Jini Ashok Bhanushali

Final Project

Data:

The data includes sales and promotional information from a sampling of stores over 156 weeks, beginning January 2009 through December 2011. These are the sales metrics of products in 4 categories: mouthwash, pretzels, frozen pizza, and boxed cereal, the top 5 products from each of the top 3 brands in each category.

The data consists of three excel sheets : Store Lookup, Product Lookup and Transaction Data. Additionally we are also provided with an MSA dataset that helps us to identify the locations based on the US national msa location encoding.
We have successfully connected all these sheets and the data in them with a relation for each.

Let's first look and understand our data:

Store Lookup: This sheet gives all the information related to the stores, like their location, name, area/space occupied and sales, etc.

STORE_ID	STORE_NAME	ADDRESS_CITY_NAME	ADDRESS_STATE_PROV_CODE	MSA	SEG_VALUE_NAME	PARKING_SPACE_QTY	SALES_AREA_SIZE_NUM	AVG_WEEKLY_BASKETS	Latitude	Longitude
389	SILVERLAKE	ERLANGER	KY	17140	MAINSTREAM	408	46073	24767	39.01994	-84.5936
2277	ANDERSON TOWNE CTR	CINCINNATI	OH	17140	UPSCALE		81958	54053	39.058412	-84.3396
4259	WARSAW AVENUE	CINCINNATI	OH	17140	VALUE		48813	31177	39.11161	-84.578
6379	KINGWOOD	KINGWOOD	TX	26420	MAINSTREAM		50237	20620	30.067048	-95.1871
6431	AT WARD ROAD	BAYTOWN	TX	26420	VALUE	350	43698	24322	29.740178	-94.9523
13853	RICHARDSON	RICHARDSON	TX	19100	MAINSTREAM		53586	15787	32.938364	-96.7222
13859	MCKINNEY	MCKINNEY	TX	19100	MAINSTREAM		47824	15345	33.213761	-96.6305
15755	KROGER JUNCTION S/C	PASADENA	TX	26420	VALUE	351	34373	18291	29.664894	-95.1898
23327	HOUSTON	HOUSTON	TX	26420	MAINSTREAM		50722	30258	29.740102	-95.3401
23345	FRISCO	FRISCO	TX	19100	UPSCALE		46369	21947	33.14972	-96.8188
25001	PINEWOOD	DAYTON	OH	19380	VALUE	351	34177	18035	39.764658	-84.1507
26983	DAYTON	DAYTON	OH	19380	VALUE		48782	31461	39.797962	-84.2452
623	HOUSTON	HOUSTON	TX	26420	MAINSTREAM		46930	36741	29.74097	-95.3913
2541	NORWOOD	ARLINGTON	TX	19100	MAINSTREAM	300	29231	11019	32.679872	-97.1996
4489	SWEETWATER PLAZA	SUGAR LAND	TX	26420	UPSCALE		49030	21345	29.586305	-95.6323
10019	AT EASTEX FRWY	HOUSTON	TX	26420	VALUE	221	35675	19068	30.004891	-95.2835
13609	VANDALIA	VANDALIA	OH	19380	MAINSTREAM		55663	21232	39.901261	-84.1989
15531	GOSHEN	GOSHEN	OH	17140	VALUE		33667	13344	39.229923	-84.1572
17599	KEARNEY	MESQUITE	TX	19100	VALUE	300	36994	15134	32.770944	-96.5993
19265	LATONIA	COVINGTON	KY	17140	MAINSTREAM		55685	31578	38.94711	-84.49
21485	HOUSTON	KATY	TX	26420	MAINSTREAM		46369	26472	29.717926	-95.767
26981	SPRINGFIELD	SPRINGFIELD	OH	44220	MAINSTREAM		42081	25122	39.937167	-83.8351
4503	ROCKWALL	ROCKWALL	TX	19100	MAINSTREAM		56230	25289	32.882209	-96.4728

Products Lookup: This sheet gives all the information about the products or items available in these stores like their description, manufacturer, category the product belongs to, sub category and size, etc.

UPC	DESCRIPTION	MANUFACTURER	CATEGORY	SUB_CATEGORY	PRODUCT_SIZE
1111009477	PL MINI TWIST PRETZELS	PRIVATE LABEL	BAG SNACKS	PRETZELS	15 OZ
1111009497	PL PRETZEL STICKS	PRIVATE LABEL	BAG SNACKS	PRETZELS	15 OZ
1111009507	PL TWIST PRETZELS	PRIVATE LABEL	BAG SNACKS	PRETZELS	15 OZ
1111035398	PL BL MINT ANTSPTC RINSE	PRIVATE LABEL	ORAL HYGIENE PRODUCTS	MOUTHWASHES (ANTISEPTIC)	1.5 LT
1111038078	PL BL MINT ANTSPTC RINSE	PRIVATE LABEL	ORAL HYGIENE PRODUCTS	MOUTHWASHES (ANTISEPTIC)	500 ML
1111038080	PL ANTSPTC SPG MNT MTHWS	PRIVATE LABEL	ORAL HYGIENE PRODUCTS	MOUTHWASHES (ANTISEPTIC)	500 ML
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
1111087395	PL SR CRUST SUPRM PIZZA	PRIVATE LABEL	FROZEN PIZZA	PIZZA/PREMIUM	32.7 OZ
1111087396	PL SR CRUST 3 MEAT PIZZA	PRIVATE LABEL	FROZEN PIZZA	PIZZA/PREMIUM	30.5 OZ
1111087398	PL SR CRUST PEPPRN PIZZA	PRIVATE LABEL	FROZEN PIZZA	PIZZA/PREMIUM	29.6 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
2066200530	NWMN OWN PEPPERONI PIZZA	KING	FROZEN PIZZA	PIZZA/PREMIUM	13.2 OZ
2066200531	NWMN OWN 4 CHEESE PIZZA	KING	FROZEN PIZZA	PIZZA/PREMIUM	13.3 OZ
2066200532	NWMN OWN SUPREME PIZZA	KING	FROZEN PIZZA	PIZZA/PREMIUM	14.7 OZ

Transaction Data: This data consists of information regarding all the transactions that are carried out, like the store number, units, number of visits, sales, price, display items, featured items etc.

