Data analyst project

RETAIL TRANSACTIONS

DATA ANALYTICS DASHBOARD

Dashboard:
https://lookerstudio.google.com/reporting/68ef15f4-d651-43ac-ad0c-123d664fcc92

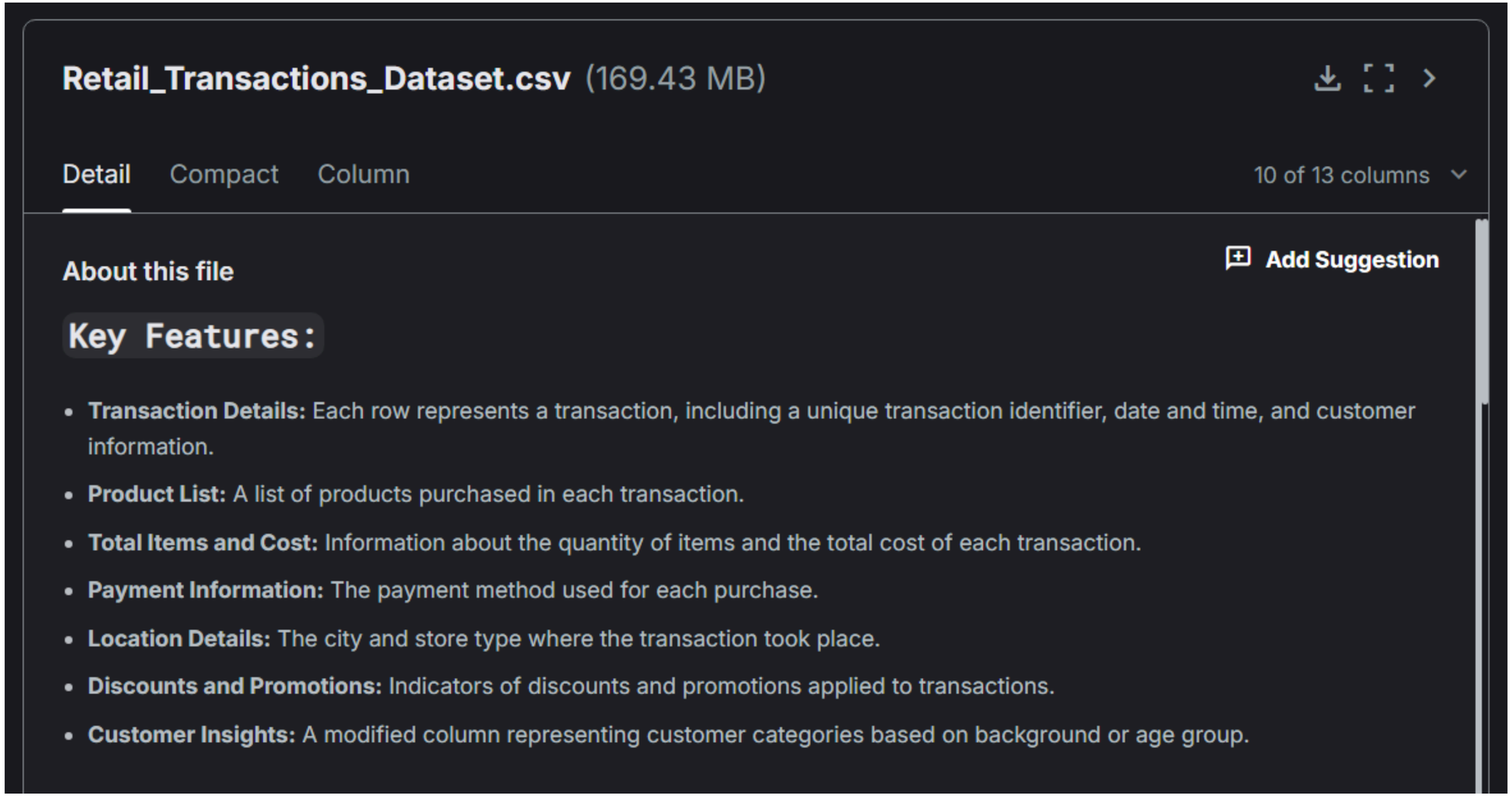
by Siwat Srisuddee

siwat.srisuddee@gmail.com

https://www.linkedin.com/in/siwat-sri/

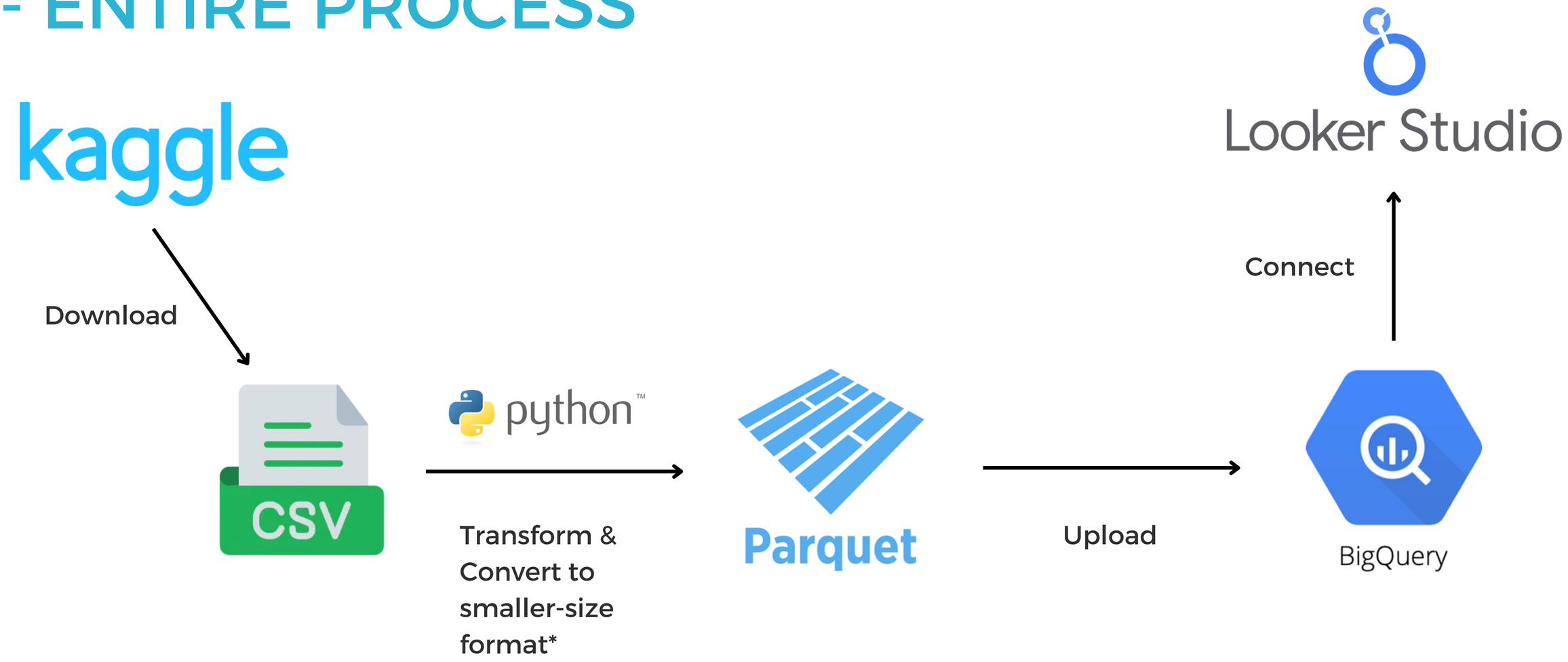
ABOUT THE DATA SET

Data source: https://www.kaggle.com/datasets/prasad22/retail-transactions-dataset



The data has 1,000,000 records

DATA PREPARATION
- ENTIRE PROCESS



*because Looker Studio accept data sources which must not more than 100 Mb.

