DATA VISUALIZATION USING TABLEAU

FINAL PROJECT

December 08, 2022.

INDIANA UNIVERSITY
(BLOOMINGTON)

Authored by: VARSHA R
IU ID: 2000751388
Sem: Fall 2022

Introduction

Data visualization is achieved using Tableau, a software package focusing on business intelligence (BI). The word tableau refers to a graphic representation or description. As a novice with Tableau, I have experimented with the software and built a few visualizations. I have tried to implement some of my learnings from this tableau course. The purpose of this report is to discuss these visualizations and derive meaningful recommendations based on sales and revenue. For each graph, the following information will be discussed:

- Plots built using the dataset.
- Plot analysis.

Dataset

For the visualizations, we will use the following given dataset. The data contained in the given file 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 four categories: mouthwash, pretzels, frozen pizza, and boxed cereal, the top five products from each of the top three brands in each category.

Name: #CollegeLife Data Challenge: A Time Series Analysis

Data glossary

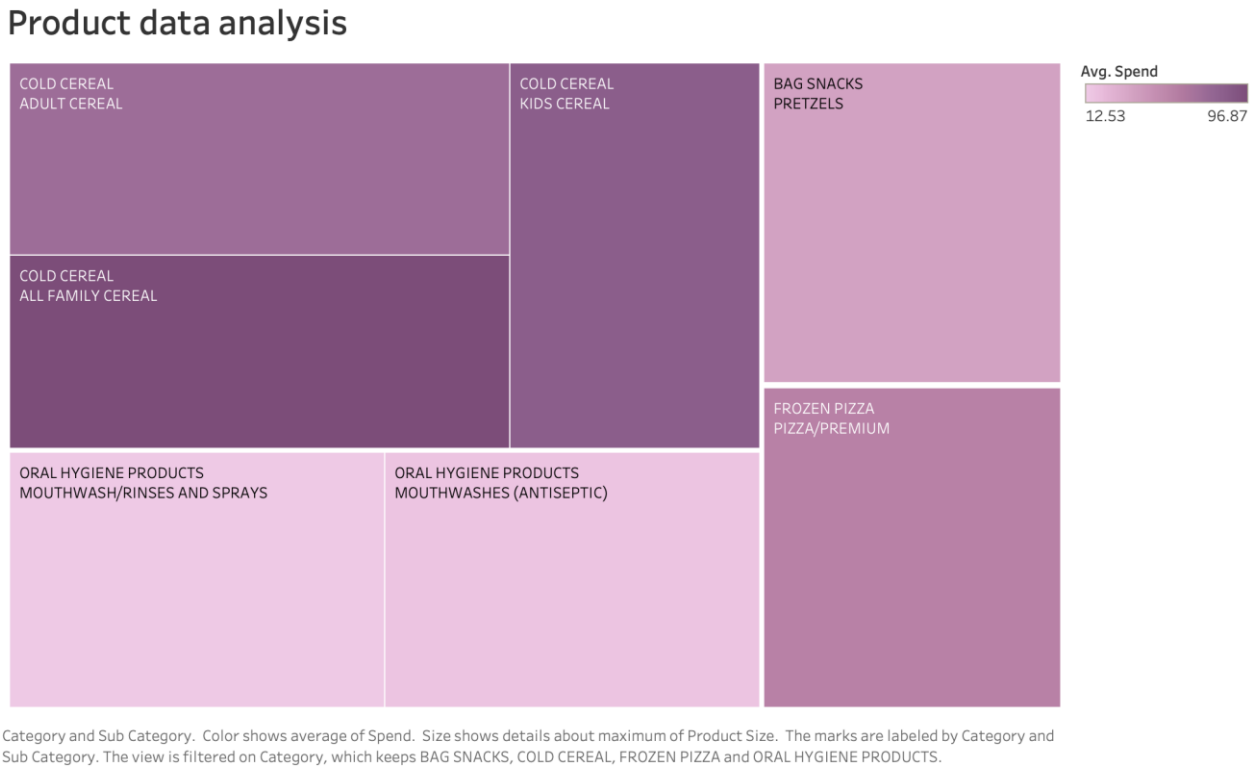
VARIABLE NAME	TABLE	DESCRIPTION
ADDRESS_CITY_NAME	store lookup	city
ADDRESS_STATE_PROV_CODE	store lookup	state
AVG_WEEKLY_BASKETS	store lookup	average weekly baskets sold in the store
BASE_PRICE	data	base price of item
MANUFACTURER	products lookup	manufacturer
CATEGORY	products lookup	category of product
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

Visualizations

Sheets:

1)