WEEK_END_DATE	STORE_NUM	UPC	UNITS	VISITS	HHS	SPEND	PRICE	BASE_PRICE	FEATURE	DISPLAY	TPR_ONLY
14,Jan,09	387	1111035398	4	4	4	14	3.5	4.49	0	0	1
14,Jan,09	387	1111038078	5	5	5	12.7	2.54	2.54	0	0	0
14,Jan,09	387	1111038080	6	5	5	15.3	2.55	2.55	0	0	0
14,Jan,09	387	1111085319	28	30	30	51.24	1.83	1.83	0	0	0
14,Jan,09	387	1111085345	38	35	35	71.44	1.88	1.88	0	0	0
14,Jan,09	387	1111085350	87	71	71	168.78	1.94	1.94	0	0	0
14,Jan,09	387	1111087395	19	18	18	62.13	3.27	3.91	0	0	1
14,Jan,09	387	1111087396	18	16	16	62.1	3.45	4	0	0	1
14,Jan,09	387	1111087398	24	16	16	79.44	3.31	3.97	0	0	1
14,Jan,09	387	1600027527	74	65	65	226.44	3.06	3.06	0	0	0
14,Jan,09	387	1600027528	52	51	49	220.48	4.24	4.24	0	0	0
14,Jan,09	387	1600027564	65	53	53	163.15	2.51	2.82	1	0	0
14,Jan,09	387	2840002333	10	9	9	28.9	2.89	2.89	0	0	0
14,Jan,09	387	2840004768	21	19	17	60.69	2.89	2.89	0	0	0
14,Jan,09	387	2840004770	7	7	7	20.23	2.89	2.89	0	0	0

MSA data: This is a text file, that helps us to identify the locations based on the US national msa location encoding,

```
10180,"Abilene, TX (Metropolitan Statistical Area)",48059,"Callahan, TX"
10180,"Abilene, TX (Metropolitan Statistical Area)",48253,"Jones, TX"
10180,"Abilene, TX (Metropolitan Statistical Area)",48441,"Taylor, TX"
10420,"Akron, OH (Metropolitan Statistical Area)",39133,"Portage, OH"
10420,"Akron, OH (Metropolitan Statistical Area)",39153,"Summit, OH"
10500,"Albany, GA (Metropolitan Statistical Area)",13007,"Baker, GA"
10500,"Albany, GA (Metropolitan Statistical Area)",13095,"Dougherty, GA"
10500,"Albany, GA (Metropolitan Statistical Area)",13177,"Lee, GA"
10500,"Albany, GA (Metropolitan Statistical Area)",13273,"Terrell, GA"
10500,"Albany, GA (Metropolitan Statistical Area)",13321,"Worth, GA"
10540,"Albany, OR (Metropolitan Statistical Area)",41043,"Linn, OR"
10580,"Albany-Schenectady-Troy, NY (Metropolitan Statistical Area)",36001,"Albany, NY"
```

Meaning of each attribute used:

VARIABLE NAME	TABLE	DESCRIPTION
ADDRESS_CITY_NAME	store lookup	city
ADDRESS_STATE_PROV_CODE	store lookup	state
AVG_WEEKLY_BASKETS	store lookup	average weekly baskets sold in the store
BASE_PRICE	data	base price of item
MANUFACTURER	products lookup	manufacturer
CATEGORY	products lookup	category of product
DESCRIPTION	products lookup	prodcut description
DISPLAY	data	product was a part of in-store promotional display
FEATURE	data	product was in in-store circular
HHS	data	# of purchasing households
MSA_CODE	store lookup	(Metropolitan Statistical Area) geographic region with a high core population density and close economic ties throughout the surrounding areas
PARKING_SPACE_QTY	store lookup	number of parking spaces in the store parking lot
PRICE	data	actual amount charged for the product at shelf
WEEK_END_DATE	data	week ending date
SALES_AREA_SIZE_NUM	store lookup	square footage of store
STORE_APPEAL	store lookup	Retailer's designated store appeal
SPEND	data	total spend (i.e., $ sales)
STORE_NUM	data, store lookup	store number
SUB_CATEGORY	products lookup	sub-category of product
TPR_ONLY	data	temporary price reduction only (i.e., shelf tag only, product was reduced in price but not on display or in an advertisment)
UNITS	data	units sold
UPC	data, products lookup	(Universal Product Code) product specific identifier
VISITS	data	number of unique purchases (baskets) that included the product
PRODUCT_SIZE	products lookup	package size or quantity of product

Connectivity:

Let's see how we connected all this data to be used in a single project. We used relational factor to connect all of them.

CollegeLife Data Challenge

Connection
⦿ Live   ◯ Extract

Products Lookup		Transaction Data		Store Lookup		msadata.txt

To connect Products Lookup and Transaction data we used the following relation:

Products Lookup	Operator	Transaction Data
# UPC	=	# UPC (Transaction Data)

To connect Transaction data and Store Lookup we used the following relation:

Transaction Data	Operator	Store Lookup
# Store Num	=	# Store Id

To connect Store Lookup and msadata.txt we used the following relation:

| Store Lookup	Operator	msadata.txt
# MSA	=	# F1

Sub Category Analysis
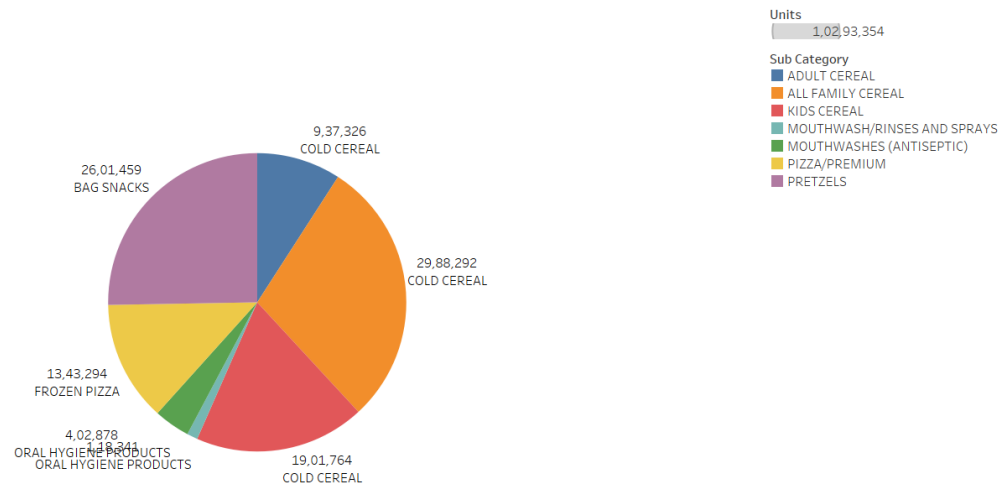
Sum of Units and Category. Color shows details about Sub Category. Size shows sum of Units. The marks are labeled by sum of Units and Category.