DATA PREPARATION
- TRANSFORMATION DETAILS IN PYTHON

First, read the raw data and inspect the data

```python
import pandas as pd

# Load the data
data = pd.read_csv("Retail_Transactions_Dataset.csv")

print('Columns:', ', '.join(list(data.columns)))
data.head()
```

✓ 3.3s

Output: next ->

DATA PREPARATION
- TRANSFORMATION DETAILS IN PYTHON

Columns: Transaction_ID, Date, Customer_Name, Product, Total_Items, Total_Cost, Payment_Method, City, Store_Type, Discount_Applied, Customer_Category, Season, Promotion

	Transaction_ID	Date	Customer_Name	Product	Total_Items	Total_Cost	Payment_Method	City	Store_Type	Discount_Applied	Customer_Category	Season	Promotion
0	1000000000	2022-01-21 06:27:29	Stacey Price	['Ketchup', 'Shaving Cream', 'Light Bulbs']	3	71.65	Mobile Payment	Los Angeles	Warehouse Club	True	Homemaker	Winter	NaN
1	1000000001	2023-03-01 13:01:21	Michelle Carlson	['Ice Cream', 'Milk', 'Olive Oil', 'Bread', 'P...	2	25.93	Cash	San Francisco	Specialty Store	True	Professional	Fall	BOGO (Buy One Get One)
2	1000000002	2024-03-21 15:37:04	Lisa Graves	['Spinach']	6	41.49	Credit Card	Houston	Department Store	True	Professional	Winter	NaN
3	1000000003	2020-10-31 09:59:47	Mrs. Patricia May	['Tissues', 'Mustard']	1	39.34	Mobile Payment	Chicago	Pharmacy	True	Homemaker	Spring	NaN
4	1000000004	2020-12-10 00:59:59	Susan Mitchell	['Dish Soap']	10	16.42	Debit Card	Houston	Specialty Store	False	Young Adult	Winter	Discount on Selected Items

The problem is Product column is list of products that stored in “object” data type, we need to separate the products to single product

DATA PREPARATION
- TRANSFORMATION DETAILS IN PYTHON

Process to transform the Product column to be available for usage
by read raw data to csv but add the converters parameter which selected the listed column.
Then, use “explode” method to divide the element of each list

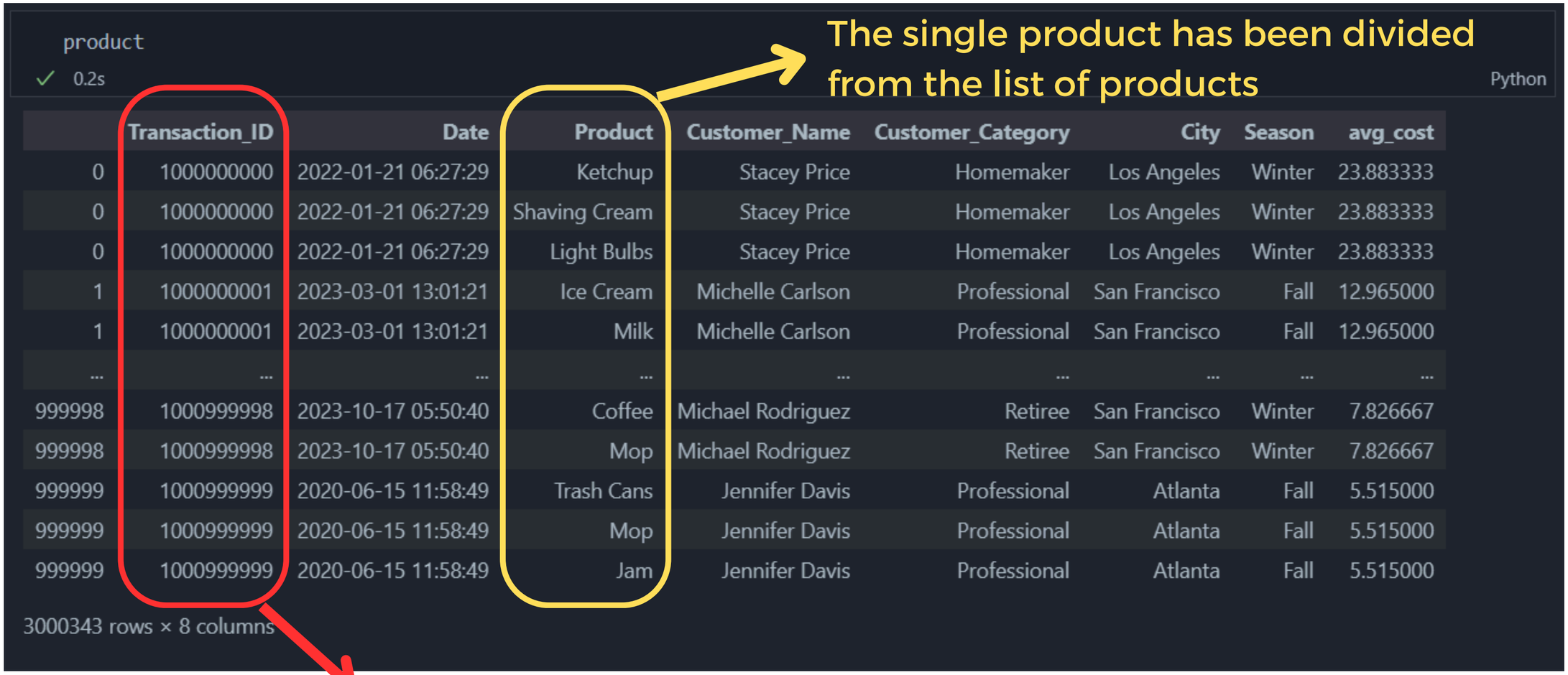
```python
from ast import literal_eval

# Read the raw data again but using ast.literal_eval to convert Product columns can be used as a list
data = pd.read_csv("Retail_Transactions_Dataset.csv", converters={"Product": literal_eval})

# Calculate average cost per product before explode the list of products
data['avg_cost'] = data['Total_Cost'] / data['Total_Items']
```

✓ 16.2s Python

```python
product = data.explode('Product') \
    [['Transaction_ID','Date','Product','Customer_Name','Customer_Category','City','Season','avg_cost']]
    # Select Product and other columns:
        # Transaction_ID as foreign key
        # other columns for futher analysis
```

✓ 2.3s Python

DATA PREPARATION
- TRANSFORMATION DETAILS IN PYTHON



	Transaction_ID	Date	Product	Customer_Name	Customer_Category	City	Season	avg_cost
0	1000000000	2022-01-21 06:27:29	Ketchup	Stacey Price	Homemaker	Los Angeles	Winter	23.883333
0	1000000000	2022-01-21 06:27:29	Shaving Cream	Stacey Price	Homemaker	Los Angeles	Winter	23.883333
0	1000000000	2022-01-21 06:27:29	Light Bulbs	Stacey Price	Homemaker	Los Angeles	Winter	23.883333
1	1000000001	2023-03-01 13:01:21	Ice Cream	Michelle Carlson	Professional	San Francisco	Fall	12.965000
1	1000000001	2023-03-01 13:01:21	Milk	Michelle Carlson	Professional	San Francisco	Fall	12.965000
...	...	...	...	...	...	...	...	...
999998	1000999998	2023-10-17 05:50:40	Coffee	Michael Rodriguez	Retiree	San Francisco	Winter	7.826667
999998	1000999998	2023-10-17 05:50:40	Mop	Michael Rodriguez	Retiree	San Francisco	Winter	7.826667
999999	1000999999	2020-06-15 11:58:49	Trash Cans	Jennifer Davis	Professional	Atlanta	Fall	5.515000
999999	1000999999	2020-06-15 11:58:49	Mop	Jennifer Davis	Professional	Atlanta	Fall	5.515000
999999	1000999999	2020-06-15 11:58:49	Jam	Jennifer Davis	Professional	Atlanta	Fall	5.515000