- **Plot built using this dataset**



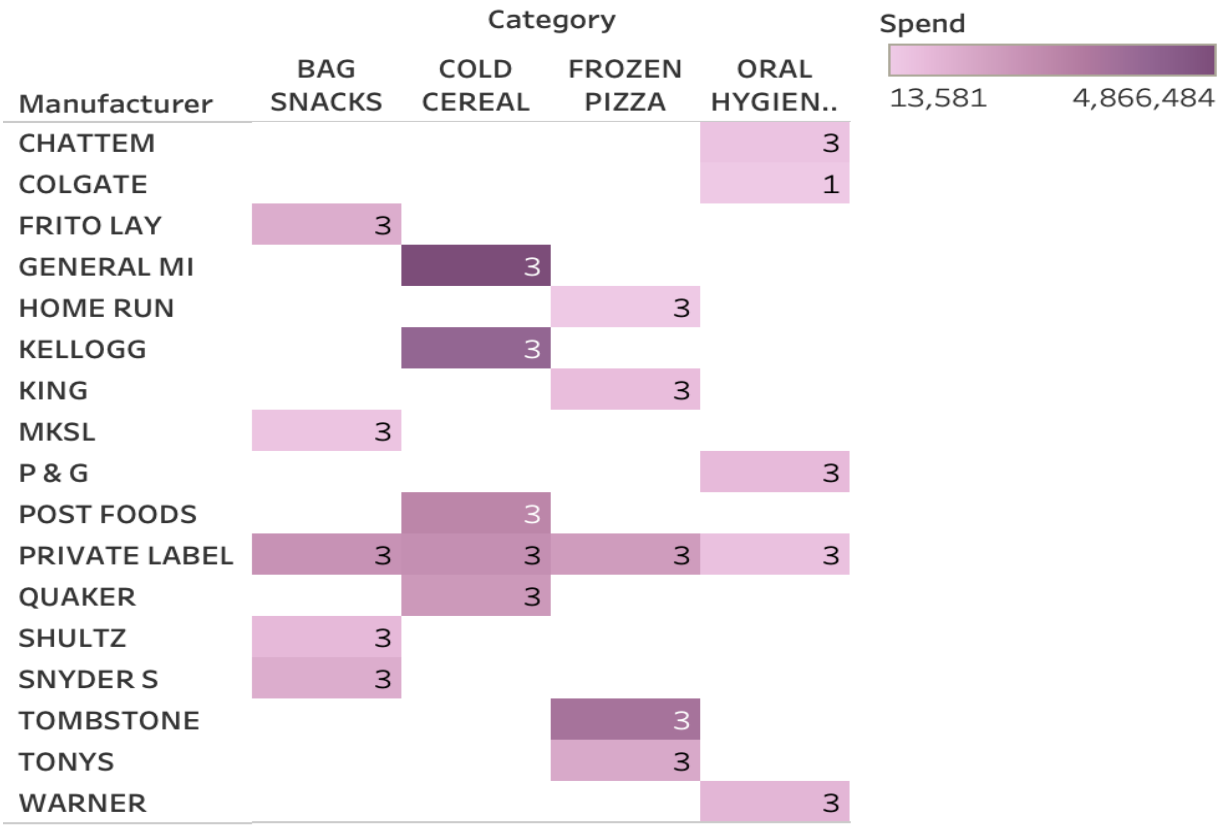
- **Plot analysis**

Attributes used for this plot: Category, sub-category, spend, product size
Marks: Labels – category, sub-category, Size – MAX(product size)Color – AVG(spend)
Filter - Category

Here we are trying to get a gist of what are the sizes of each product and also the products that most money was spent on by the customers (revenue).

2)

- **Plot built using this dataset**



Count of Manufacturer broken down by Category vs. Manufacturer. Color shows sum of Spend. The marks are labeled by count of Manufacturer. The view is filtered on Manufacturer, which keeps 17 of 17 members.