- The above pie chart is used to analyze sub categories.
- In the data we have 7 sub categories, here we go a step ahead to see how many units each sub category has and under which category does the subcategory fall into.
- Here we can observe that All family cereal which falls under Cold Cereal has the highest number of units sold.
- The trend is followed by the Pretzels category which falls under Bag Snacks category with 26,01,459 units sold.
- We can infer here that the oral hygiene category which consists of two subcategories Mouthwashes and Mouthwashes and Mouthwash/Rinses and sprays has the least amount of units sold overall, combining both the subcategories.
- We here come to know about the customers preferences of buying the items from the store. They frequently and in huge quantities buy cereals (the most, which may include cold, adult and kids cereal), followed by Snacks, Pizza and lastly oral hygiene products.
- We observe that the edible items are most frequently required by the customers and their consumption is way more than the oral hygiene products.

Statewise Store Analysis
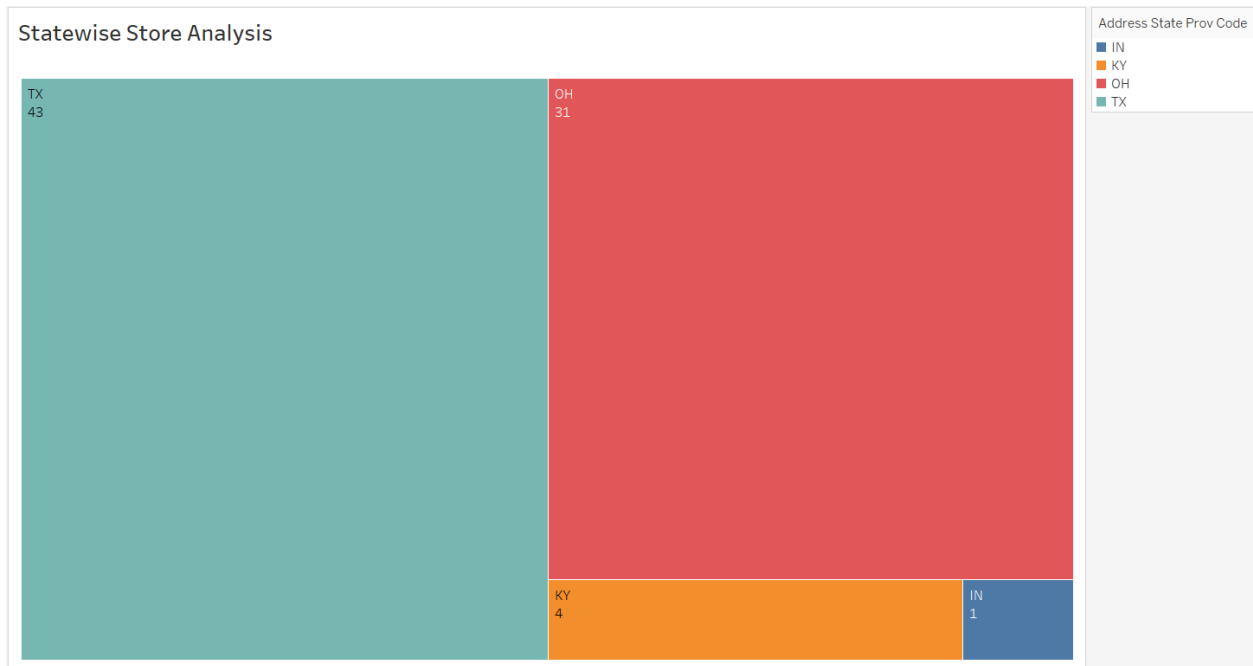
- In our data we have information of stores which are located in 4 different states of the USA.
- We can observe that maximum stores are present in Texas followed by Ohio, Kentucky and Indiana.
- We also found out that Ohio has the highest household items sold followed by texas, kentucky and indiana.
- Even with fewer stores in Ohio than Texas the number of household items sold in Ohio are more than kentucky.
- We can infer that sales in Ohio for households are more than in kentucky .
- People shop household more in ohio than kentucky

Visits VS Base Price VS Spend

Visit VS Base Price VS Spend
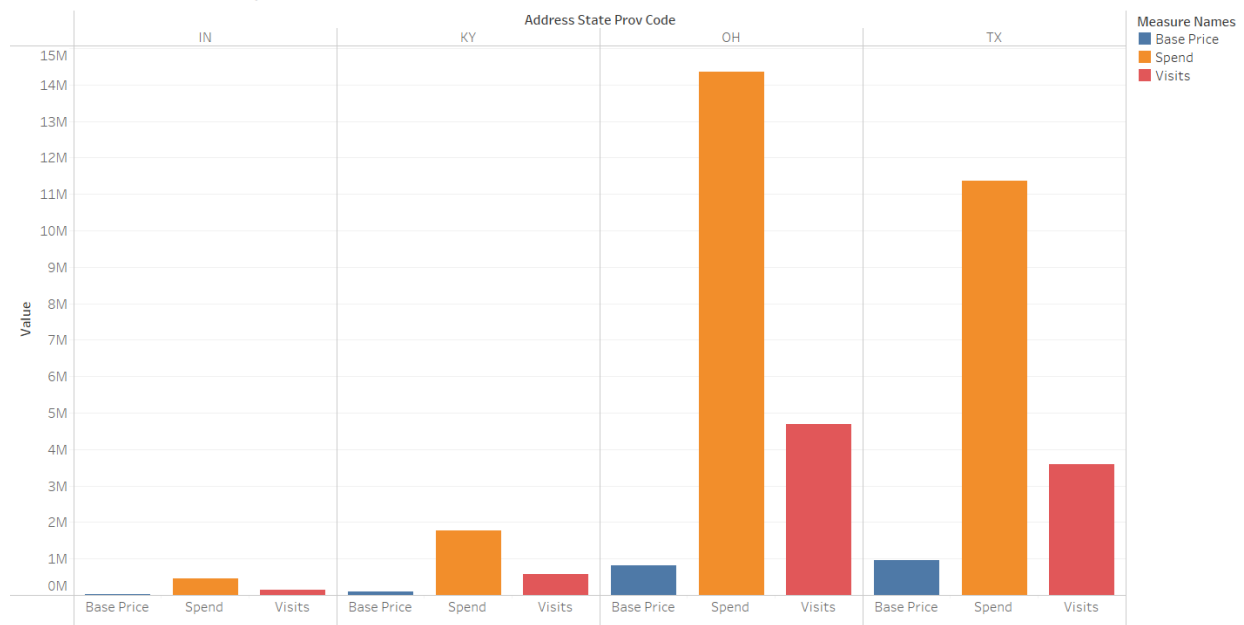
Base Price, Spend and Visits for each Address State Prov Code. Color shows details about Base Price, Spend and Visits.