3000343 rows × 8 columns

Transaction_ID is the foreign key that related to Transaction_ID in the original table

DATA PREPARATION
- TRANSFORMATION DETAILS IN PYTHON

Save the parquet files

```python
# Load to parque format
data.to_parquet("Retail_Transactions_Dataset_pq.parquet")
product.to_parquet("Transactions_by_products.parquet")
```

The result of converting the data from .csv to .parquet:
169 Mb --> 91 Mb (total of 2 parquet files)

Then, upload these parquet files on BigQuery to be the data source on Looker Studio

DASHBOARD CONTENTS

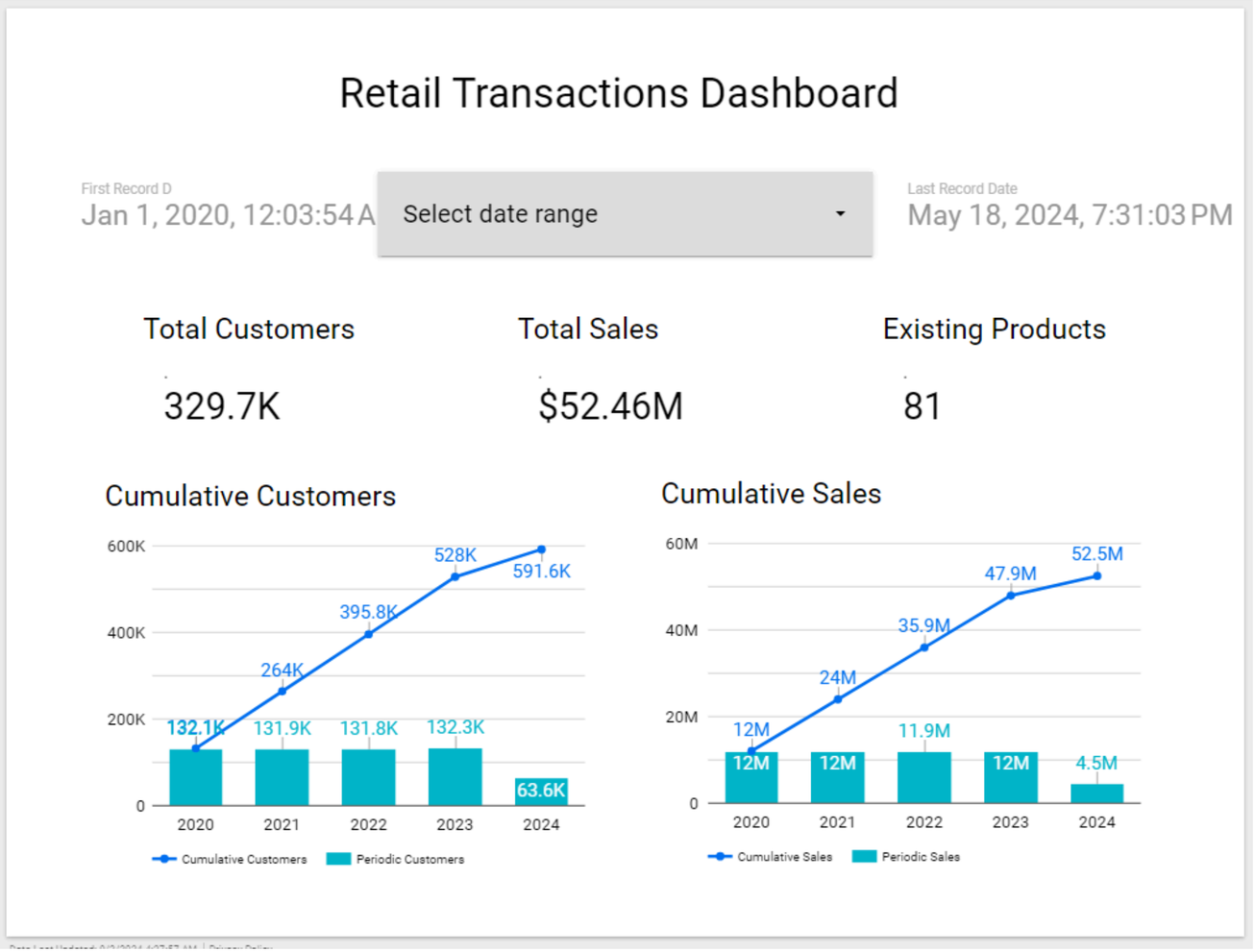
Page 1 : Overview

Page 2 : Sales performance

Page 3 : Product

Page 4 : Region

Page 5 : Customer behavior



DASHBOARD CONTENTS

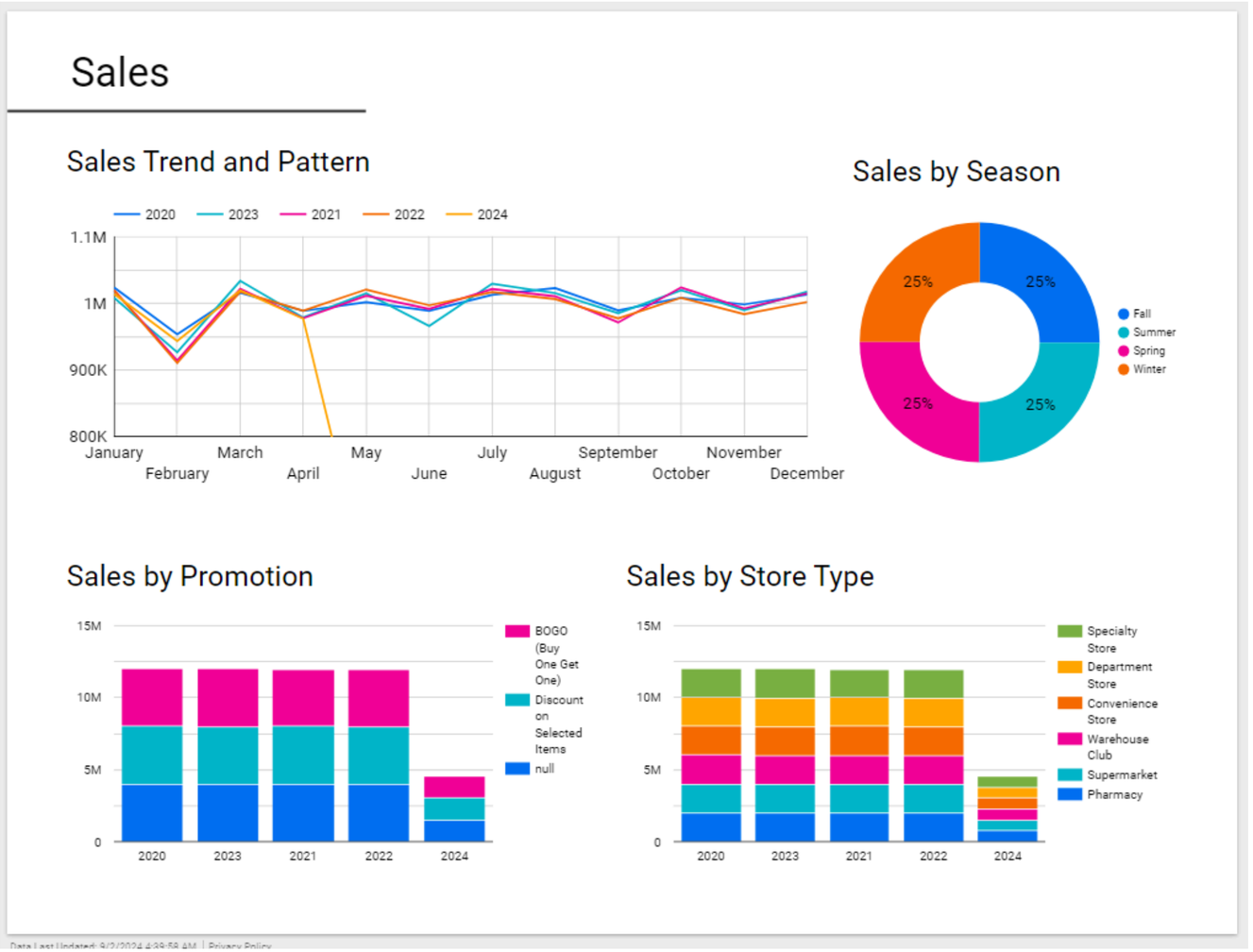
Page 1 : Overview

Page 2 : Sales performance

Page 3 : Product

Page 4 : Region

Page 5 : Customer behavior



DASHBOARD CONTENTS

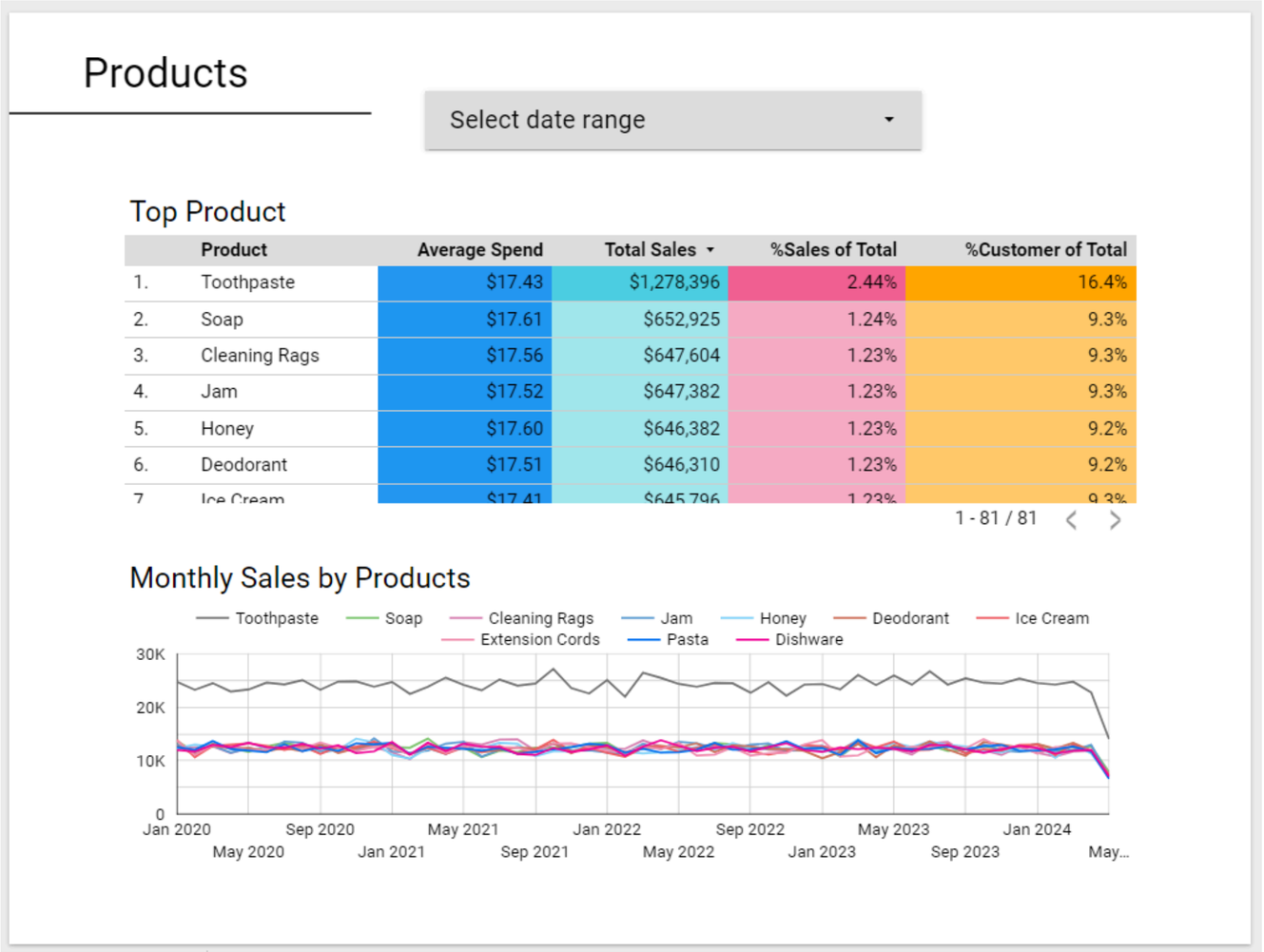
Page 1 : Overview

Page 2 : Sales performance

Page 3 : Product