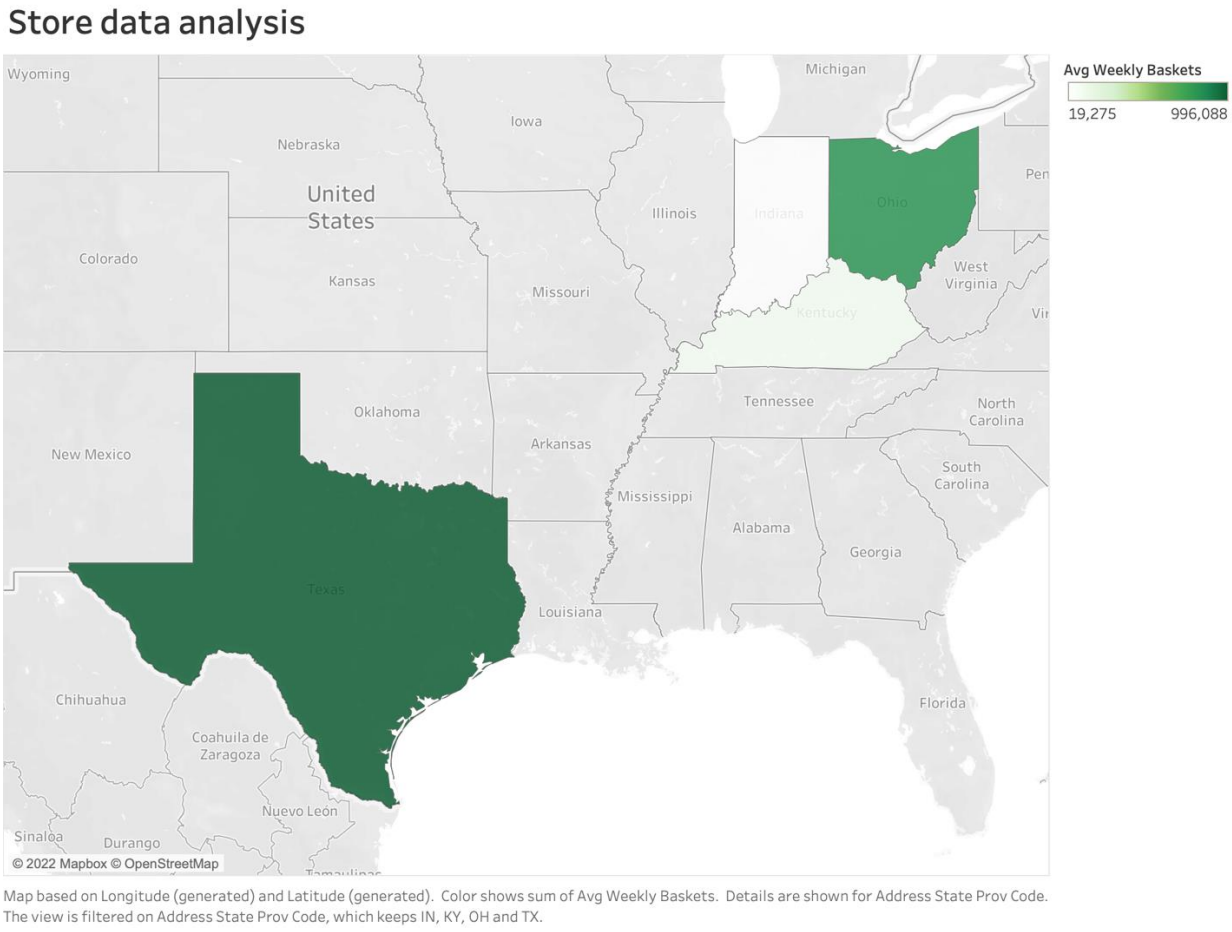
- **Plot analysis**

Attributes used for this plot: Category, manufacturer, spend
Columns: category
Rows: manufacturer
Marks: Color – SUM(spend), label - CNT(manufacturer)
Filter: Manufacturer

Here we are trying to get a count of how many products are being sold by a particular manufacturer and also which products are making the most revenue.

3)

- **Plot built using this dataset**



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Avg Weekly Baskets. Details are shown for Address State Prov Code. The view is filtered on Address State Prov Code, which keeps IN, KY, OH and TX.

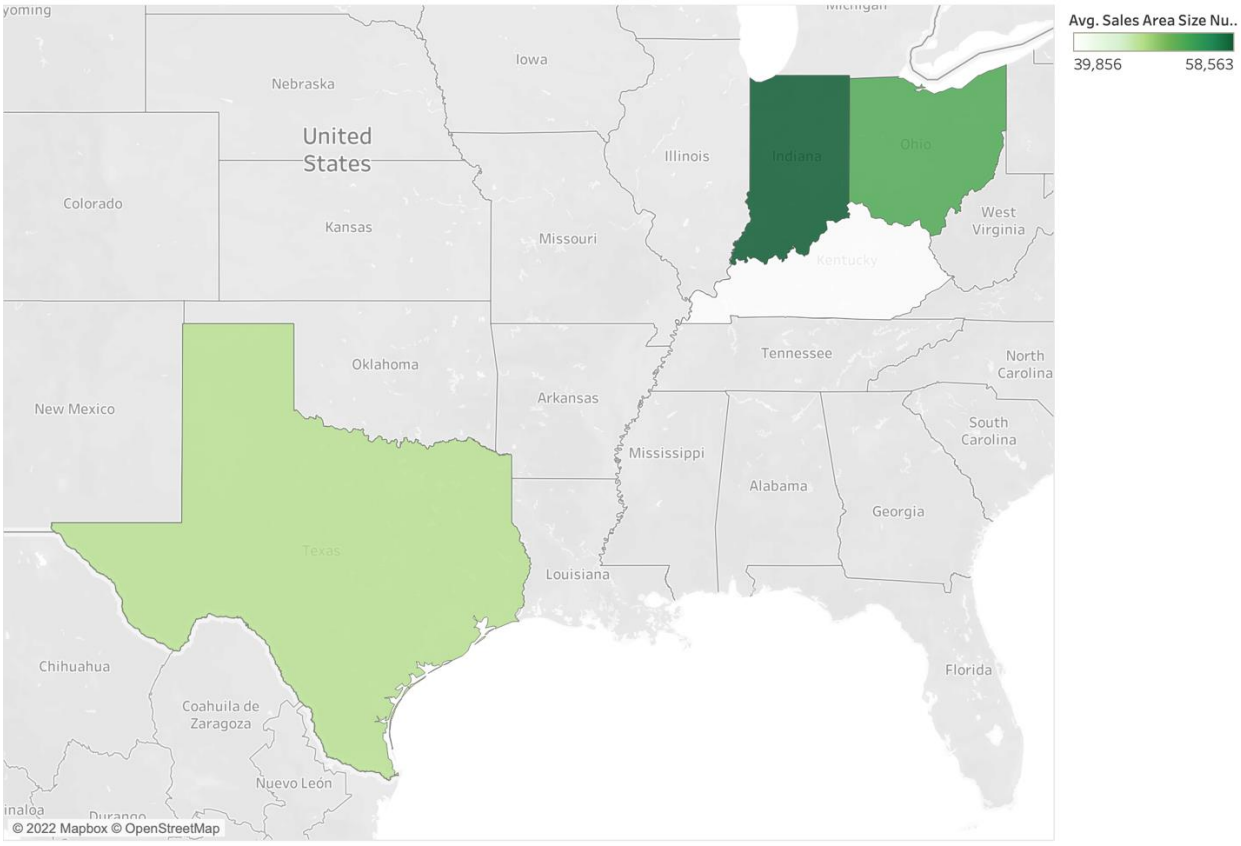
- **Plot analysis**

Attributes used for this plot: longitude, latitude, address state prov code, average weekly basket
Columns: Longitude
Rows: Latitude
Marks: Color – AVG(average weekly basket), Details - address state prov code
Filter: address state prov code

Here we are trying to check to see which states are having the highest on average weekly basket sales. According to the graph, Texas has the highest weekly basket sales on average.

4)

- **Plot built using this dataset**



Map based on Longitude (generated) and Latitude (generated). Color shows average of Sales Area Size Num. Details are shown for Address State Prov Code. The view is filtered on Address State Prov Code, which keeps IN, KY, OH and TX.