- Here for 4 states we have seen the visits , base price and spend for all the stores in each state.
- We observe that Ohio has the maximum sales, followed by Texas, Kentucky and Indiana.
- The unique number of items are also picked the most in Ohio followed by Texas, Kentucky and Indiana.
- The base price for the products is more in Kentucky than compared to Ohio.
- From the above two analyses we can infer that the maximum sales are observed in Ohio compared to Texas.
- Although Texas has more stores than Ohio it doesn't make more sales.
- This can be possible due to higher base prices of products in Texas than Ohio.
- One recommendation that can be given is to increase stores in Ohio to increase even more sales and decrease the base price of products in Texas to lure more customers.
- On the other hand Indiana and Kentucky observe way less sales with less base prices. The reasons that can be inferred are that Indiana and Kentucky being smaller states might have few buyers or due to less number of stores the availability and accessibility of items may be less.

Effect of Feature, Display & TPR_ONLY on Visits

Effect of Feature, Display & TPR_ONLY on Visits

|	Visits
Display	57,815
Feature	44,251
Tpr Only	70,734
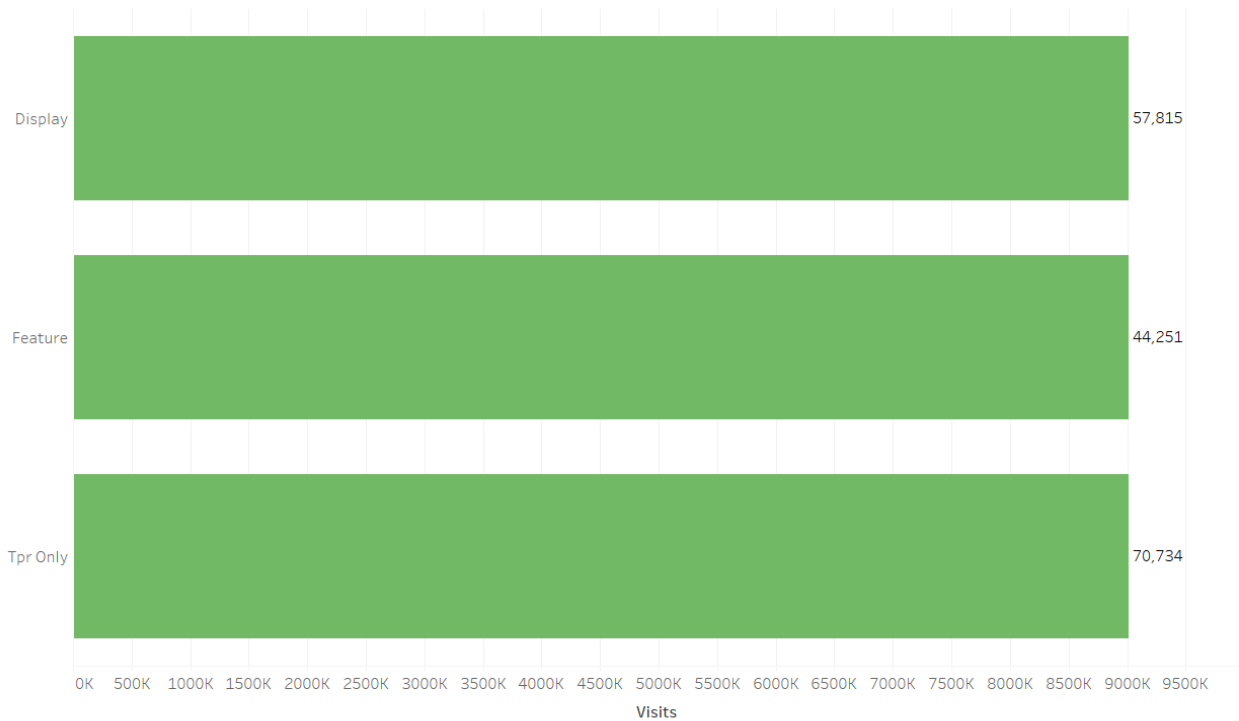
Sum of Visits for each Display, Feature and Tpr Only. The marks are labeled by Display, Feature and Tpr Only.

- In the above analysis we are observing the effects of features, display and TPR_only on Visits.
- Visits is the number of distinct purchases in the basket.
- Here we observe that the featured products are the ones that are sold the least.
- The products with temporary price reduction contribute to the maximum visits, certainly contributing to more sales.
- It can be inferred that the reduction in prices lure more customers and increase the sale of the same item that was once featured or displayed.

Sub Category/Manufacturer
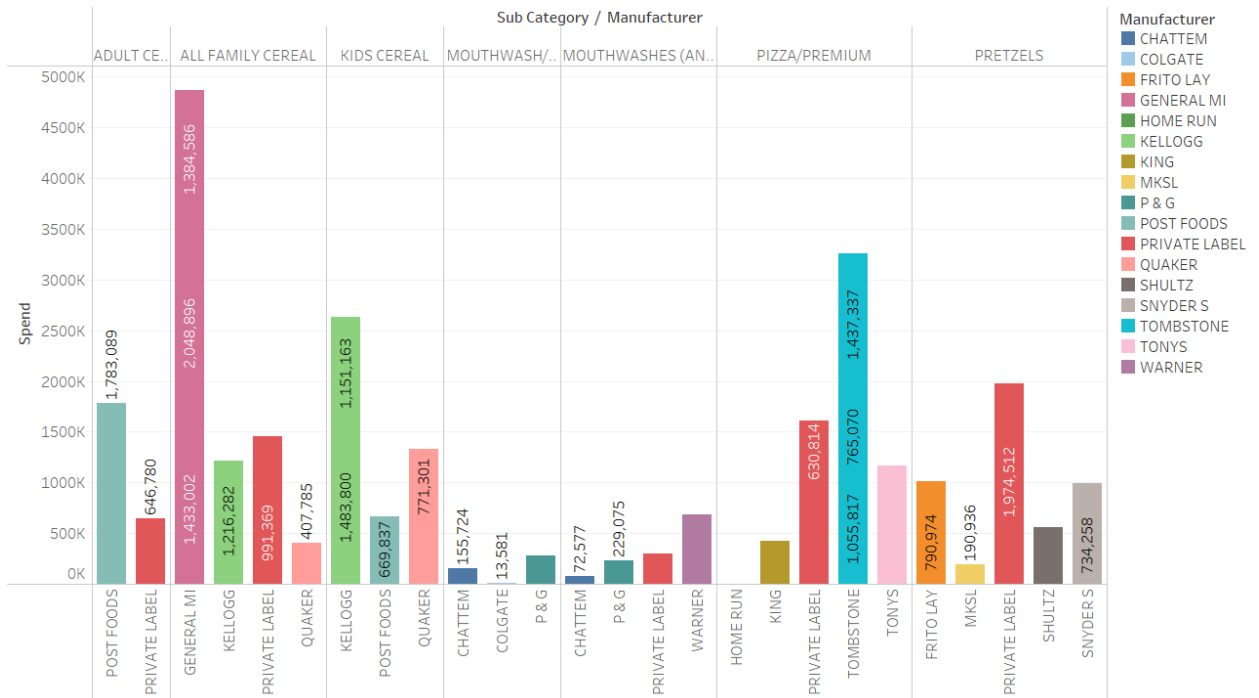
Sum of Spend for each Manufacturer broken down by Sub Category. Color shows details about Manufacturer. The marks are labeled by sum of Spend. Details are shown for Product Size.