Page 4 : Region

Page 5 : Customer behavior



DASHBOARD CONTENTS

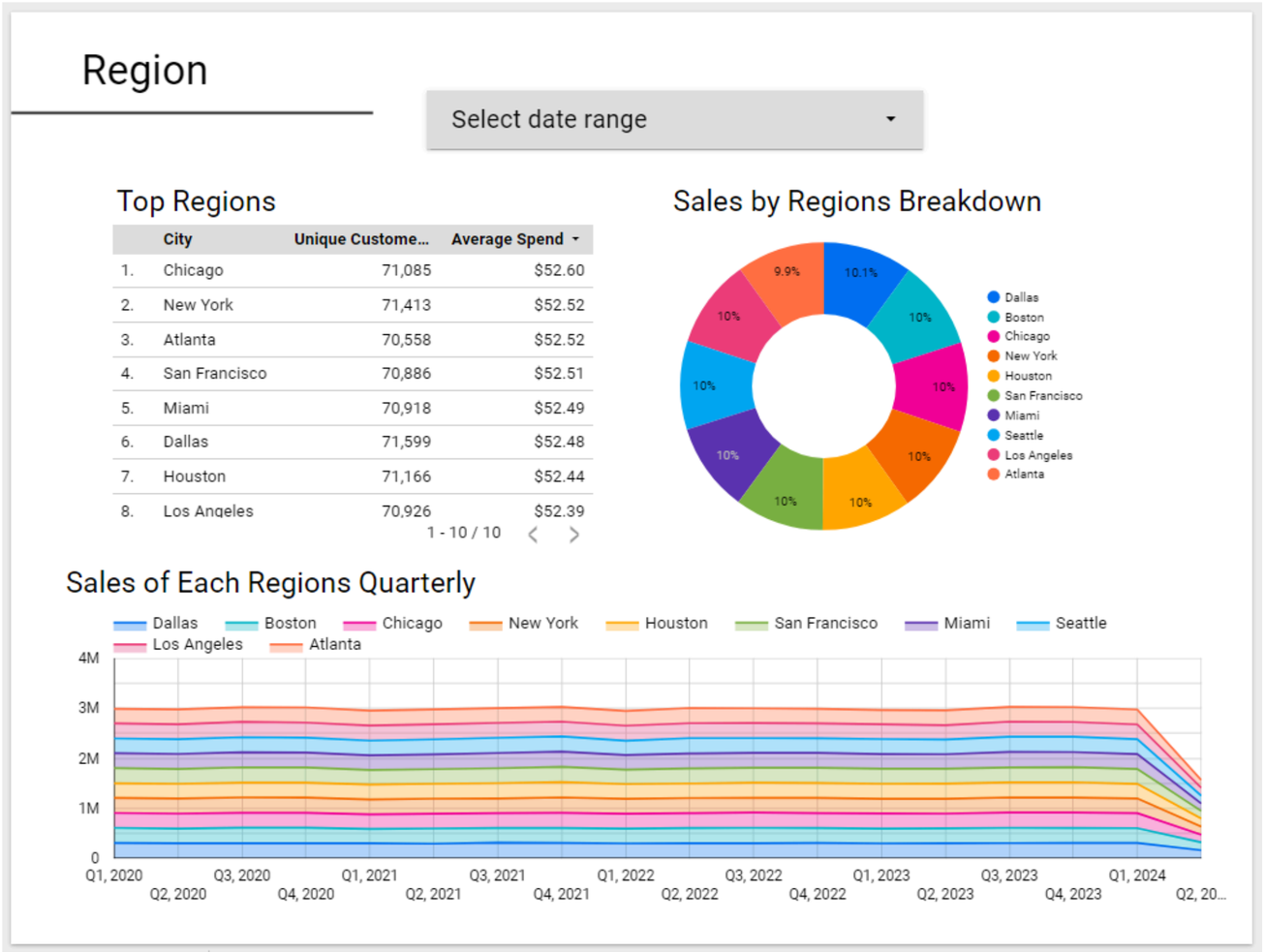
Page 1 : Overview

Page 2 : Sales performance

Page 3 : Product

Page 4 : Region

Page 5 : Customer behavior



DASHBOARD CONTENTS

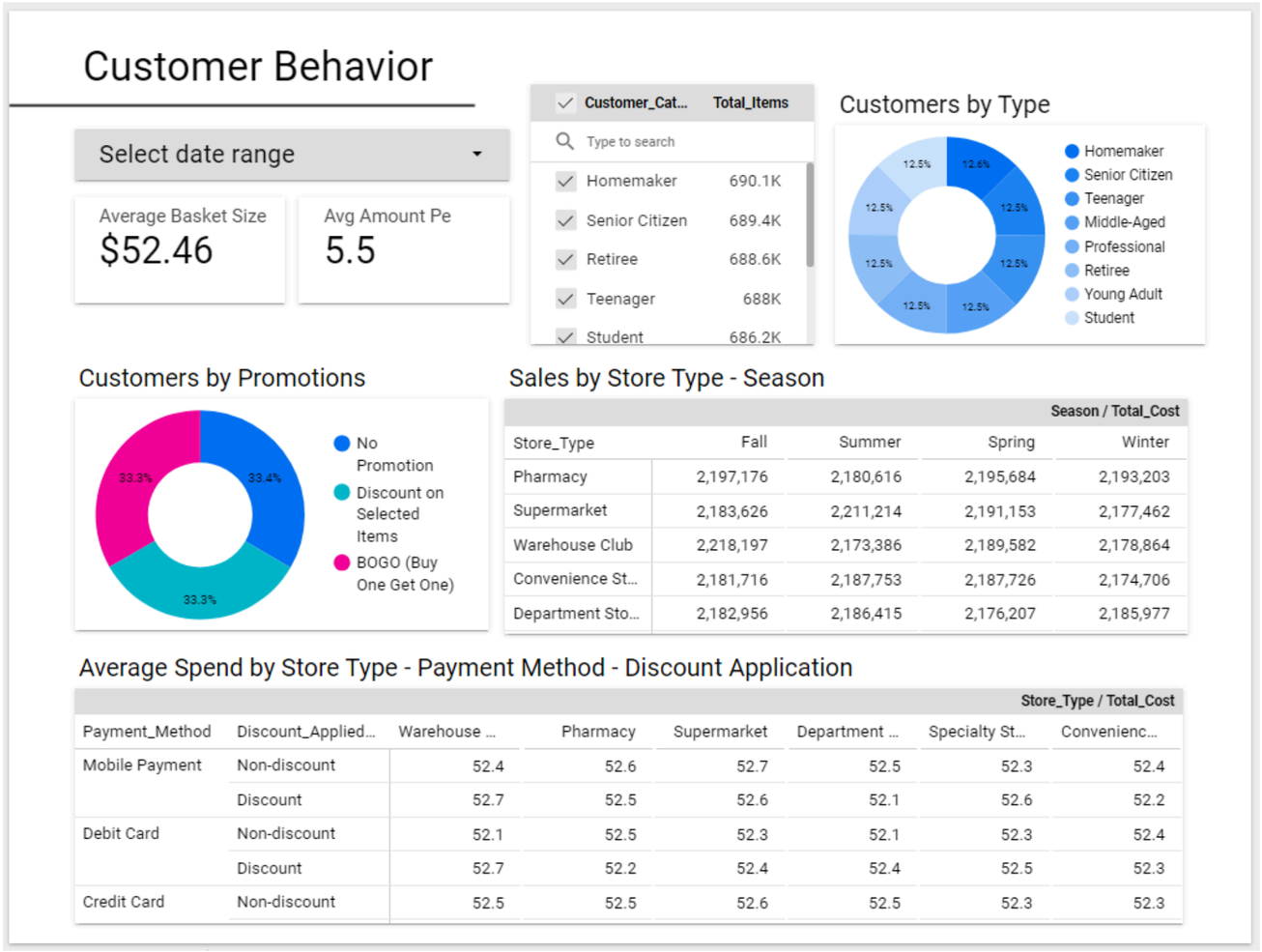
Page 1 : Overview

Page 2 : Sales performance

Page 3 : Product

Page 4 : Region

Page 5 : Customer behavior



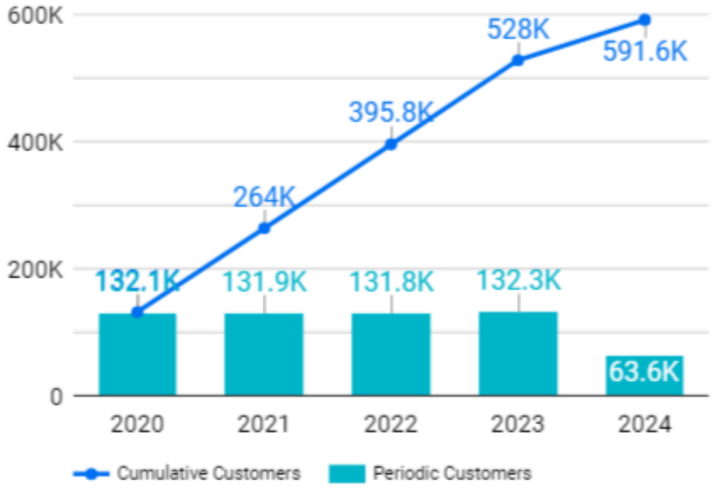
EXECUTIVE SUMMARY FROM THE DASHBOARD

next ->

OVERVIEW (Dashboard page 1)

