCH	QUAKER	COLD CEREAL	KIDS CEREAL	14 OZ
3800031829	KELL BITE SIZE MINI WHEAT	KELLOGG	COLD CEREAL	ALL FAMILY CEREAL	18 OZ
3800031838	KELL FROSTED FLAKES	KELLOGG	COLD CEREAL	KIDS CEREAL	15 OZ
3800039118	KELL FROOT LOOPS	KELLOGG	COLD CEREAL	KIDS CEREAL	12.2 OZ
88491201426	POST HNY BN OTS HNY RSTD	POST FOODS	COLD CEREAL	ADULT CEREAL	18 OZ
88491201427	POST FM SZ HNYBNCH OT ALM	POST FOODS	COLD CEREAL	ADULT CEREAL	18 OZ
88491212971	POST FRUITY PEBBLES	POST FOODS	COLD CEREAL	KIDS CEREAL	11 OZ

Figure 1: The subcategories, manufacturers and products for cold cereal (4)

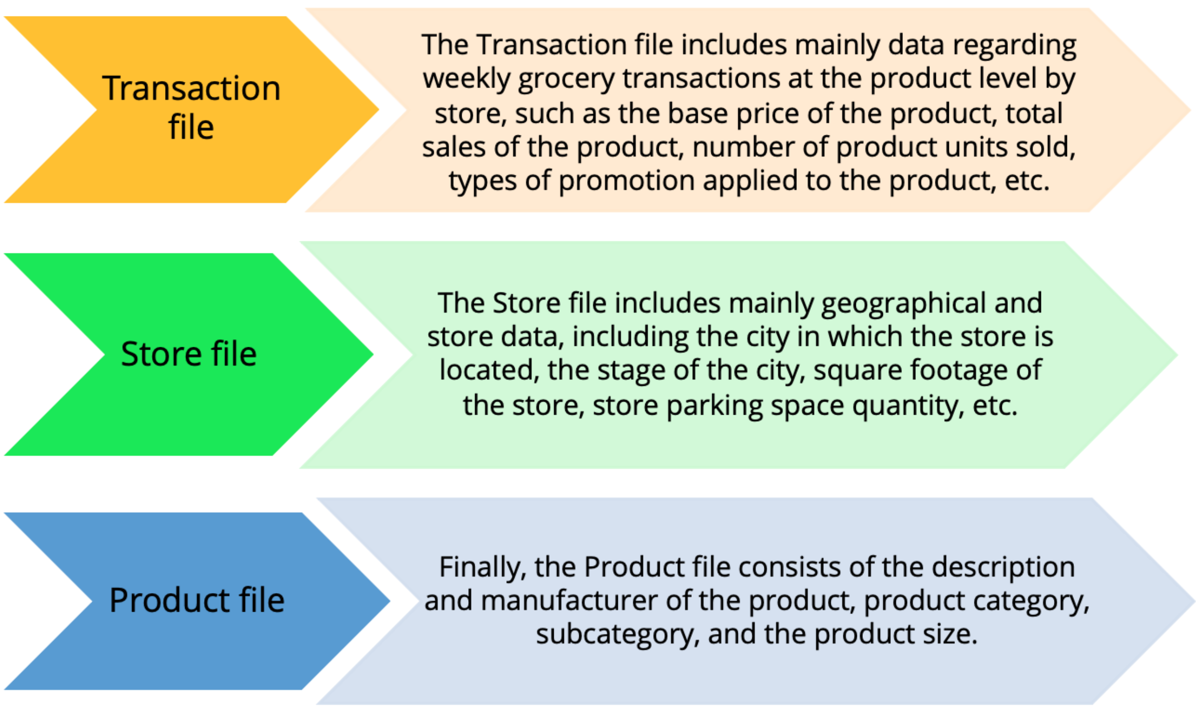
US CITIES DATASET

This dataset (4) (updated in 2021) is built using authoritative sources such as the US. Geological Survey and US Census Bureau. It contains the coordinates (longitude and latitudes) of all the cities in the US, which allows matching the coordinates of the cities in Dunnhumby data. These coordinates play an important role in Part 4: Clustering.

2.2 DATA PRE-PROCESSING

MERGING DATA

The Dunnhumby data comes in 3 individual Excel files: Transaction, Store, and Products.



These datasets were merged using UPC and STORE ID. This is followed by performing a left join (Dunhumby on the left) with the US cities dataset.

MISSING VALUES

PRICE, BASE PRICE and PARKING_SPACE_QTY variables had missing values and 3 different treatments were applied for each variable.

Firstly, the 2 missing values of PRICE were set to be equal to the BASE PRICE variable as none of the promotion types is applied (FEATURE, DISPLAY, TPR_ONLY). This is because the difference between BASE PRICE and PRICE comes from whether the promotion is applied or not.

Secondly, the 10 missing values of BASE PRICE were removed, as there is at least one promotion applied to the product and hence it could not be set equal to the PRICE.

Thirdly, the NAs from the PARKING_SPACE_QTY was set to 0, as it means there is no parking space available in the store.

RENAMING CITIES NAMES AND FORMAT

Besides, 3 cities in the US Cities Database were renamed before being merged with the Dunnhumby database, where the wording of the city names was different (Saint Mary's VS St. Mary's). Moreover, city names in the US Cities Database had to be converted to upper-case to match those in the Dunnhumby database.

3. METHODS

3.1 PRICE

For a better understanding of whether there is price discrimination or not, a single UPC (QUAKER LIFE ORIGINAL) was chosen for the analysis. The subset function was used to extract the rows containing said UPC (3000006340) and a new data frame (UPC_Single) was created for visuals and price elasticity.

UPC	DESCRIPTION	MANUFACTURER	CATEGORY	SUB_CATEGORY	PRODUCT_SIZ
3000006340	QKER LIFE ORIGINAL	QUAKER	COLD CEREAL	ALL FAMILY CEREAL	13 OZ

MEASURING PRICE ELASTICITY

The price elasticity is calculated by applying log() to both UNITS and PRICE variables, as it allows the calculation to be measured in percentage of changes in UNITS and PRICE. The formula of price elasticity used in R is shown below:

```
PED = lm(log(UNITS) ~ log(PRICE), data = .)
```

3.2 PROMOTION

ISOLATING ATTRIBUTES

To truly understand the impact of each type of promotion, it is important to isolate each promotion independently from the two others. If for example, the focus is on the FEATURE promotion, both DISPLAY and TPR_ONLY are set to 0.

The formula of treatment effect (TE), shown below, has been used to calculate the impact of each type of promotion on revenues and sales. After summarising the average TE for the 3 promotions techniques applied alone, **FEATURE promotion** was selected as it had the highest overall TE.

Treatment Effect of promotion = (mean(sales) [promotion type=1] - mean(sales) [promotion type=0]

IDENTIFYING POTENTIAL STRATEGY

The TE was found by grouping 3 variables - *store price tier*, *state* and *city*, and the one with the highest TE is selected to be a potential promotional strategy.

CLUSTERING

To find a targeting strategy that considers the interaction effect between several variables, the **k-means clustering** is used followed by plotting the TE for each cluster to target the one with the highest TE.

For the **2D clustering**, as the latitude and longitude of the cities were used, it is unnecessary to scale the data. To obtain areas that are smaller than the states (4) but larger than the cities (51) in the dataset, **k=10** is selected.