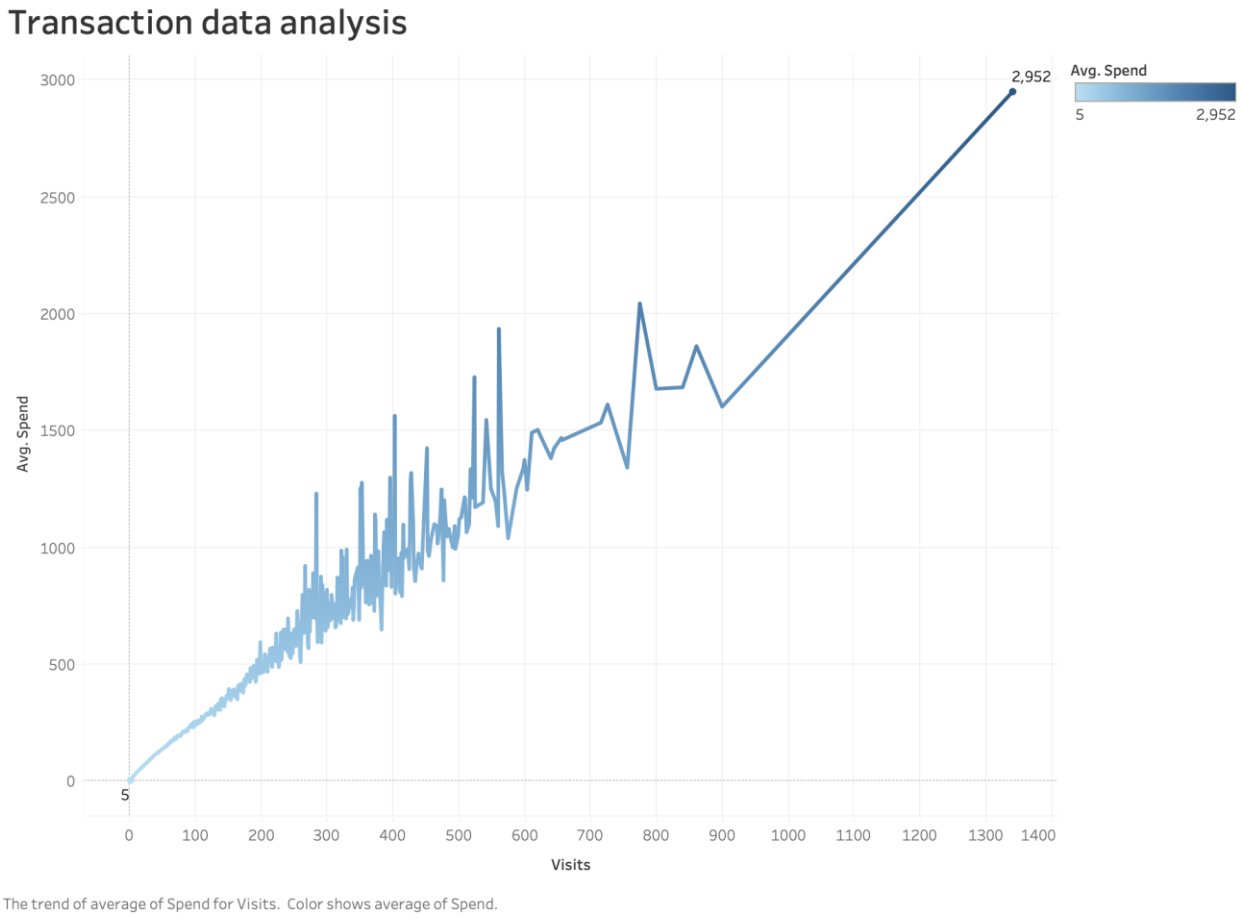
- **Plot analysis**

Attributes used for this plot: longitude, latitude, address state prov code, sales area size number, address state prov code
Columns: Longitude
Rows: Latitude
Marks: Color – AVG(sales area size number), Details - address state prov code
Filter: address state prov code

Here we are trying to see which state has larger sized stores on average and the results say that Indiana has the larger stores compared to the other states.

5)