- In the above graph we have analyzed Sub categories and Manufacturers.
- We observe that the general Mi manufacturer has the highest sales in the all family cereal category.
- Kellog and Quaker manufactures cereals for all families as well as kids.
- Post manufactures for adults and kids
- Cereal in general have the highest sales.
- Tombstone Pizzas are famous and sold the most.
- Small size pizzas have the highest revenue followed by large and then medium.
- For product size we observe that the customers tend to buy large or small over the medium sized items.

Visits with Respect to Area

Visits with respect to Area

Map based on Longitude (generated) and Latitude (generated). Color shows sum of Visits. Size shows details about Sales Area Size Num. The marks are labeled by Address State Prov Code. Details are shown for Address City Name and Address State Prov Code. The view is filtered on Address State Prov Code, Latitude (generated) and Longitude (generated). The Address State Prov Code filter excludes Null. The Latitude (generated) filter keeps all values. The Longitude (generated) filter keeps all values.

- We have carried out geospatial analysis here.
- More unique items sold are represented by dark green and less with yellow.
- We observe that Ohio, Kentucky, Indiana side people tend to buy more unique items.
- Near Texas people are comfortable with their daily items and don't try other brands or unique items.
- One recommendation given over here could be that to try out unique items in the market, one should keep them in stores near Ohio and Kentucky, as unique item sales are there and there are chances people try out these as well.
- Whereas in Texas the items that do a good sale should be continued as people are more comfortable with them.

State wise Sale Analysis

State wise Sale Analysis

Address State Prov Code. Color shows details about Address State Prov Code. Size shows sum of Spend. The marks are labeled by Address State Prov Code.
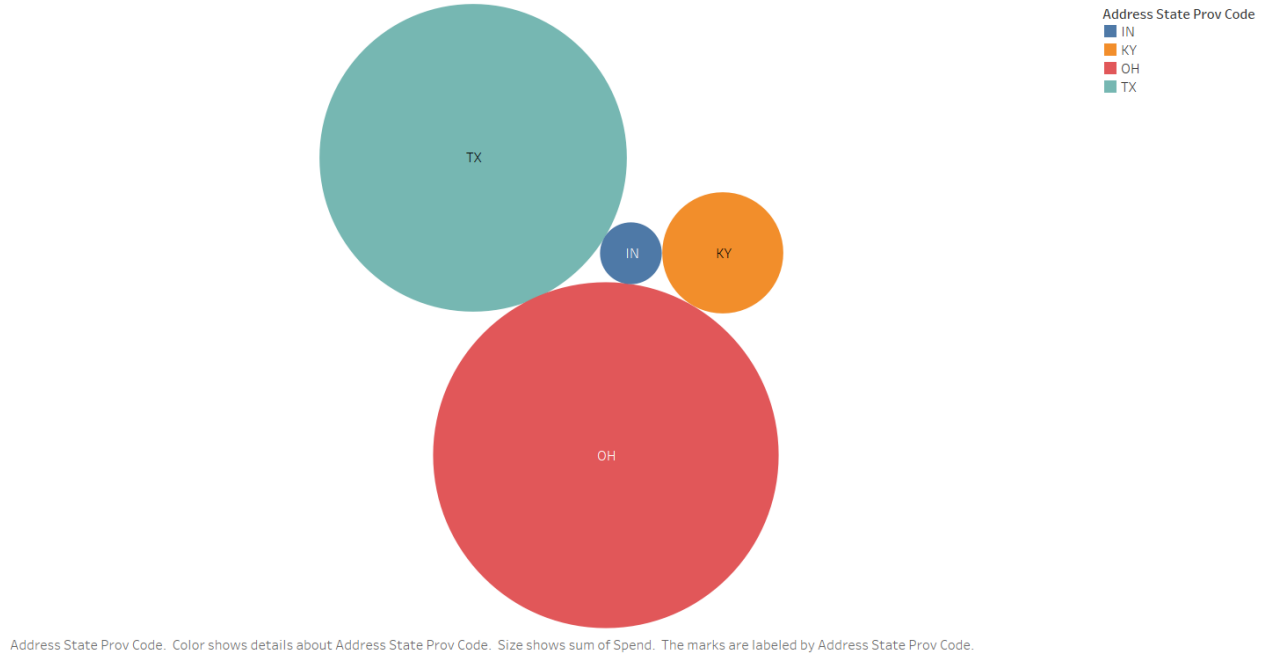
- We observe that the sales in Ohio and Texas are very close but Ohio has more sales.
- Although with less sales, the display, feature items are more in Texas as compared to ohio.
- Indiana has the least items and least sales.
- Recommendation that can be given here is that as observed Ohio has a better selling market more products should be focused (displayed and featured) here.