For **3D clustering**, as STORE PRICE TIER was used (coded with numbers from 1 to 3) and city coordinates (with a coordinate range of -180,180), it is necessary to scale the data. As there are 3 types of stores and 2 geographical metrics, **k=6** is selected (one cluster for each type of store in each geographical area). Based on figure 7 under section 4.2.2, 2 geographical areas were identified as follows: the one below (dark blue) represents Texas, the one above (light blue) represents Kentucky, Ohio and Indiana.

After identifying 5 targeting strategies, the TE for each of them is plotted followed by recommendations for the best one.

4. FINDINGS AND IMPLICATIONS

4.1 PRICE DISCRIMINATION

4.1.1 BY STATES FOR A SINGLE UPC

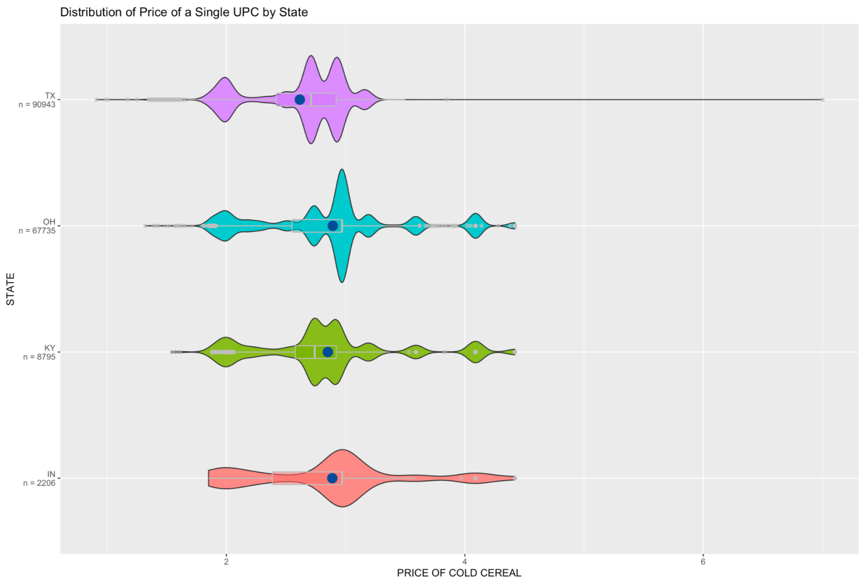


Figure 2: Distribution of Price by State (R) for one specific UPC

The violin plot with boxplot and mean showcases the distribution of prices of a single UPC (QUAKER LIFE ORIGINAL) among the four states. While the mean prices are almost similar in the three Northern states, there is slight price discrimination between them. Furthermore, it shows that Texas has the lowest average price, and it is also found that the price elasticity in Texas is inelastic (most inelastic among all 4 states). It suggests that there should be higher price discrimination in Texas for retailers to absorb more revenue. Figure 3a shows the price elasticity of each state whereas 3b shows the price elasticity by a single UPC across all states. The purpose of generating their elasticity is merely to support the findings from price discrimination by observing their relationship.

	ADDRESS_STATE_PROV_CODE	term	estimate	std.error	statistic	p.value
1	IN	log(PRICE)	-1.1051284	0.05945747	-18.58687	9.420927e-72
2	KY	log(PRICE)	-1.0422801	0.03399576	-30.65912	3.254571e-196
3	OH	log(PRICE)	-0.9048202	0.01255836	-72.04925	0.000000e+00
4	TX	log(PRICE)	-0.2251337	0.01080617	-20.83381	3.590867e-96

Figure 3a: PED for each states

	ADDRESS_STATE_PROV_CODE	term	estimate	std.error	statistic	p.value
1	IN	log(PRICE)	-2.893811	0.30385195	-9.523754	1.032120e-16
2	KY	log(PRICE)	-3.153059	0.16906735	-18.649724	3.774359e-60
3	OH	log(PRICE)	-3.287542	0.06018498	-54.623958	0.000000e+00
4	TX	log(PRICE)	-3.226829	0.06633476	-48.644624	0.000000e+00

Figure 3b: PED of each states by a single UPC

4.1.2 BY CITIES FOR A SINGLE UPC

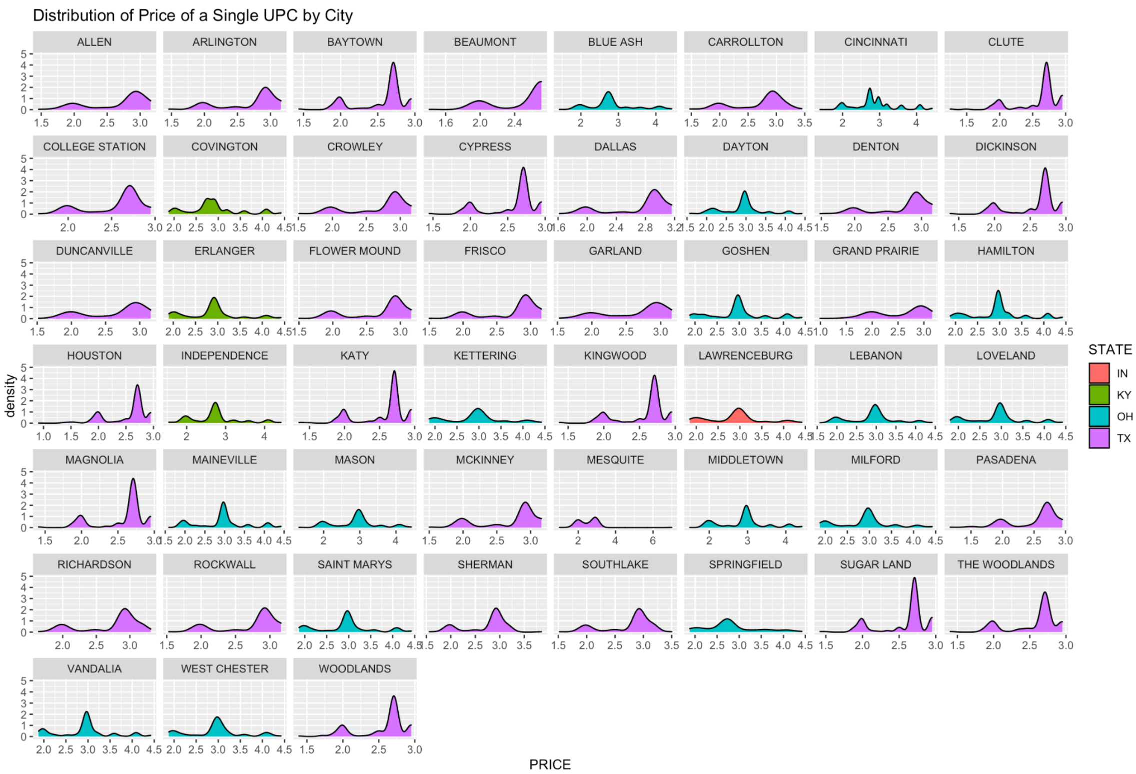


Figure 4: Distribution of Price by City (R) for one specific UPC

Figure 4 describes the distribution of prices across cities in the four states for one specific UPC (QUAKER LIFE ORIGINAL). It shows that small, similar groups are formed within Texas, hinting at price discrimination between said groups. For example, some Texas charts show significantly different and earlier peaks (such as Baytown, Katy or Magnolia) than other Texas cities (such as Garland and Grand Prairie). However in the cities of the other 3 states, prices are charged similarly.