- **Plot built using this dataset**



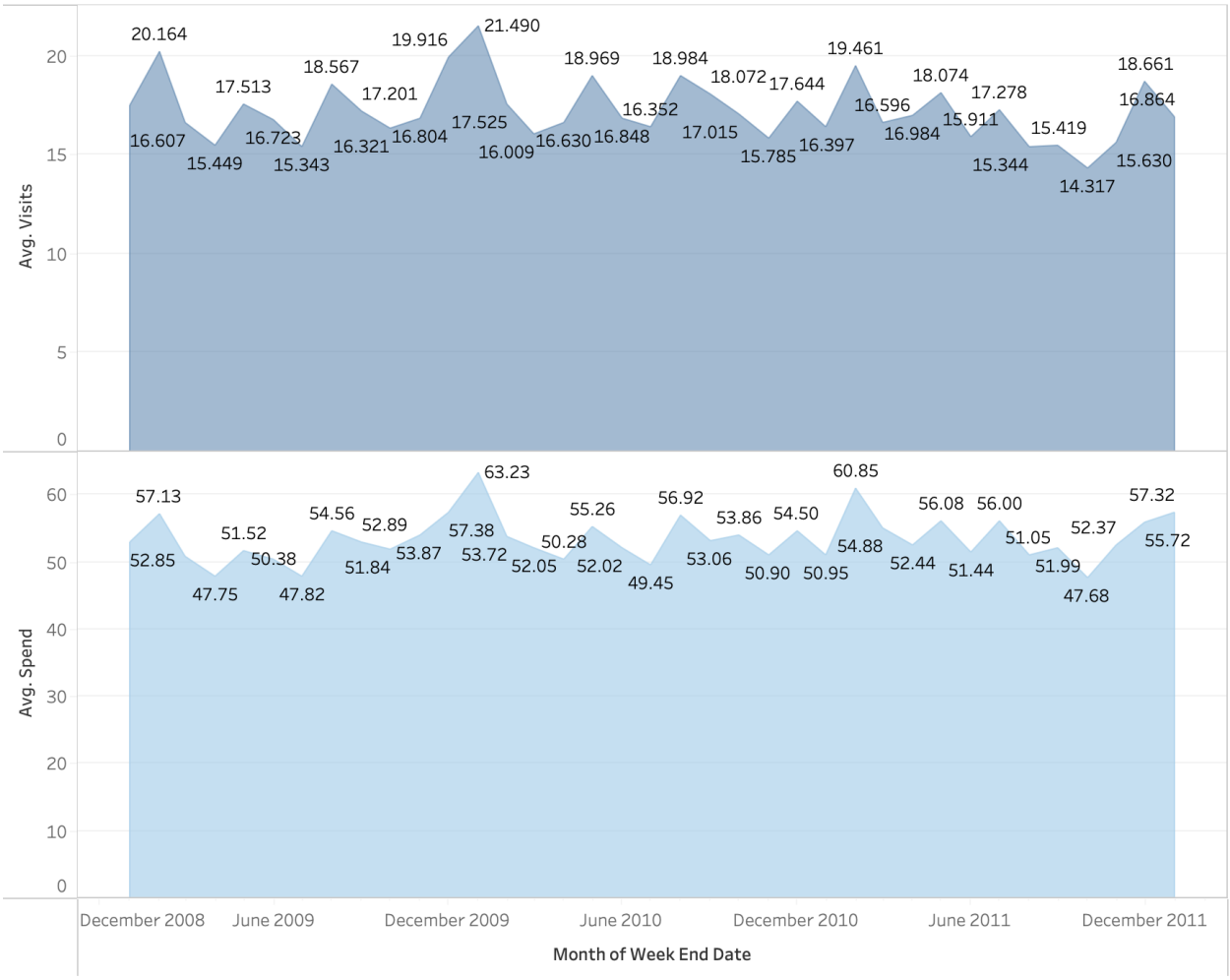
- **Plot analysis**

Attributes used for this plot: visits, spend
Columns: visits
Rows: AVG(spend)
Marks: Color – AVG(spend)

Here we are trying to see how number of visits increases the number of purchases by the customer by analyzing the amount of money spent by them. As a clear result, we can see that as the number of visits increase, the amount of spent accumulates to being a huge amount.

6)

- ## Plot built using this dataset



The plots of average of Visits and average of Spend for Week End Date Month. The data is filtered on Week End Date Year and Week End Date Month. The Week End Date Year filter keeps 2009, 2010, 2011 and 2012. The Week End Date Month filter keeps 12 of 12 members.

- ## Plot analysis

Attributes used for this plot: Weekend date, visits, spend
Columns: MONTH(weekend date)
Rows: AVG(visits), AVG(spend)
Filter: YEAR(weekend date), MONTH(weekend date)

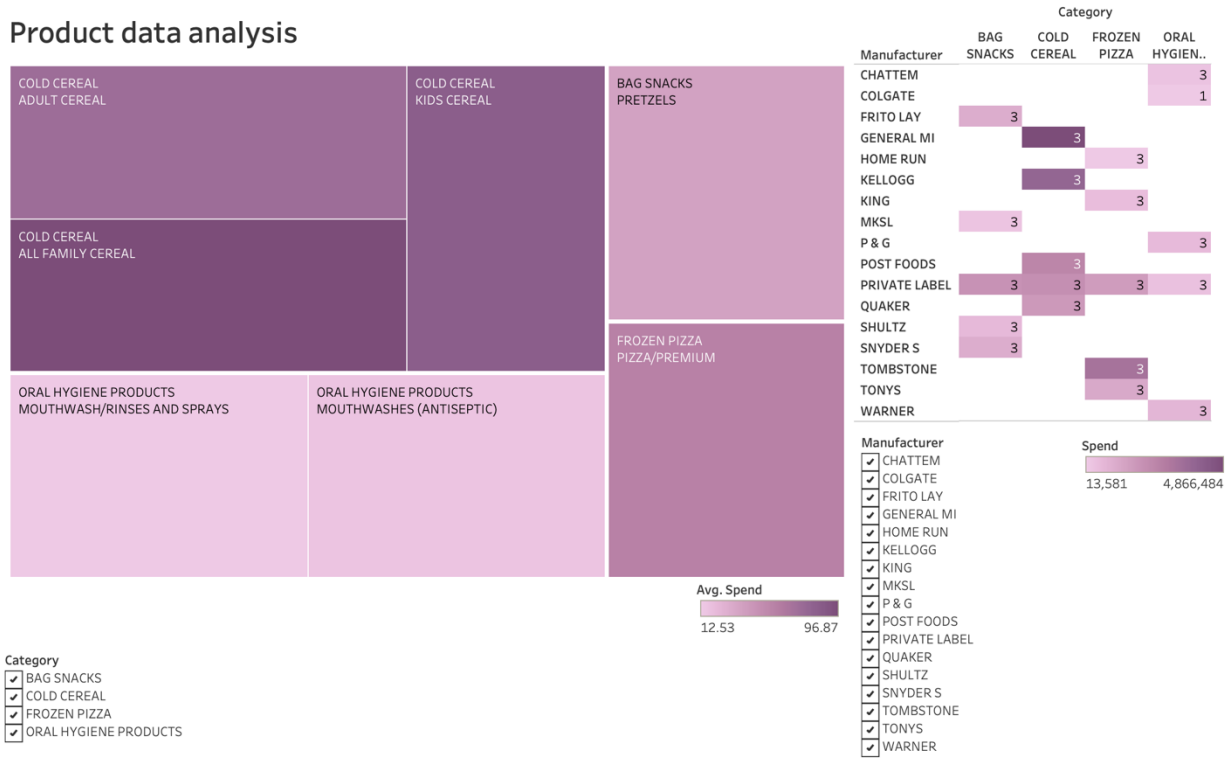
We are trying to plot and analyze the number of visits and the spending through different times of the year for around 3 years. We use time-series analysis for this purpose.