Sales Analysis

Sales Analysis
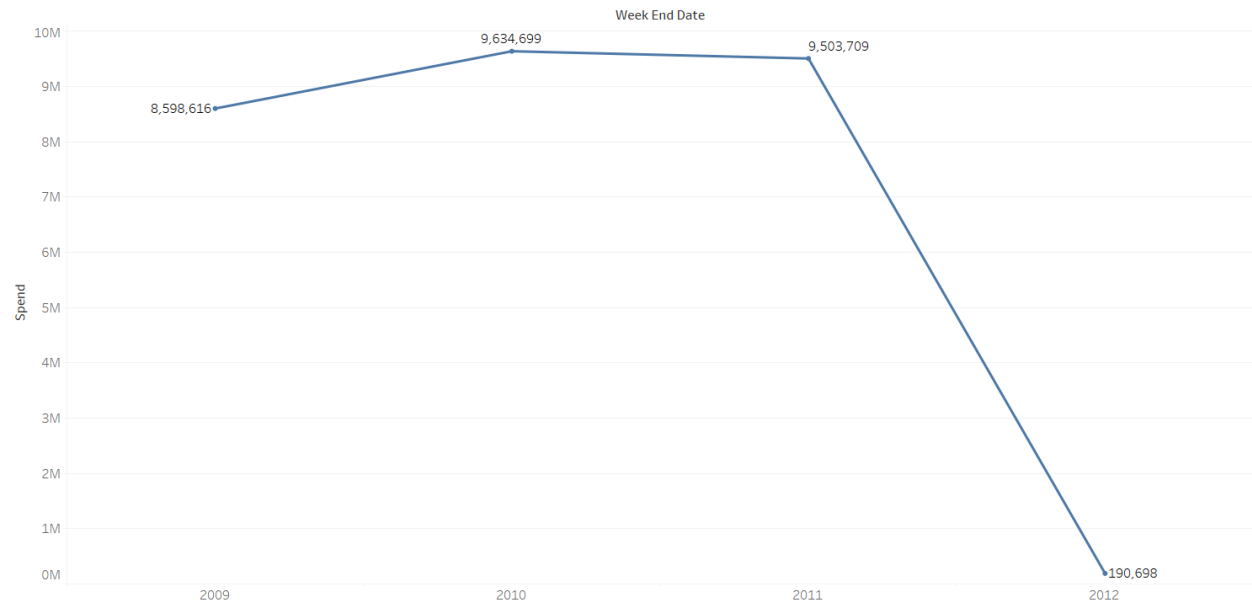
Week End Date

The trend of sum of Spend for Week End Date Year. The marks are labeled by sum of Spend.

- We observe a time series analysis for Week End Date Year of sales in this sheet.
- We can observe that for the weekend date for the year 2012 the sales are least.
- The drop is from 8598616 ro 190698
- The maximum was observed for 2010.
- A great loss can be inferred in 2012 from the drop seen.

3 Segment Analysis

3 Segment Analysis
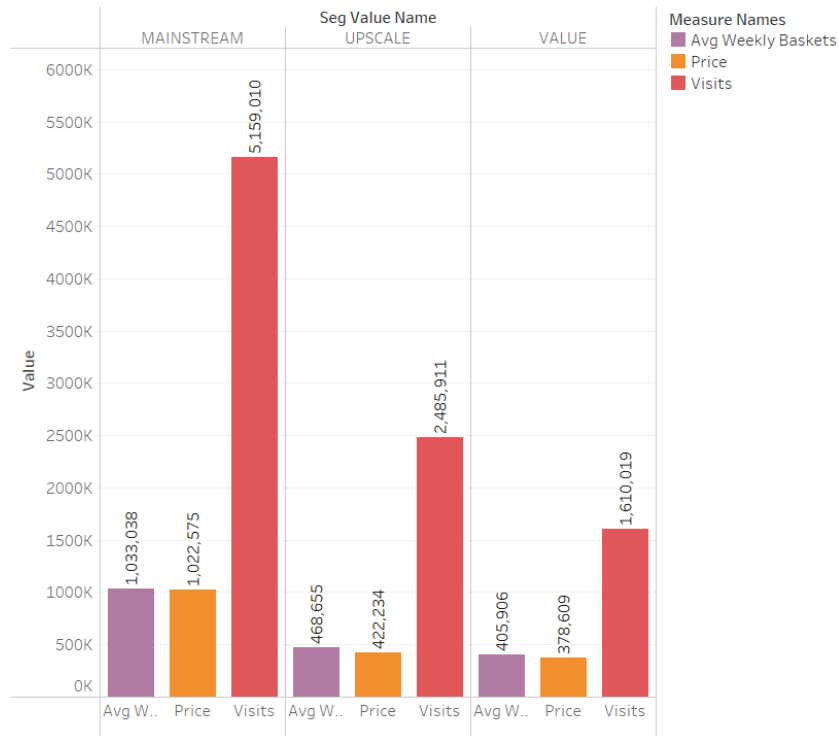
Avg Weekly Baskets, Price and Visits for each Seg Value Name. Color shows details about Avg Weekly Baskets, Price and Visits. The marks are labeled by Avg Weekly Baskets, Price and Visits.

- We have three segments : Mainstream, Upscale and Value.
- Mainstream has highest Visits followed by Upscale and Value.
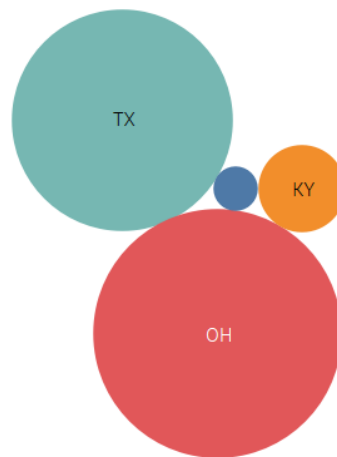- We observe that average weekly baskets are directly proportional to the price, for all three segments somewhat the same.