4.1.3 IMPLICATIONS

The findings are supported by Stefano DellaVigna and Matthew Gentzkow's paper "Uniform Pricing in U.S. retail chains," which shows most U.S. retailers charge nearly-uniform prices across stores despite wide variations in consumer demographics, competition, and geography. (5)

The paper proposes **2** major reasons for the findings:

Managerial inertia, which encompasses both agency frictions and behavioural characteristics that prevent businesses from implementing optimal policies, even when the benefits outweigh the traditional economic costs. (5)

The second is **brand image concerns**, which include a variety of ways through which differing prices among locations could cause negative consumer reactions, lowering the long-term demand for a chain. Factors like tacit collusion, menu costs, and engineering costs could also play a role in such price discriminatory decisions. (5)

4.2 PROMOTION

4.2.1 AVERAGE TREATMENT EFFECT ON SALES AND REVENUE

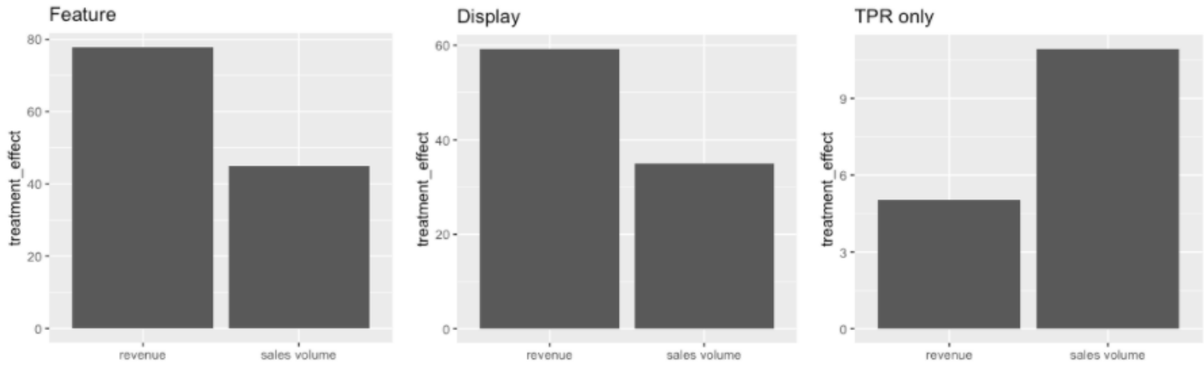


Figure 5 : Average TE on sales and revenue per each promotion technique (R)

Figure 5 shows that FEATURE has the highest TE, hence, for simplicity, this report will focus on finding a targeting marketing strategy for FEATURE only.

4.2.2 TREATMENT EFFECT OF FEATURE BY STATE, CITY AND STORE PRICE TIERS

TE BY STATE

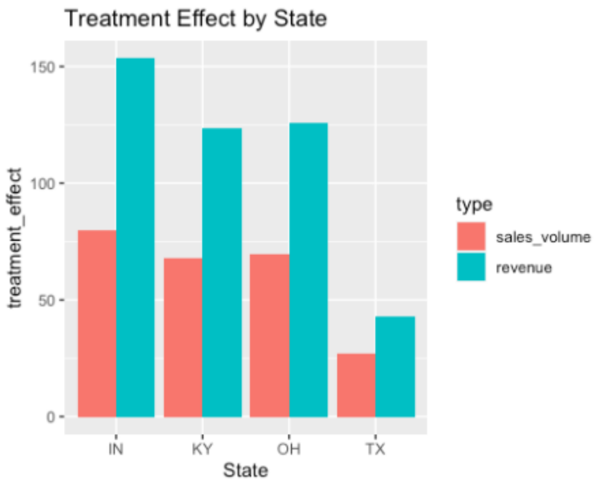


Figure 6 : TE by state (R)

The TE of each state is analysed, and the result indicates that Indiana has the highest TE of feature whereas Texas has the lowest. Hence, when analysing the TE of feature by cities, those in Texas will be excluded. The ***first potential strategy is to focus promotion on Indiana***.

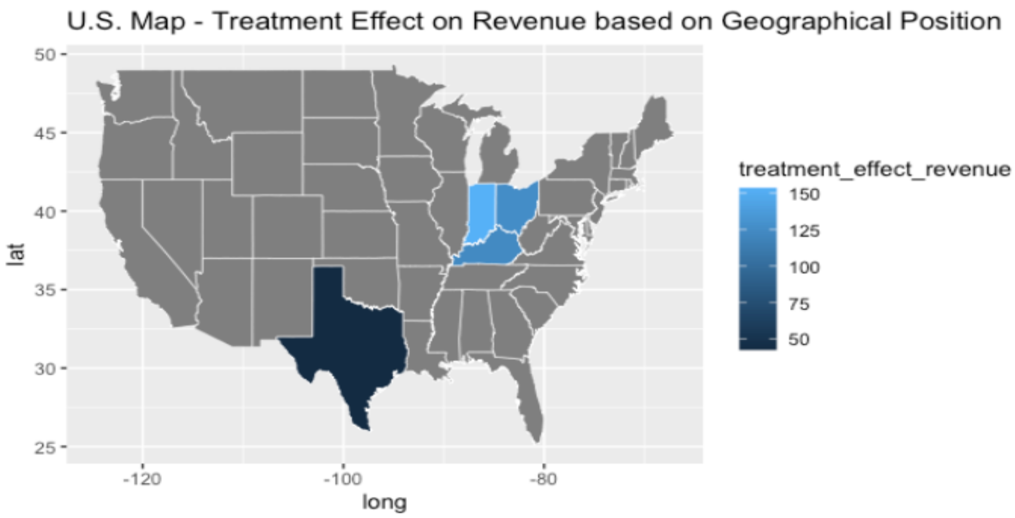


Figure 7: U.S Map that shows TE on revenue based on geographical position (R)

The map suggests a correlation between the TE and geographical position and TE tends to increase when moving towards the North of the US. However, this can be misleading because only 4 out of 50 states were analysed.

TE OF FEATURE BY CITY (TEXAS EXCLUDED)

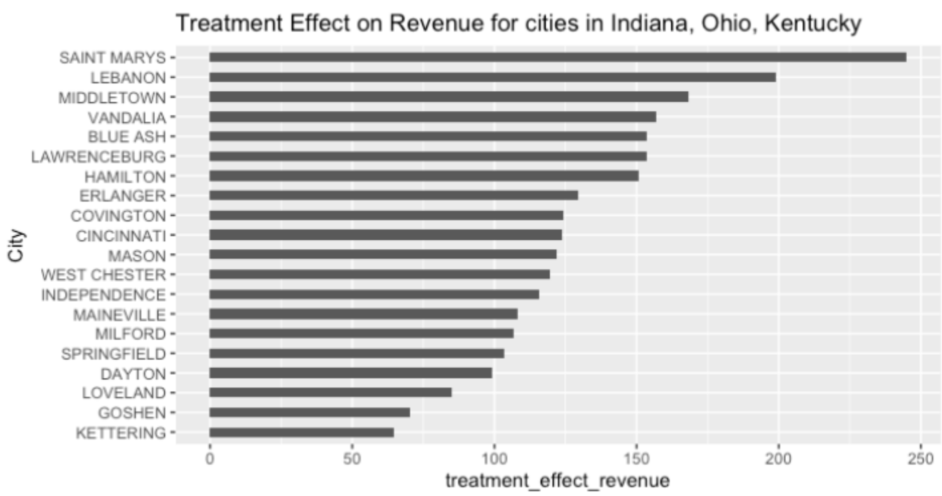


Figure 8: TE per city (R)

Figure 8 shows that Saint Marys and Lebanon are the top 2 highest TE of feature on revenue, suggesting an insight to target these cities for a more effective promotion strategy. Hence, the ***second potential strategy is to focus on Saint Marys and Lebanon.*** Interestingly, these three cities are in Ohio, and the state with the highest TE (Indiana) has a city (Lawrenceburg) ranked in the top 6. Although Indiana has the highest TE, it only has one city in the entire dataset.