Dashboards:

Dashboards are a combinations on one or more worksheets. Multiple worksheets can be merged into a single dashboards. Tableau worksheets are the primary building blocks for the dashboard. In this assignment, there are 3 dashboards created after grouping a couple of sheets for each dashboard based on their similarity of focus attribute.

1)

- ### Dashboard



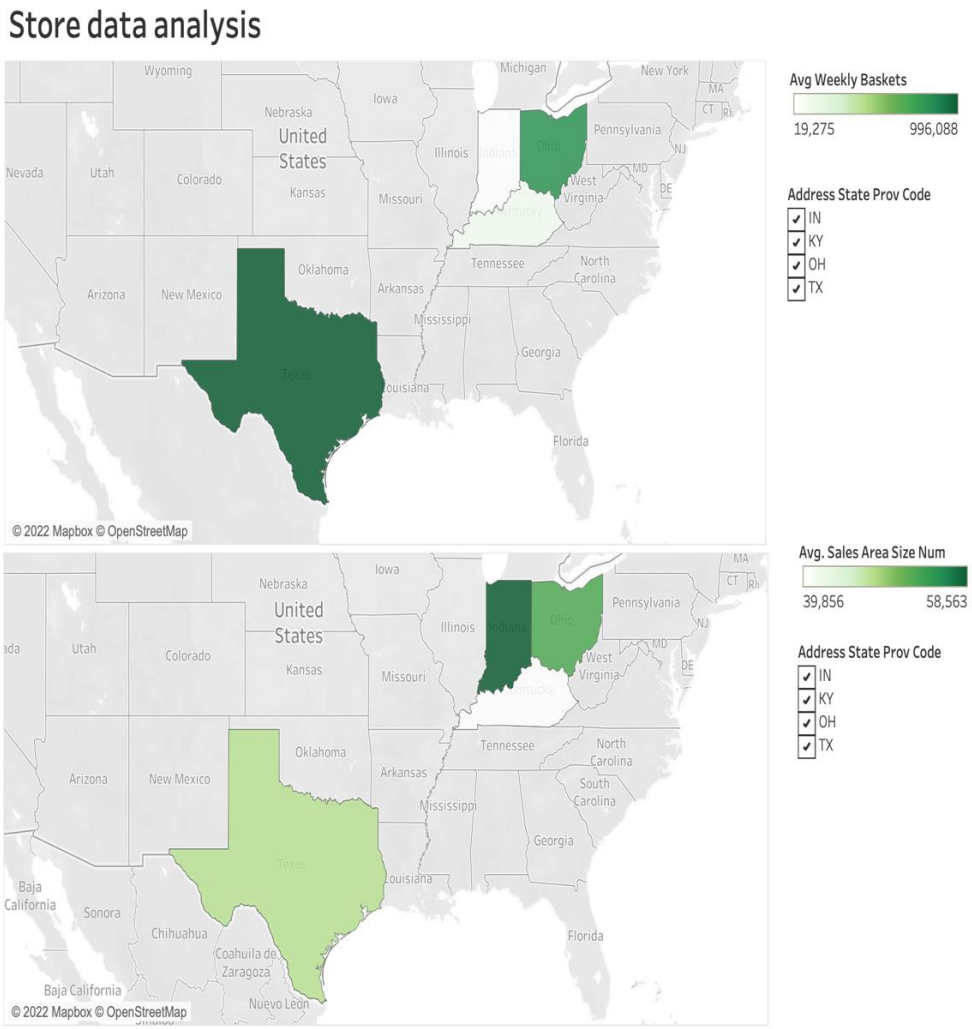
- ### Analysis

Grouping attribute: Products

As you can see, we are conducting a product data analysis using the product dataset. By these visualizations, we can derive insights such as cereals being the highest selling products compared to the other products. We can also see that GENERAL MI are the highest earning manufacturers on single product basis whereas PRIVATE LABEL is the overall bestselling manufacturer having products under every category.

2)