Total Customers	Total Sales	Existing Products
329.7K	$52.46M	81

Cumulative Customers

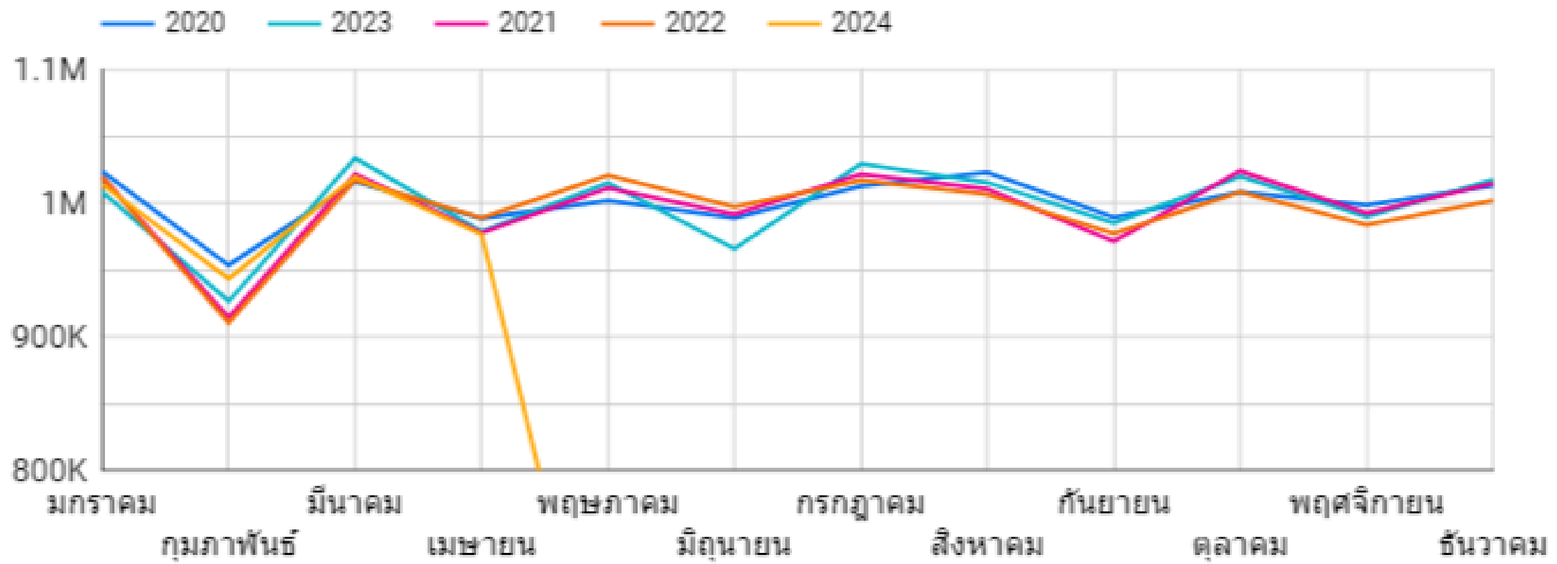


Cumulative Sales

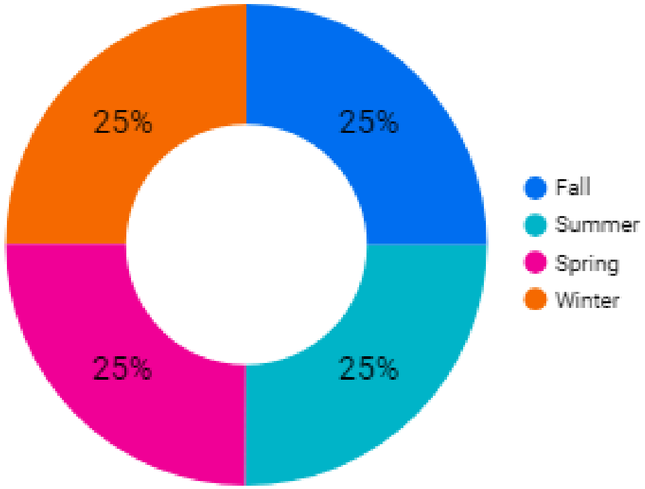
- Time range of data is started at January 2020 and ended at 18 May 2024.
- The all-time statistics shows that the business gained more than 50M $ total sales, 300k unique customers, and has a lot of products (81) for sale.
- On two charts below, we can see that from 2020 to 2023, sales and amount of customers are at the same level or having no significant growth. This is the sign of the maturity of the business or facing some limitations.
- In 2024, Sales and customers seem to be the same as before.

SALES PERFORMANCE BREAKDOWN (Dashboard page 2)