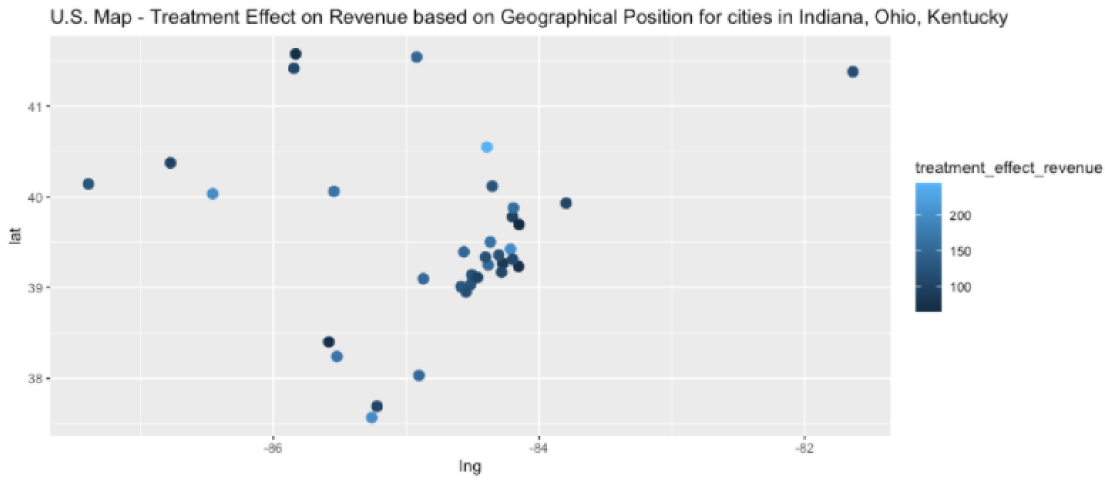


Figure 9: TE on revenue based on longitude and latitude (R)

The diagram above shows no obvious correlation between the geographical position and TE.

TE OF FEATURE BY STORE PRICE TIER

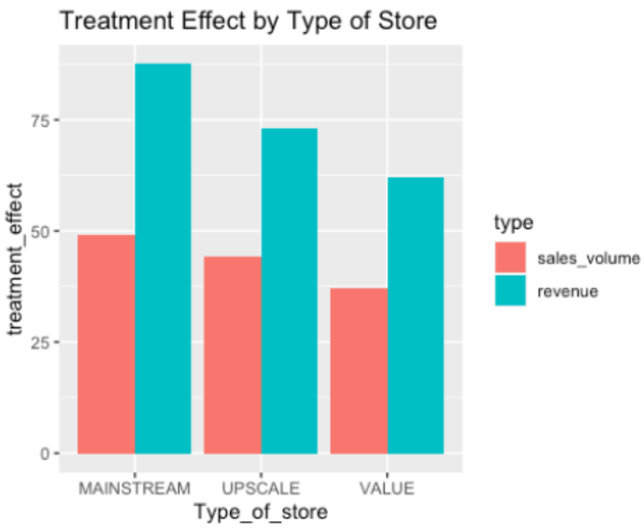


Figure 10 : TE by type of store (R)

There are 3 main ways to categorize the type of stores in Dunnhumby' dataset: Value, Mainstream and Upscale. Firstly, the Value stores are designed for consumers who are highly sensitive to price change. Secondly, the Upscale stores are designed to serve consumers who emphasize more on the shopping experience and variety of products. Finally, the Mainstream stores are just the middle of both extremes. Figure 6 illustrates that Mainstream stores have the highest TE compared to Upscale and Value stores, hence the ***third potential promotion strategy would be to focus on Mainstream stores***.

4.2.3 CLUSTERING

CLUSTERING ON LATITUDE AND LONGITUDE

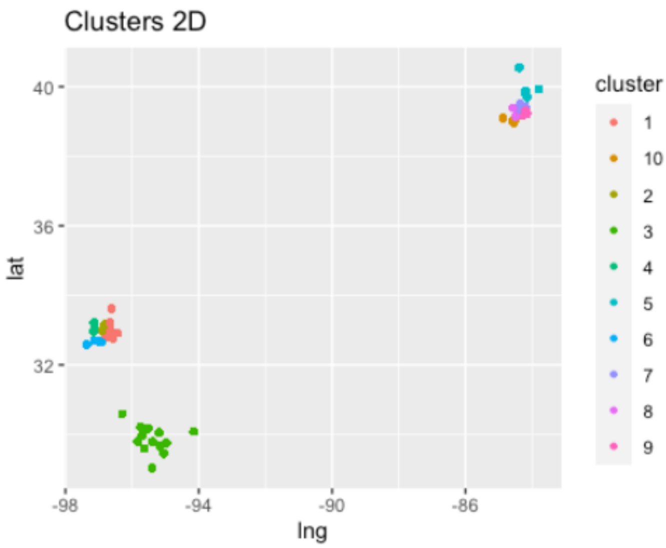


Figure 11: Clustering on Lat/long of cities (R)

The diagram above shows a 2D K-means clustering by longitude and latitude, specifically for 10 clusters across 51 cities. The outcome represents a US map, which shows that there are 5 clusters in Texas and another 5 in the remaining states. Figure 12 shows the TE of feature on revenue for each identified cluster, where cluster 7 has the highest TE and cities belonging to cluster 7 are found to be: Mason, Middletown, Lebanon and West Chester. This suggests the ***fourth potential strategy would be to target cluster 7***.