- **Dashboard**



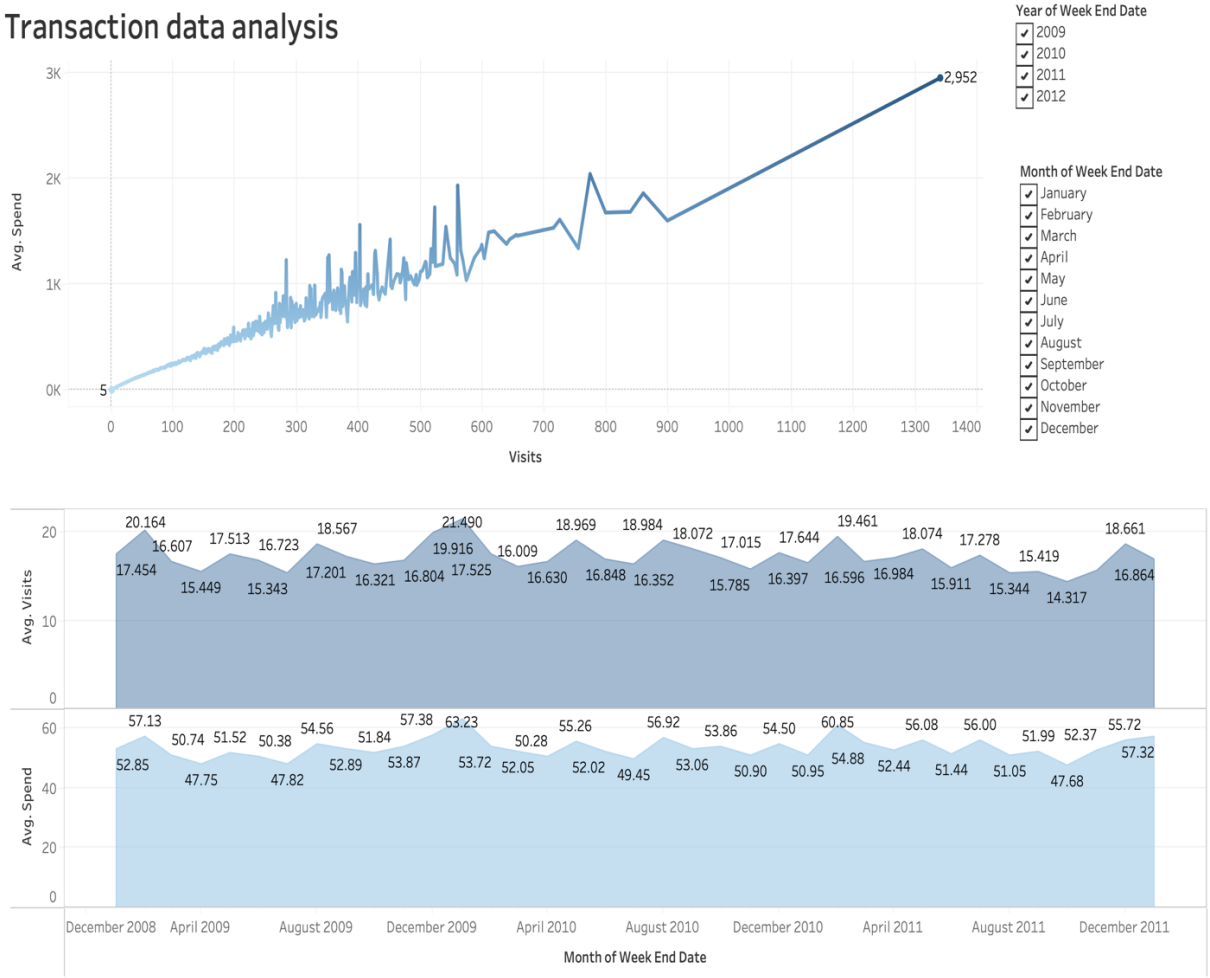
- **Analysis**

Grouping attribute: Stores

Here we are analyzing store data from the store-lookup dataset. In this dashboard I have placed both visualizations besides each other so that it becomes easy to compare and derive insights. We can see that there is a weird pattern on how on an average even though Indiana has more larger stores, Texas has the most weekly basket sales. After comparing both data visuals, it is safe to say that Texas and Ohio are the safe zones for number of sales and revenue.

3)

○ Dashboard



○ Analysis

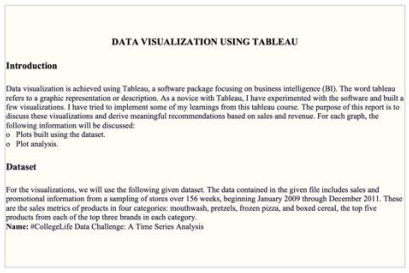
Grouping attribute: Transactions

Here we are trying to get a relationship between number of visits and the amount spent by the customers. The visualizations clearly show that as the number of visits increase, the more revenue was made. The earning somehow stays consistent throughout the years which is not a bad thing but it is also not a good thing since the stores and manufacturers would want to earn more as time goes. They should lure more customers by providing extra offers, which will lead to increased revenue.

Storyboard:

In Tableau, a story is a sequence of visualization that work together to convey information. Either sheets or dashboards can be put together in the storyboard to present the information in the most efficient and convenient way. In my assignment, the best way to build the storyboard was to put all the dashboards together in 3 story points: products, stores, and transactions. The other 3 story points are text based: Introduction, analysis overview and finishing. Here is a summary of the storyboard put together as images.

DATA ANALYSIS OVERVIEW



DATA VISUALIZATION USING TABLEAU

Introduction

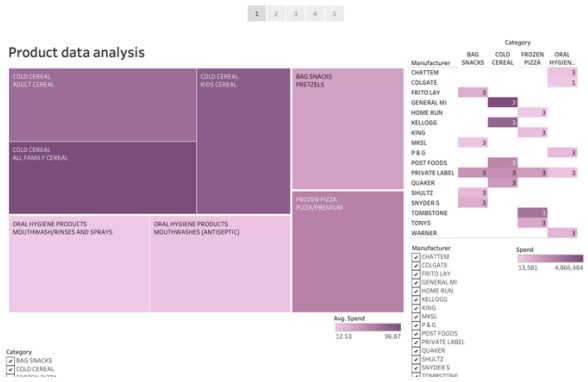
Data visualization is achieved using Tableau, a software package focusing on business intelligence (BI). The word tableau refers to a graphic representation or description. As a novice with Tableau, I have experimented with the software and built a few visualizations. I have tried to implement some of my learnings from this tableau course. The purpose of this report is to discuss these visualizations and derive meaningful recommendations based on sales and revenue. For each graph, the following information will be discussed:

- Plots built using the dataset.
- Plot analysis.