Sales Trend and Pattern



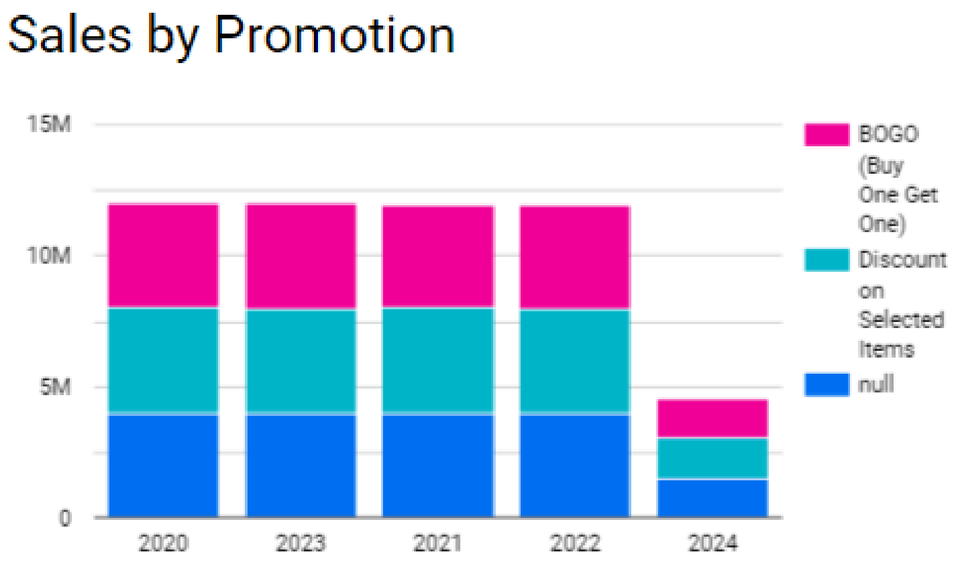
Sales by Season

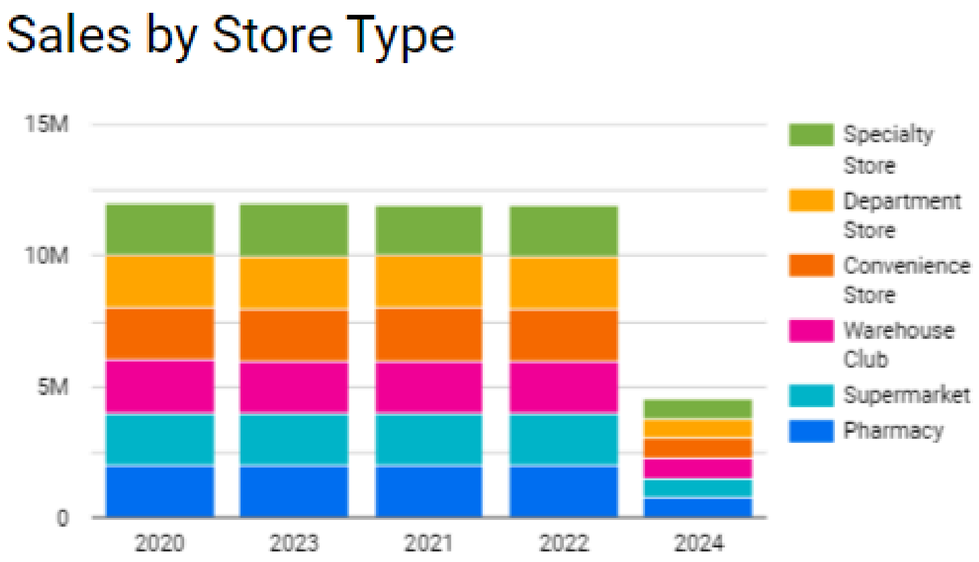


- From the plot of Sales for each month by years, wecansee that sales is around 100k each month without up or down trend.
- The sales are > 1M $ at months that have 31 days and < 1M $ at months that have 30 days and droped on every Febuary. **Sales are tend to be equal daily**
- In 2024, the first 4 months are repeated the pattern (not included may because it is only half of the month). Is can be projected till the end of 2024 that sales tend to repeat the pattern which means **the total sales of 2024 will be around the same level**

- When devide sales data as seasonal, there are no significant difference between each season

SALES PERFORMANCE BREAKDOWN (Dashboard page 2)





- The Breakdown of sales on promotion show indifferent of sales between Buy1get1 promotion, Discounted promotions, or buy without promotions
- But if we consider only on buying with promotions or not, it means that **2/3 of sales came from promotions**

- The Breakdown of sales on type of stores show indifferent of sales between store

PRODUCT BREAKDOWN (Dashboard page 3)

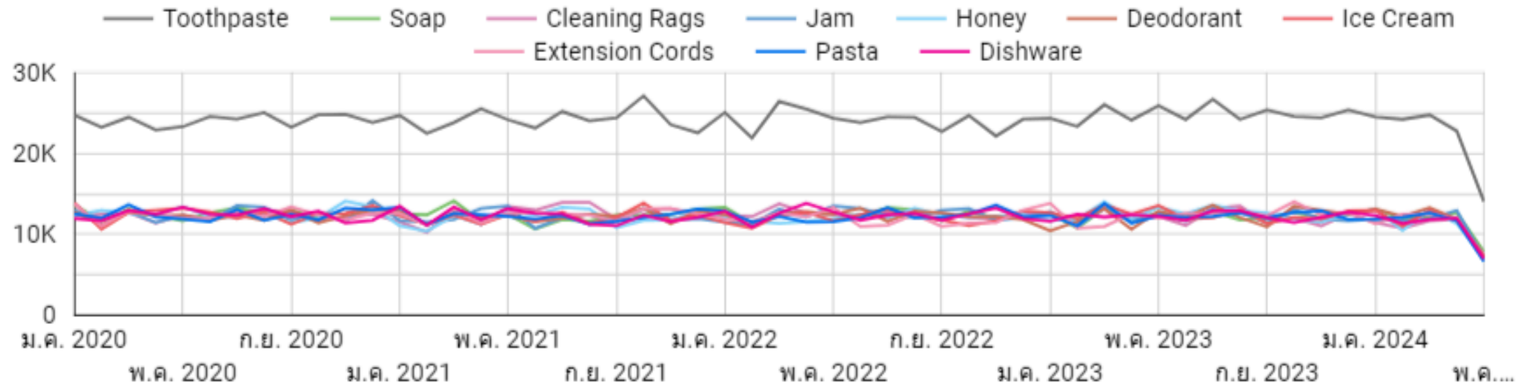
Top Product

	Product	Average Spend	Total Sales	%Sales of Total	%Customer of Total
1.	Toothpaste	$17.43	$1,278,396	2.44%	16.4%
2.	Soap	$17.61	$652,925	1.24%	9.3%
3.	Cleaning Rags	$17.56	$647,604	1.23%	9.3%
4.	Jam	$17.52	$647,382	1.23%	9.3%
5.	Honey	$17.60	$646,382	1.23%	9.2%
6.	Deodorant	$17.51	$646,310	1.23%	9.2%
7.	Ice Cream	$17.41	$645,796	1.23%	9.3%

1 - 81 / 81

- Every product has the same level of avarage purchase
- **Toothpaste** is different from the other 80 products. It contributed 16.4% of customer who ever buy it and sales 2.4% of total which is more than the other 2 times

Monthly Sales by Products



- Every product sold at the same level of itself every months. No products that have seasonal or periodical patterns.

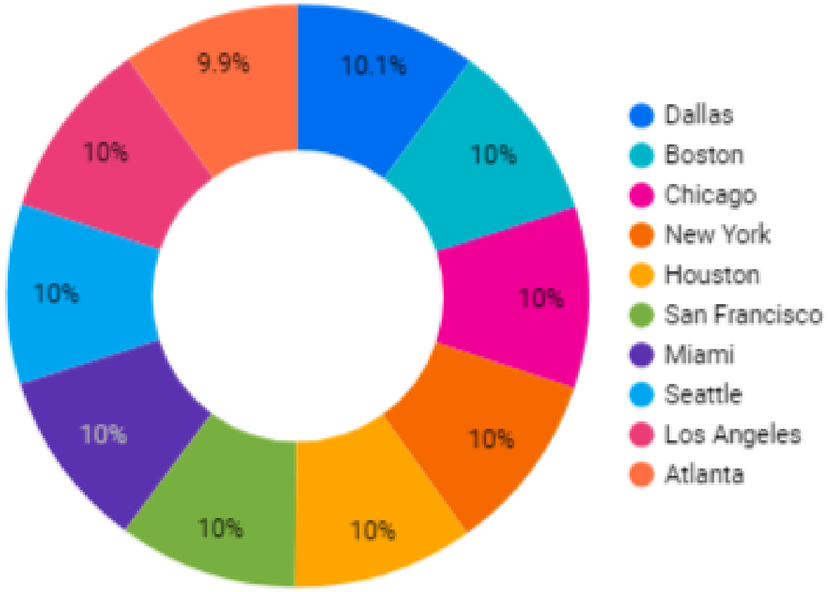
REGION BREAKDOWN (Dashboard page 4)

Top Regions

	City	Unique Custome...	Average Spend ▾
1.	Chicago	71,085	$52.60
2.	New York	71,413	$52.52
3.	Atlanta	70,558	$52.52
4.	San Francisco	70,886	$52.51
5.	Miami	70,918	$52.49
6.	Dallas	71,599	$52.48
7.	Houston	71,166	$52.44
8.	Los Angeles	70,926	$52.39