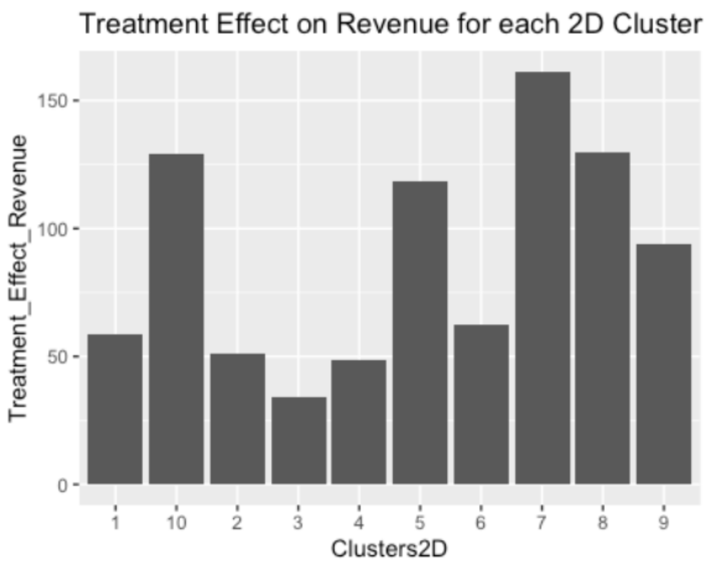


Figure 12: TE on revenue for each 2D cluster (R)

CLUSTERING WITH LATITUDE, LONGITUDE, (OF CITIES) SEG_VALUE (STORE PRICE TIER)

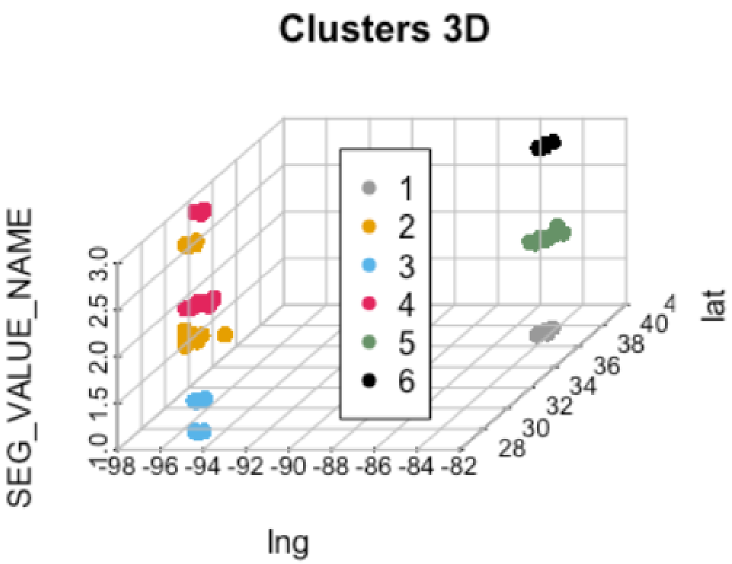


Figure 13: 3D Clustering (R)

The store price tiers are changed from characters to numerical, by assigning 1 for Value, 2 for Mainstream and 3 for Upscale stores. Based on figure 13 and results from R, clusters 2, 3 and 4 are located in Texas whereas clusters 1, 5 and 6 are located in the other three states. The diagram below shows cluster 6 has the highest TE of feature. Cluster 6 consists of the Upscale stores in the following cities: Kettering, Cincinnati, Blue Ash, Mason, Loveland and Middletown. Hence, the ***fifth potential strategy would be to target cluster 6.***

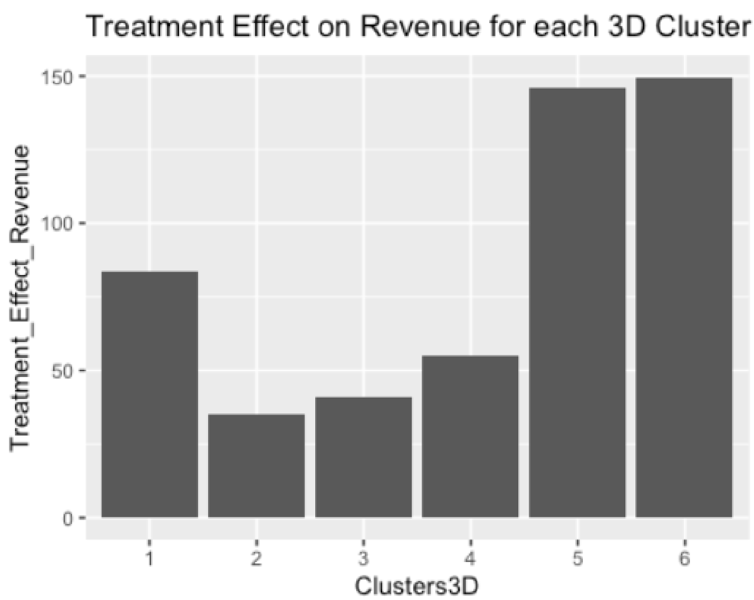


Figure 14: TE on revenue for each 3D Cluster (R)

4.2.4 IMPLICATIONS

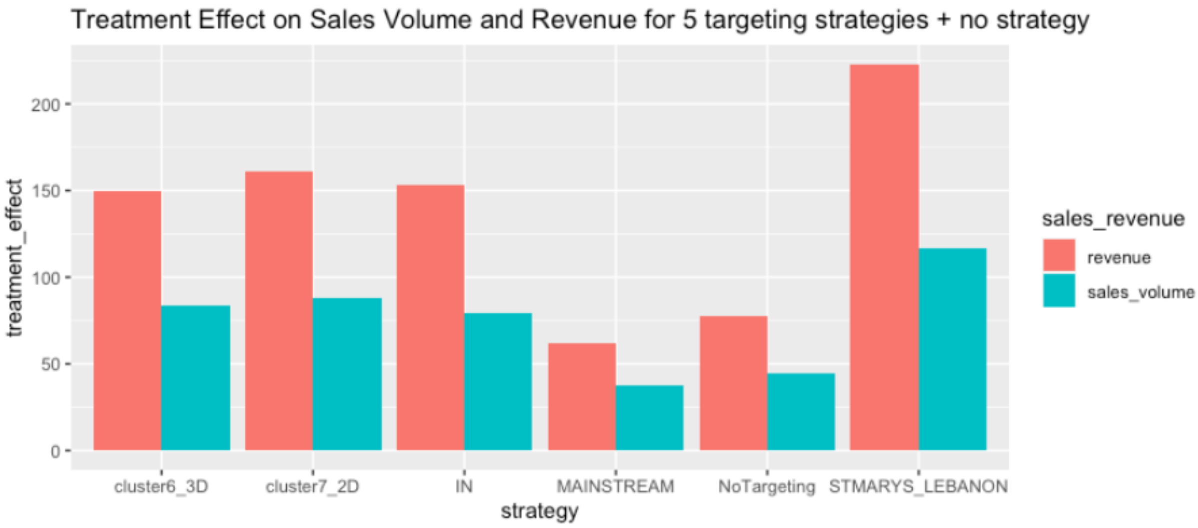


Figure 15: TE on sales and revenue for different targeting strategies (R)

Figure 15 shows a comparison between all five potential targeting strategies on promotion (by FEATURE only). As a result, targeting **Saint Marys** and **Lebanon** shows the highest TE of feature, indicating that grocery retailers should firmly target these cities to achieve the greatest result from the FEATURE promotion strategy. Interestingly, it is found that Saint Marys has the highest elasticity across all 51 cities, with a value of -1.36, while Lebanon ranked top 7 (see appendix). Their high elasticity partly explains the strong TE of in-store circular promotion by offering lower prices.

5. DIRECTIONS FOR FUTURE RESEARCH

5.1 ANALYSING ALL PRODUCT CATEGORIES

Only the cold cereal product category was analysed. Consequently, further research could be done by analysing all product categories to generate additional insights and suggest more informed marketing strategies.

5.2 FURTHER INVESTIGATION ON THE OTHER PROMOTION TECHNIQUES

There are three types of promotions but the feature promotion is the only one that is used for targeting as it has the highest individual TE, compared to DISPLAY and TPR. However, it might be possible that DISPLAY or TPR would give a higher TE when targeting is applied. This suggests that future research could be done to offer targeting for all 3 types of promotions, maximising the TE (i.e. target St. Marys for feature, target value stores with TPR, target Texas with display).