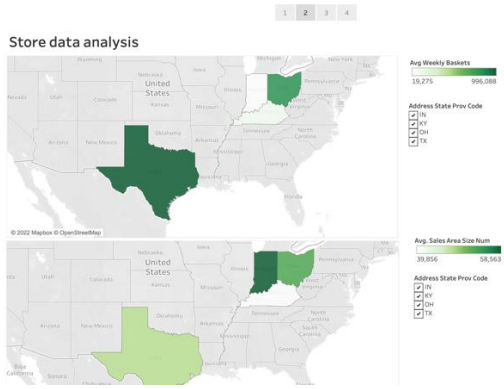
Dataset

For the visualizations, we will use the following given dataset. The data contained in the given file 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 four categories: mouthwash, pretzels, frozen pizza, and boxed cereal, the top five products from each of the top three brands in each category.
Name: #CollegeLife Data Challenge: A Time Series Analysis

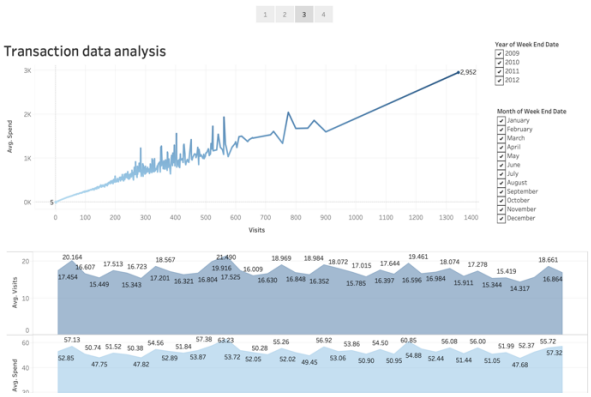
DATA ANALYSIS OVERVIEW



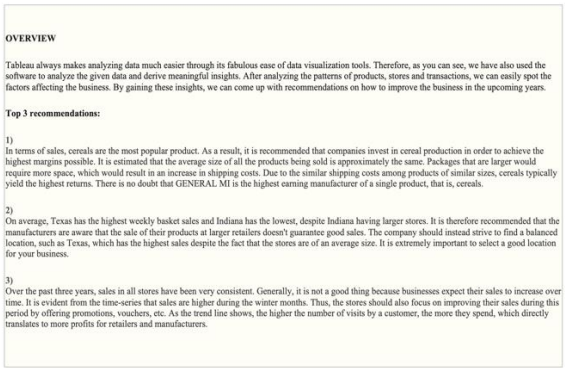
DATA ANALYSIS OVERVIEW



DATA ANALYSIS OVERVIEW



DATA ANALYSIS OVERVIEW



OVERVIEW

Tableau always makes analyzing data much easier through its fabulous ease of data visualization tools. Therefore, as you can see, we have also used the software to analyze the given data and derive meaningful insights. After analyzing the patterns of products, stores and transactions, we can easily spot the factors affecting the business. By gaining these insights, we can come up with recommendations on how to improve the business in the upcoming years.

Top 3 recommendations:

1)
In terms of sales, cereals are the most popular product. As a result, it is recommended that companies invest in cereal production in order to achieve the highest margins possible. It is estimated that the average size of all the products being sold is approximately the same. Packages that are larger would require more space, which would result in an increase in shipping costs. Due to the similar shipping costs among products of similar sizes, cereals typically yield the highest returns. There is no doubt that GENERAL MI is the highest earning manufacturer of a single product, that is, cereals.

2)
On average, Texas has the highest weekly basket sales and Indiana has the lowest, despite Indiana having larger stores. It is therefore recommended that the manufacturers are aware that the sale of their products at larger retailers doesn't guarantee good sales. The company should instead strive to find a balanced location, such as Texas, which has the highest sales despite the fact that the stores are of an average size. It is extremely important to select a good location for your business.

3)
Over the past three years, sales in all stores have been very consistent. Generally, it is not a good thing because businesses expect their sales to increase over time. It is evident from the time-series that sales are higher during the winter months. Thus, the stores should also focus on improving their sales during this period by offering promotions, vouchers, etc. As the trend line shows, the higher the number of visits by a customer, the more they spend, which directly translates to more profits for retailers and manufacturers.

DATA ANALYSIS OVERVIEW

The analysis was made and was mentioned in the tableau workbook including the three top recommendations. I will mention the same in this report as well. Therefore, here are the 3 top recommendations that are mentioned in the workbook.

Top 3 recommendations:

- In terms of sales, cereals are the most popular product. As a result, it is recommended that companies invest in cereal production in order to achieve the highest margins possible. It is estimated that the average size of all the products being sold is approximately the same. Packages that are larger would require more space, which would result in an increase in shipping costs. Due to the similar shipping costs among products of similar sizes, cereals typically yield the highest returns. There is no doubt that GENERAL MI is the highest earning manufacturer of a single product, that is, cereals.

- On average, Texas has the highest weekly basket sales and Indiana has the lowest, despite Indiana having larger stores. It is therefore recommended that the manufacturers are aware that the sale of their products at larger retailers doesn't guarantee good sales. The company should instead strive to find a balanced location, such as Texas, which has the highest sales despite the fact that the stores are of an average size. It is extremely important to select a good location for your business.

- Over the past three years, sales in all stores have been very consistent. Generally, it is not a good thing because businesses expect their sales to increase over time. It is evident from the time-series that sales are higher during the winter months. Thus, the stores should also focus on improving their sales during this period by offering promotions, vouchers, etc. As the trend line shows, the higher the number of visits by a customer, the more they spend, which directly translates to more profits for retailers and manufacturers.

FIN.