Dashboard 1

Sales + Store Analysis

Statewise Store Analysis
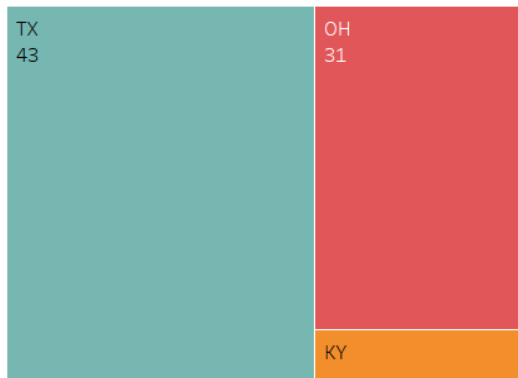
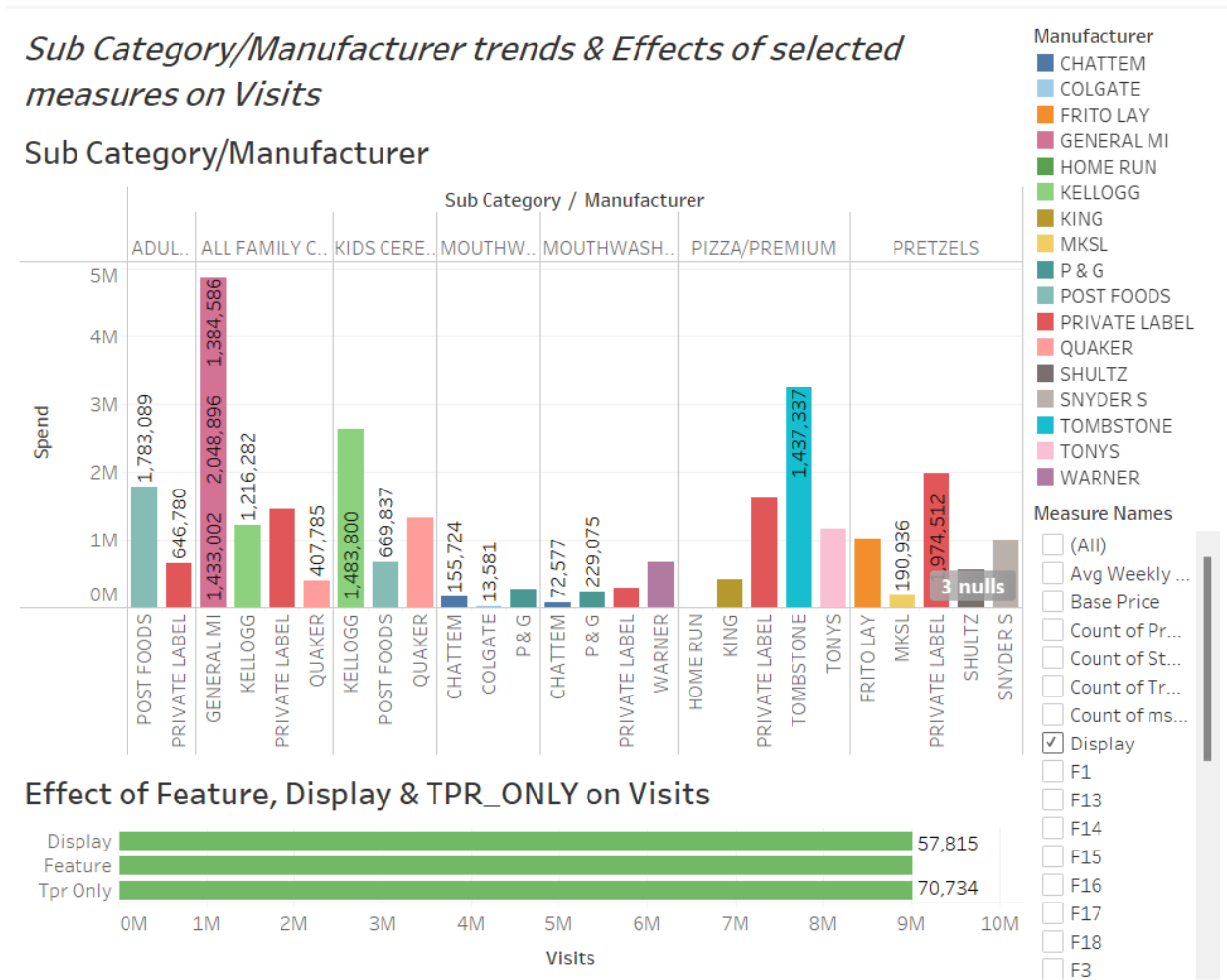### State wise Sale Analysis

Address State Prov ..
- IN
- KY
- OH
- TX

Sales Analysis
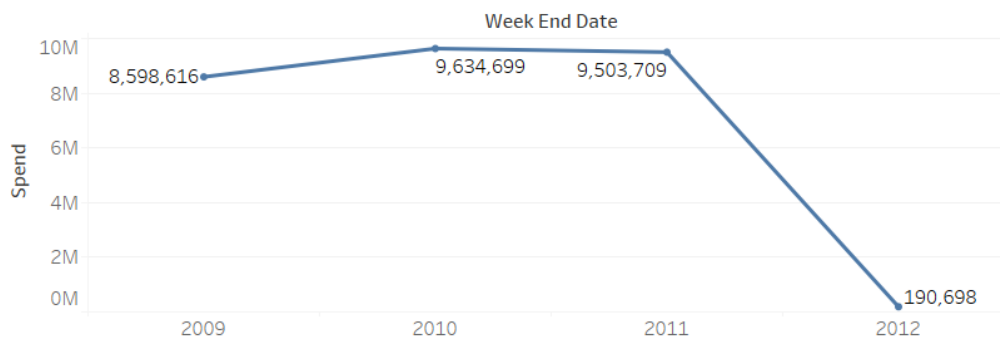
- The first dashboard represents sales and store analysis.
- We observe the sales trend and storewise analysis of sales in this dashboard.
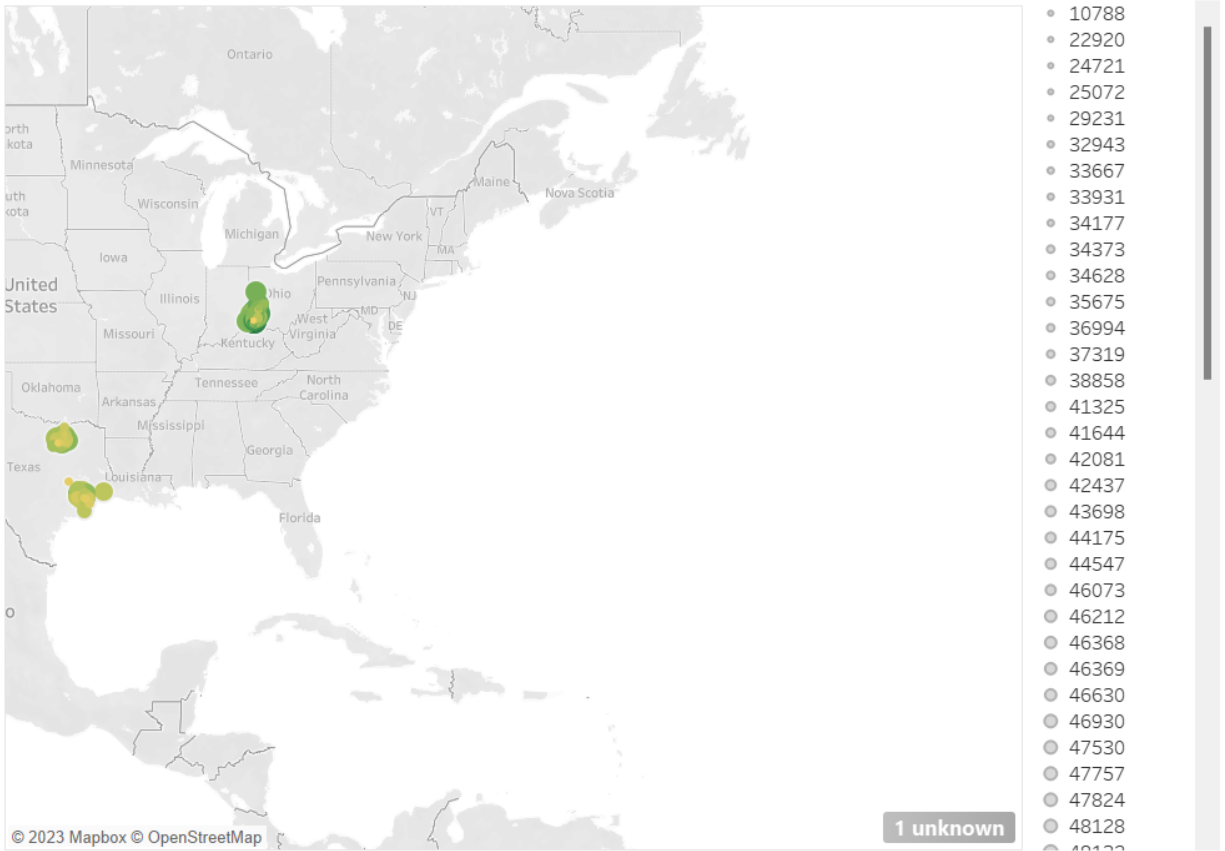
Dashboard 2