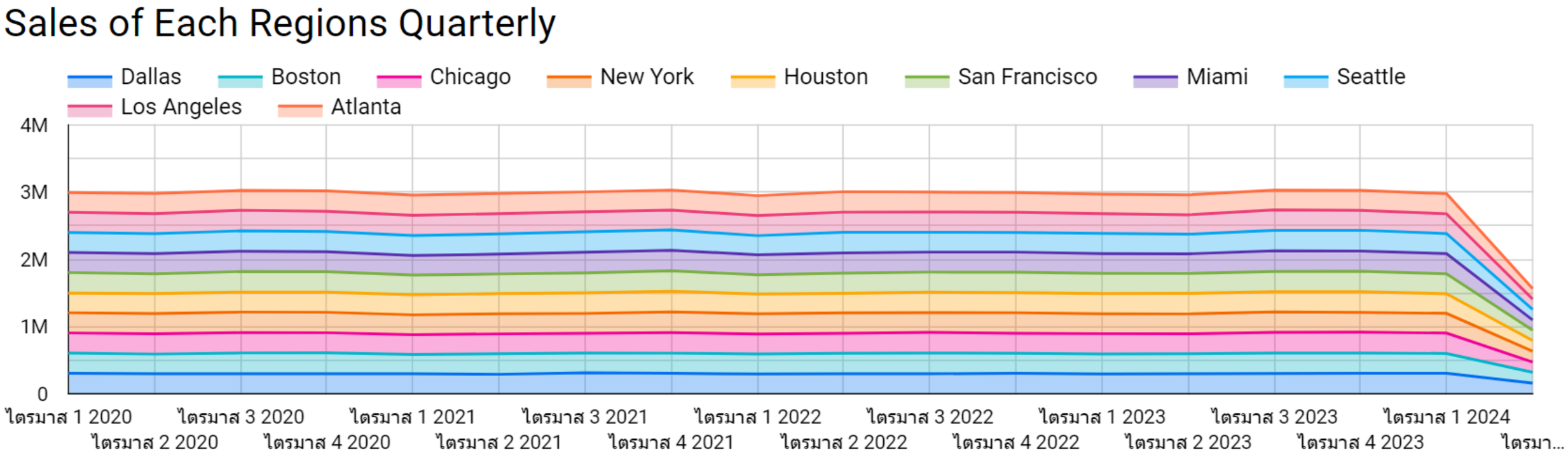
1 - 10 / 10

Sales by Regions Breakdown



- The are total 10 cities that the business operates and all cities have the same level of unique customers, avaverge spending, and sales

REGION BREAKDOWN (WITH TIME) (Dashboard page 4)

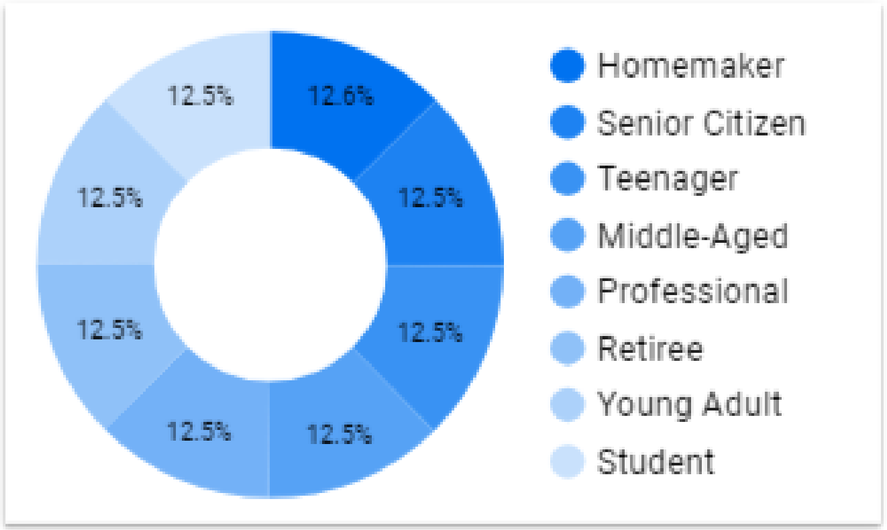


- And there also no pattern of different sales on the time-breakdown

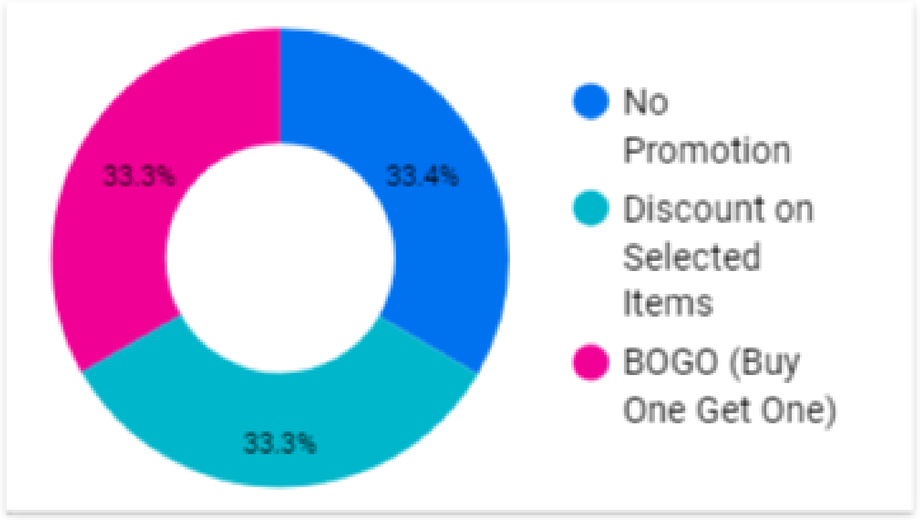
CUSTOMER BEHAVIOR (Dashboard page 5)

Average Basket Size	Avg Amount Pe
$52.46	5.5

Customers by Type



Customers by Promotions



CUSTOMER BEHAVIOR (Dashboard page 5)

- The unit breakdown analysis to find the customer insights is not found the specific form of behavior that significantly affect sales to different
- **So that every customer segments tend to have the same prefrerence, desicion making, and how they response the information**

Sales by Store Type - Season

	Season / Total_Cost			
Store_Type	Fall	Summer	Spring	Winter
Supermarket	2,183,626	2,211,214	2,191,153	2,177,462
Warehouse Club	2,218,197	2,173,386	2,189,582	2,178,864
Convenience St...	2,181,716	2,187,753	2,187,726	2,174,706
Department Sto...	2,182,956	2,186,415	2,176,207	2,185,977
Specialty Store	2,173,242	2,177,292	2,172,886	2,178,180

Average Spend by Store Type - Payment Method - Discount Application

		Store_Type / Total_Cost					
Payment_Method	Discount_Applied...	Warehouse ...	Pharmacy	Supermarket	Department ...	Specialty St...	Convenienc...
Mobile Payment	Non-discount	52.4	52.6	52.7	52.5	52.3	52.4
	Discount	52.7	52.5	52.6	52.1	52.6	52.2
Debit Card	Non-discount	52.1	52.5	52.3	52.1	52.3	52.4
	Discount	52.7	52.2	52.4	52.4	52.5	52.3
Credit Card	Non-discount	52.5	52.5	52.6	52.5	52.3	52.3

SUMMARY FROM THE DATA

Customer

- **Any segment of customers are the same** amount, basket size (average spending 52.5 $ per purchase), and equally distributed activity or behavior

Product

- **Toothpaste is the only one outstanding product** which contributed around 2 times compared to all other products and contributed 16.4% of customer
- Other 80 products have approximately equal sales and customers

Promotions

- **2/3 of sales came from using promotions.** Promotions contain with discount and buy1get1

Time, Season, Store type, Payment method, Region

- **Equal customers and sales contribution**

RECOMMENDATIONS

Focus on the outstanding products (toothpaste)

- Toothpaste have potential to be developed for new product lines or different quality and pricings

Push the existing and new markets with promotion

- Because 2/3 of sales came from promotions
- Because of customers behavior that tend to be the same at all, the promotion effort doesn't have to differenciate to specific segment of customers

Region expansion to other major cities (cities which aquire large population)

- The current and the past of the business is considerably stable sales and customers, and still stable in units of analysis (eg. season, channel, payment method)
- This retail business is existing in major cities for major states of USA, but just 10 cities. There are many opportunitie to open new market in other major cities in USA such as Philadelphia, Phoenix

THE END