5.3 INTERACTION BETWEEN ALL THREE TYPES OF PROMOTIONS

Future research might include finding the TE for interactions between all 3 types of promotion. However, in the dataset, TPR is only applied alone, so it is only possible to find the interaction effect between FEATURE (F) and DISPLAY (D). This is done by setting TPR = 0 and:

TE(D+F) = mean(sales(D =1 & F =1)) - mean(sales(D=0 & F=0))

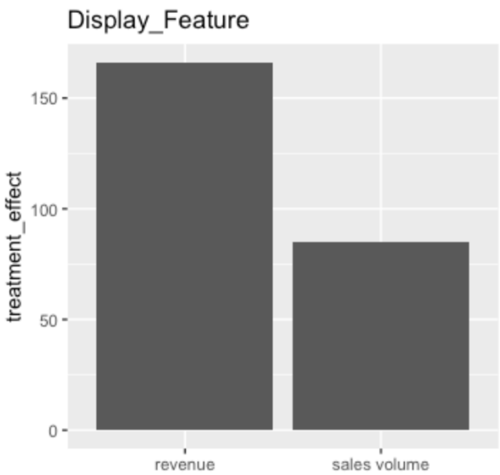


Figure 16: Treatment Effect between Feature and Display

Interestingly, the TE is larger compared to the promotions applied alone or the TE as shown earlier. Future research should allow a comprehensive study on a targeted marketing strategy for DISPLAY and FEATURE.

5.4 Targeting Cities in Indiana or with more states in the dataset

The report could focus on targeting cities in Indiana - state with highest TE - if there were more information as the dataset gives only information on one city of this state. Besides, with the availability of more states in the US, clearer insights can be generated alongside with longitude and latitude data that this report has utilised.

5.5 Price Discrimination

Finally, future research could be done on price discrimination by collecting information on customers' visits, their willingness to spend by cities and the time spent at the store.

6. APPENDIX

DUNNHUMBY - DESCRIPTION

Dunnhumby is a global customer data science company that offers solutions to clients across several sectors including the grocery retail industry. It provides marketing services such as pricing and promotion consulting, through utilising large amounts of data, with more than 9 billion data records processed every week. In 30 years, It analysed an accumulated $600 billion worldwide retail revenue.

HOW CAN PROMOTION ANALYSIS HELP THE GROCERY RETAILERS IN MAKING DECISIONS ?

Customer awareness of a product or brand is increased through promotion, which leads to increased sales and brand loyalty. As a result, promotion analysis, particularly promotion effectiveness, has become a popular analytical topic. In an uncertain economy, the majority of customers are programmed to only buy things on sale. Because promotions can be costly to a firm, their performance must be effectively understood and implemented in order to meet business objectives. With the influx of data, retailers have understood that promotions should not be viewed as a one-time attempt to boost traffic and lower inventory levels; rather, promotions should be a part of their overall planning process. Promotional analysis can assist grocery stores in determining the most successful product and promotion mix in order to improve business performance and answer critical questions like as

1. What is the impact on sales of promotions, displays, or being featured in the circular?

IMPROVEMENTS TO THE BY CITIES GRAPH BASED ON ADAM'S FEEDBACK (UNDER 4.1.2)

Figure 17: Distribution of Price by City (R) for all cold cereal categories (UPC)

The chart studies the distribution of prices across cities in the four states for all categories of cold cereal (UPC). In this case, the distribution between cities may be explained by the fact that there are different categories of cold cereals or by the fact that customers in different cities have different cereal preferences, which explains the category load. Thus, it would be more accurate to focus on one specific product (UPC) to study the variation of this specific product category between cities. This is why it has been decided to focus on only one UPC in the main report.

PRICE ELASTICITY OF EACH CITY

	ADDRESS_CITY_NAME	term	estimate	std.error	statistic	p.value
1	ALLEN	log(PRICE)	-0.42199997	0.05222489	-8.0804373	1.047329e-15
2	ARLINGTON	log(PRICE)	-1.00023193	0.05834808	-17.1424990	6.119697e-62
3	BAYTOWN	log(PRICE)	-0.18062194	0.06564894	-2.7513305	5.983731e-03
4	BEAUMONT	log(PRICE)	-0.35300216	0.07485516	-4.7158029	2.556844e-06
5	BLUE ASH	log(PRICE)	-0.96602317	0.05385790	-17.9365171	2.959500e-67
6	CARROLLTON	log(PRICE)	-0.18210610	0.06220093	-2.9277072	3.449413e-03
7	CINCINNATI	log(PRICE)	-0.74311380	0.02810647	-26.4392455	2.230921e-151
8	CLUTE	log(PRICE)	-0.12712860	0.06171753	-2.0598460	3.952934e-02
9	COLLEGE STATION	log(PRICE)	-0.37484514	0.06859833	-5.4643478	5.167162e-08
10	COVINGTON	log(PRICE)	-0.92062477	0.05145353	-17.8923532	3.631680e-69
11	CROWLEY	log(PRICE)	-0.75476048	0.05635137	-13.3938274	2.108152e-39
12	CYPRESS	log(PRICE)	0.07973666	0.06396751	1.2465180	2.127057e-01
13	DALLAS	log(PRICE)	-0.14290690	0.05710406	-2.5025697	1.240084e-02
14	DAYTON	log(PRICE)	-0.53583621	0.04326545	-12.3848532	1.183430e-34
15	DENTON	log(PRICE)	-0.74215624	0.06047812	-12.2714829	1.518940e-33
16	DICKINSON	log(PRICE)	-0.34312550	0.06463158	-5.3089450	1.212776e-07
17	DUNCANVILLE	log(PRICE)	-0.60715614	0.06562428	-9.2520049	5.024588e-20
18	ERLANGER	log(PRICE)	-1.05090809	0.05602222	-18.7587742	6.231101e-73
19	FLOWER MOUND	log(PRICE)	-0.18183219	0.05595467	-3.2496336	1.172793e-03
20	FRISCO	log(PRICE)	0.45107914	0.06018566	7.4947945	9.532096e-14
21	GARLAND	log(PRICE)	-0.49454313	0.06872160	-7.1963271	8.429569e-13
22	GOSHEN	log(PRICE)	-1.10855209	0.05882970	-18.8434085	1.502420e-73
23	GRAND PRAIRIE	log(PRICE)	-0.56670555	0.06302939	-8.9911325	5.134948e-19
24	HAMILTON	log(PRICE)	-0.70074619	0.04222775	-16.5944471	4.862209e-60
25	HOUSTON	log(PRICE)	0.13384598	0.02716900	4.9264219	8.450598e-07
26	INDEPENDENCE	log(PRICE)	-1.23023725	0.05727022	-21.4812737	4.463151e-93