Sub Category/Manufacturer trends & Effects of selected measures on Visits

Sub Category/Manufacturer

- Here we observe the sab category/ manufacturer analysis on sales.
- We also observe the effects of different measures on Visits in this dashboard.

Dashboard 3

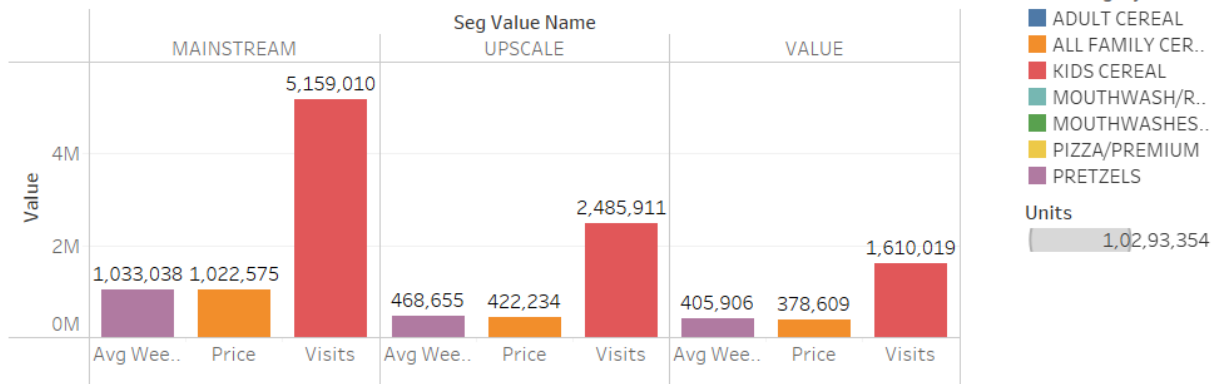
Geospatial Analysis of Stores

Visits with respect to Area

Visits

38,688   3,13,822

Sales Area Size Num
- 10788
- 22920
- 24721
- 25072
- 29231
- 32943
- 33667
- 33931
- 34177
- 34373
- 34628
- 35675
- 36994
- 37319
- 38858
- 41325
- 41644
- 42081
- 42437
- 43698
- 44175
- 44547
- 46073
- 46212
- 46368
- 46369
- 46630
- 46930
- 47530
- 47757
- 47824
- 48128

1 unknown

© 2023 Mapbox © OpenStreetMap

This Dashboard shows the geospatial analysis of stores in 4 states of USA and the visits in each state is shown by a gradient.
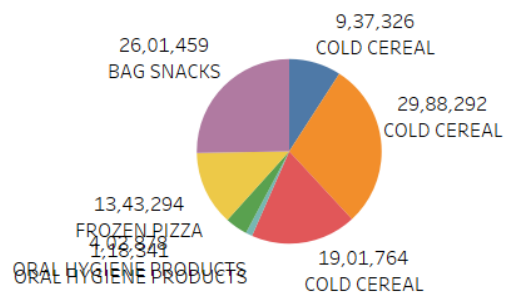
Dashboard 4

Sub category and segment analysis

3 Segment Analysis

Sub Category Analysis

Measure Names
- Avg Weekly Bas..
- Price
- Visits

Sub Category
- ADULT CEREAL
- ALL FAMILY CER..
- KIDS CEREAL
- MOUTHWASH/R..
- MOUTHWASHES..
- PIZZA/PREMIUM
- PRETZELS

Units
1,02,93,354

- In the above dashboard we see different segments and categories and analyze visits and sales for the same.

Dashboard 5

Recommendations

1] Stores are recommended to keep enough stock of Snack and Cereal(Edible Items)

2] Stores in Texas are recommended to decrease the base price of their item and match it with Ohio to increase the sales.

3] More Stores are recommended to be opened in Ohio, as it has maximum sales with lessser stores.

4] Stores are recommended to keep enough stock of "All Family Cereal", "Pretzels", and "Kids Cereal" sub-categories as they are highly popular and account for more than 60% of the units sold.

5] Temporary Price Reduction should be continued more often as it increases the sales and lures more customers.

6] More Unique items should be tried in Ohio, Kentucky stores as people buy more Unique items as compared to texas. Texas stores are recommended to provide items that are usually picked by the customers.
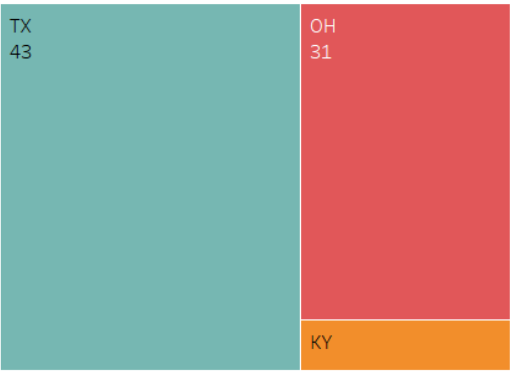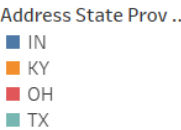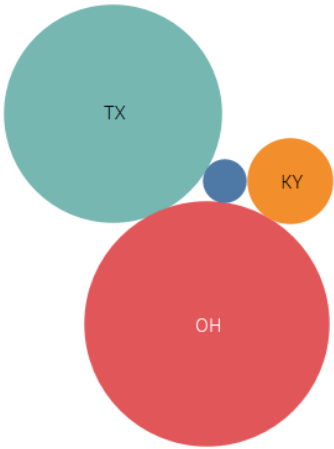
Final Story

Final Story

Sales + Store Analysis

Statewise Store Analysis

TX
43

OH
31

KY

State wise Sale Analysis

Address State Prov ..
- IN
- KY
- OH
- TX

TX

KY

OH

Sales Analysis

Week End Date

8,598,616   9,634,699   9,503,709

10M
8M
6M
4M

Spend