	ADDRESS_CITY_NAME	term	estimate	std.error	statistic	p.value
27	KATY	log(PRICE)	-0.04874749	0.04941777	-0.9864366	3.239727e-01
28	KETTERING	log(PRICE)	-0.96342780	0.04818556	-19.9941205	6.554090e-85
29	KINGWOOD	log(PRICE)	-0.29599324	0.06100969	-4.8515776	1.310092e-06
30	LAWRENCEBURG	log(PRICE)	-1.10512842	0.05945747	-18.5868719	9.420927e-72
31	LEBANON	log(PRICE)	-1.12125719	0.05785511	-19.3804346	2.141170e-77
32	LOVELAND	log(PRICE)	-0.92306133	0.04244870	-21.7453403	1.123275e-99
33	MAGNOLIA	log(PRICE)	-0.23146710	0.06141837	-3.7686953	1.683681e-04
34	MAINEVILLE	log(PRICE)	-1.24164371	0.04316697	-28.7637454	7.276582e-167
35	MASON	log(PRICE)	-0.74337947	0.05769338	-12.8850048	1.107954e-36
36	MCKINNEY	log(PRICE)	-0.71688372	0.04465793	-16.0527763	2.055217e-56
37	MESQUITE	log(PRICE)	-0.80020587	0.06003816	-13.3282873	4.890711e-39
38	MIDDLETOWN	log(PRICE)	-0.93421275	0.03412553	-27.3757726	2.045414e-156
39	MILFORD	log(PRICE)	-1.19469111	0.05642132	-21.1744609	9.976101e-91

	ADDRESS_CITY_NAME	term	estimate	std.error	statistic	p.value
40	PASADENA	log(PRICE)	-0.04993162	0.07715472	-0.6471623	5.175947e-01
41	RICHARDSON	log(PRICE)	-0.82484502	0.05860291	-14.0751557	3.644689e-43
42	ROCKWALL	log(PRICE)	-0.34583576	0.05816269	-5.9460067	3.181779e-09
43	SAINT MARYS	log(PRICE)	-1.35651824	0.06652782	-20.3902391	8.269069e-85
44	SHERMAN	log(PRICE)	-0.68549209	0.06177359	-11.0968469	6.970284e-28
45	SOUTHLAKE	log(PRICE)	-0.30591676	0.05562574	-5.4995536	4.245853e-08
46	SPRINGFIELD	log(PRICE)	-1.07788319	0.06147557	-17.5335203	1.531339e-64
47	SUGAR LAND	log(PRICE)	0.22526072	0.04996898	4.5080114	6.712479e-06
48	THE WOODLANDS	log(PRICE)	0.44517238	0.06331666	7.0308881	2.721030e-12
49	VANDALIA	log(PRICE)	-1.21142758	0.06220417	-19.4750230	4.232342e-78
50	WEST CHESTER	log(PRICE)	-0.97057382	0.06070255	-15.9890123	1.529783e-54
51	WOODLANDS	log(PRICE)	0.14122728	0.06766382	2.0871906	3.698455e-02

PRICE ELASTICITY OF EACH CITY BY A SINGLE UPC

	ADDRESS_CITY_NAME	term	estimate	std.error	statistic	p.value
1	ALLEN	(Intercept)	5.464192	0.3095642	17.651243	2.580395e-37
2	ALLEN	log(PRICE)	-2.976655	0.3114802	-9.556482	6.061804e-17
3	ARLINGTON	(Intercept)	4.616241	0.3346680	13.793491	2.340246e-27
4	ARLINGTON	log(PRICE)	-3.337275	0.3352730	-9.953904	9.333386e-18
5	BAYTOWN	(Intercept)	5.035010	0.3575703	14.081178	7.468953e-28
6	BAYTOWN	log(PRICE)	-3.776579	0.3798837	-9.941409	1.229901e-17
7	BEAUMONT	(Intercept)	6.058654	0.3913870	15.479958	1.024847e-31
8	BEAUMONT	log(PRICE)	-4.619805	0.4254692	-10.858142	3.950937e-20
9	BLUE ASH	(Intercept)	6.931632	0.1998630	34.681923	1.451775e-67
10	BLUE ASH	log(PRICE)	-3.586627	0.1918764	-18.692384	1.180983e-38
11	CARROLLTON	(Intercept)	4.792788	0.3760529	12.744982	3.509476e-25
12	CARROLLTON	log(PRICE)	-3.007514	0.3786302	-7.943144	5.829713e-13
13	CINCINNATI	(Intercept)	5.604287	0.1542857	36.324079	8.461394e-190
14	CINCINNATI	log(PRICE)	-2.962770	0.1472122	-20.125843	8.041971e-77
15	CLUTE	(Intercept)	4.937237	0.3431865	14.386456	4.969289e-29
16	CLUTE	log(PRICE)	-3.273571	0.3605495	-9.079395	1.172910e-15
17	COLLEGE STATION	(Intercept)	4.768842	0.4585132	10.400665	4.276355e-19
18	COLLEGE STATION	log(PRICE)	-3.610932	0.4876921	-7.404122	1.152273e-11
19	COVINGTON	(Intercept)	4.865383	0.2769937	17.564957	1.745824e-46
20	COVINGTON	log(PRICE)	-2.615615	0.2608916	-10.025675	2.695221e-20
21	CROWLEY	(Intercept)	5.812619	0.3246201	17.905912	6.445017e-38
22	CROWLEY	log(PRICE)	-3.720865	0.3259844	-11.414243	1.056510e-21
23	CYPRESS	(Intercept)	5.456926	0.3816592	14.297902	7.011301e-29
24	CYPRESS	log(PRICE)	-3.175454	0.4039652	-7.860711	1.044658e-12
25	DALLAS	(Intercept)	4.518855	0.2810077	16.080897	1.941406e-33
26	DALLAS	log(PRICE)	-1.708694	0.2787974	-6.128801	8.734346e-09

	ADDRESS_CITY_NAME	term	estimate	std.error	statistic	p.value
27	DAYTON	(Intercept)	4.634317	0.2478754	18.696158	8.696387e-51
28	DAYTON	log(PRICE)	-1.990108	0.2333539	-8.528282	1.034068e-15
29	DENTON	(Intercept)	6.504542	0.2927001	22.222545	9.370192e-48
30	DENTON	log(PRICE)	-4.297120	0.2912065	-14.756265	2.585701e-30
31	DICKINSON	(Intercept)	5.044091	0.3408951	14.796609	8.201478e-30
32	DICKINSON	log(PRICE)	-3.764102	0.3620645	-10.396219	7.319521e-19
33	DUNCANVILLE	(Intercept)	5.989193	0.2992881	20.011460	6.086208e-42
34	DUNCANVILLE	log(PRICE)	-4.384034	0.3040898	-14.416907	5.827480e-29
35	ERLANGER	(Intercept)	6.483177	0.2643575	24.524280	8.380009e-51
36	ERLANGER	log(PRICE)	-3.487431	0.2513629	-13.874087	1.731274e-27
37	FLOWER MOUND	(Intercept)	6.121998	0.3409199	17.957291	6.286195e-38
38	FLOWER MOUND	log(PRICE)	-3.955053	0.3408818	-11.602418	3.801184e-22
39	FRISCO	(Intercept)	5.478842	0.2762005	19.836467	4.186064e-42
40	FRISCO	log(PRICE)	-3.388692	0.2730431	-12.410831	3.580714e-24
41	GARLAND	(Intercept)	5.666858	0.3379160	16.770017	3.302234e-34
42	GARLAND	log(PRICE)	-4.255455	0.3422743	-12.432879	8.140500e-24
43	GOSHEN	(Intercept)	4.985826	0.3003969	16.597459	8.244160e-34
44	GOSHEN	log(PRICE)	-2.778787	0.2767242	-10.041719	6.952133e-18
45	GRAND PRAIRIE	(Intercept)	4.601431	0.2452854	18.759503	1.278010e-37
46	GRAND PRAIRIE	log(PRICE)	-3.379818	0.2590991	-13.044502	7.988112e-25
47	HAMILTON	(Intercept)	5.807035	0.2284888	25.414963	8.313716e-74
48	HAMILTON	log(PRICE)	-3.215377	0.2103616	-15.285001	1.698408e-38
49	HOUSTON	(Intercept)	4.694675	0.1559974	30.094577	5.199735e-144
50	HOUSTON	log(PRICE)	-3.262772	0.1672254	-19.511228	1.353444e-72
51	INDEPENDENCE	(Intercept)	6.078415	0.2621902	23.183233	3.371170e-48
52	INDEPENDENCE	log(PRICE)	-3.796609	0.2596207	-14.623679	2.573855e-29

	ADDRESS_CITY_NAME	term	estimate	std.error	statistic	p.value
53	KATY	(Intercept)	4.987989	0.3037412	16.421837	2.284170e-42
54	KATY	log(PRICE)	-3.492097	0.3231263	-10.807221	8.032257e-23
55	KETTERING	(Intercept)	5.736254	0.2182648	26.281172	2.584323e-72
56	KETTERING	log(PRICE)	-2.989904	0.2030089	-14.727944	2.462139e-35
57	KINGWOOD	(Intercept)	6.123427	0.3046573	20.099392	3.922513e-42
58	KINGWOOD	log(PRICE)	-3.750415	0.3233217	-11.599639	6.293042e-22
59	LAWRENCEBURG	(Intercept)	5.656318	0.3220239	17.564901	1.828991e-36
60	LAWRENCEBURG	log(PRICE)	-2.893811	0.3038520	-9.523754	1.032120e-16
61	LEBANON	(Intercept)	6.618523	0.3232286	20.476291	6.034333e-43
62	LEBANON	log(PRICE)	-3.774663	0.3049899	-12.376357	6.956282e-24
63	LOVELAND	(Intercept)	5.800500	0.2435706	23.814452	8.458279e-68
64	LOVELAND	log(PRICE)	-3.368705	0.2291947	-14.698007	3.579819e-36
65	MAGNOLIA	(Intercept)	5.587543	0.3137588	17.808401	5.040584e-37
66	MAGNOLIA	log(PRICE)	-3.651305	0.3321939	-10.991485	2.152370e-20
67	MAINEVILLE	(Intercept)	6.487620	0.1649405	39.333086	1.074226e-111
68	MAINEVILLE	log(PRICE)	-3.870433	0.1538266	-25.161018	9.652798e-72
69	MASON	(Intercept)	6.449260	0.2297963	28.065122	1.708792e-57
70	MASON	log(PRICE)	-3.599427	0.2171356	-16.576861	4.610521e-34
71	MCKINNEY	(Intercept)	5.796402	0.2633676	22.008791	2.356861e-62
72	MCKINNEY	log(PRICE)	-4.008011	0.2645904	-15.147985	5.262899e-38
73	MESQUITE	(Intercept)	4.283925	0.2836771	15.101410	1.499227e-30
74	MESQUITE	log(PRICE)	-2.921366	0.3068391	-9.520843	1.113276e-16
75	MIDDLETOWN	(Intercept)	6.091556	0.1714102	35.537889	7.882135e-126
76	MIDDLETOWN	log(PRICE)	-3.369584	0.1616692	-20.842459	7.545432e-66
77	MILFORD	(Intercept)	6.066191	0.2557846	23.716010	3.033876e-49
78	MILFORD	log(PRICE)	-3.464418	0.2413517	-14.354230	1.161693e-28

	ADDRESS_CITY_NAME	term	estimate	std.error	statistic	p.value
79	PASADENA	(Intercept)	4.328731	0.3511324	12.327916	7.907267e-23
80	PASADENA	log(PRICE)	-3.548228	0.3772502	-9.405505	5.872796e-16
81	RICHARDSON	(Intercept)	5.777770	0.2853465	20.248262	2.702045e-43
82	RICHARDSON	log(PRICE)	-3.760401	0.2833067	-13.273252	1.751191e-26
83	ROCKWALL	(Intercept)	6.078653	0.3578945	16.984482	1.597169e-35
84	ROCKWALL	log(PRICE)	-3.616769	0.3558517	-10.163697	1.975587e-18
85	SAINT MARYS	(Intercept)	6.867673	0.3413332	20.120143	2.558006e-42
86	SAINT MARYS	log(PRICE)	-4.014033	0.3179915	-12.623083	1.476994e-24
87	SHERMAN	(Intercept)	6.538136	0.3205510	20.396554	2.454428e-42
88	SHERMAN	log(PRICE)	-4.830469	0.3186785	-15.157813	1.603971e-30
89	SOUTHLAKE	(Intercept)	5.427033	0.2770191	19.590824	1.483669e-41
90	SOUTHLAKE	log(PRICE)	-3.190843	0.2741114	-11.640676	3.338587e-22
91	SPRINGFIELD	(Intercept)	6.115654	0.2426193	25.206789	6.888229e-52

	ADDRESS_CITY_NAME	term	estimate	std.error	statistic	p.value
92	SPRINGFIELD	log(PRICE)	-3.316432	0.2339992	-14.172838	3.793545e-28
93	SUGAR LAND	(Intercept)	5.030317	0.2505007	20.081052	1.974605e-55
94	SUGAR LAND	log(PRICE)	-3.321749	0.2649458	-12.537467	1.049198e-28
95	THE WOODLANDS	(Intercept)	5.083891	0.3398817	14.957827	1.356599e-30
96	THE WOODLANDS	log(PRICE)	-3.034551	0.3600415	-8.428336	4.295696e-14
97	VANDALIA	(Intercept)	6.417825	0.3036166	21.137923	1.656643e-44
98	VANDALIA	log(PRICE)	-3.738639	0.2843121	-13.149771	7.004798e-26
99	WEST CHESTER	(Intercept)	6.999220	0.3072423	22.780787	2.896949e-48
100	WEST CHESTER	log(PRICE)	-4.284533	0.2850352	-15.031591	1.065274e-30
101	WOODLANDS	(Intercept)	5.609050	0.3340056	16.793284	4.570237e-35
102	WOODLANDS	log(PRICE)	-3.990116	0.3553762	-11.227865	3.803931e-21

BIBLIOGRAPHY

[1] Kuijpers, Dymfke, Virginia Simmons, and Jasper van Wamelen. "Reviving Grocery Retail: Six Imperatives." McKinsey & Company. McKinsey & Company, April 13, 2020. https://www.mckinsey.com/industries/retail/our-insights/reviving-grocery-retail-six-imperatives.

[2] Aull, Bill, Anuja Desikan Perkins, Sajal Kohli, and Eric Marohn. "The State of Grocery in North America." McKinsey & Company. McKinsey & Company, June 8, 2021. https://www.mckinsey.com/industries/retail/our-insights/the-state-of-grocery-in-north-america.

[3] Dunnhumby. "Source Files - Dunnhumby." Grocery Data. dunnhumby. Accessed December 13, 2021. https://www.dunnhumby.com/source-files/

[4]"United States Cities Database." simplemaps. Accessed December 16, 2021. https://simplemaps.com/data/us-cities

[5]Dellavigna, Stefano. UNIFORM PRICING IN US RETAIL CHAINS. Accessed December 13, 2021. https://www.nber.org/papers/w23996