Exploratory Data Analysis

- StudentID: 22000655

- Name: Jeongeun Jeon

- 1st Major: AI Convergence & Entrepreneurship

- 2nd Major: Data science

My idea suggests grouping items with similar sales trends together when placing them on display in a large mart.

1. Data overview

S_data

Sample Size: `5000000`
Number of Variables: `14`
The period of this data: `from January 1, 2010, to September 10, 2020`

Data Types:

- **Region(str)**: Represents the region where the data is recorded.
- **Country(str)**: Indicates the country belonging to the region.
- **Item Type(str)**: Specifies the type of product being sold.
- **Sales Channel(str)**: Indicates the channel through which the product is sold.
- **Order Priority(str)**: Represents the priority of the order.
- **Order Date(str)**: Represents the date when the order was placed.
- **Order ID(int64)**: Identifies the order.
- **Ship Date(str)**: Represents the date when the product was shipped.
- **Units Sold(int64)**: Indicates the number of units sold.
- **Unit Price(float64)**: Represents the price per unit of the product.
- **Unit Cost(float64)**: Represents the cost per unit of the product.
- **Total Revenue(float64)**: Represents the total revenue generated from the order.
- **Total Cost(float64)**: Represents the total cost incurred for the order.
- **Total Profit(float64)**: Represents the total profit generated from the order.

2. Univariate analysis

2.1 Variable 1

This graph illustrates the total profit of each item type annually from 2010 to 2020.

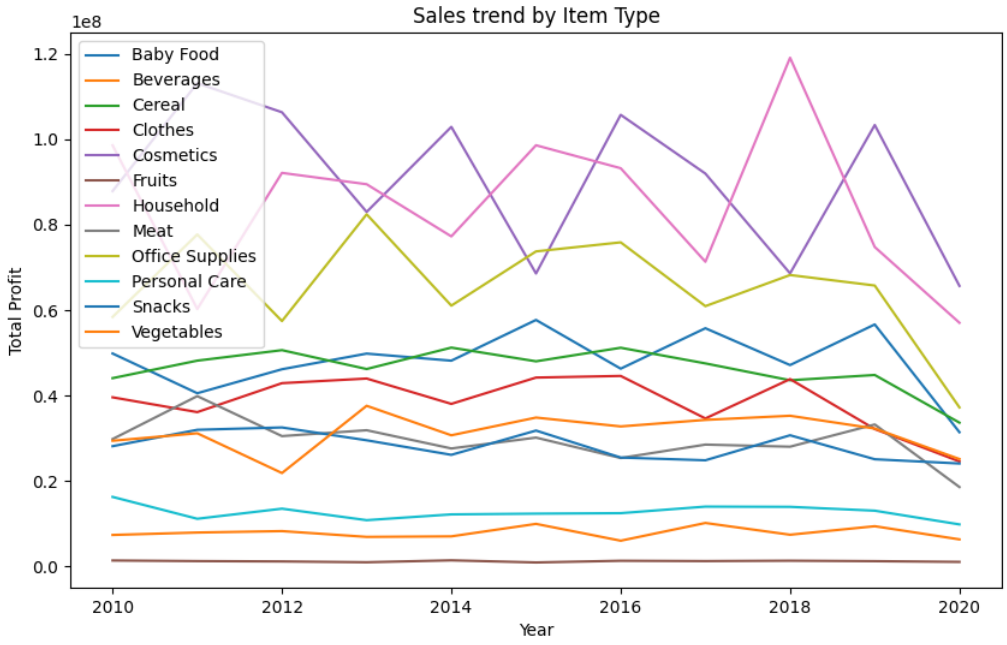

Figure 1. Sales trend by Item Type in Korea

2.2 (ex) Variable 2

This graph displays the average total profit by Item Type from 2010 to 2020.

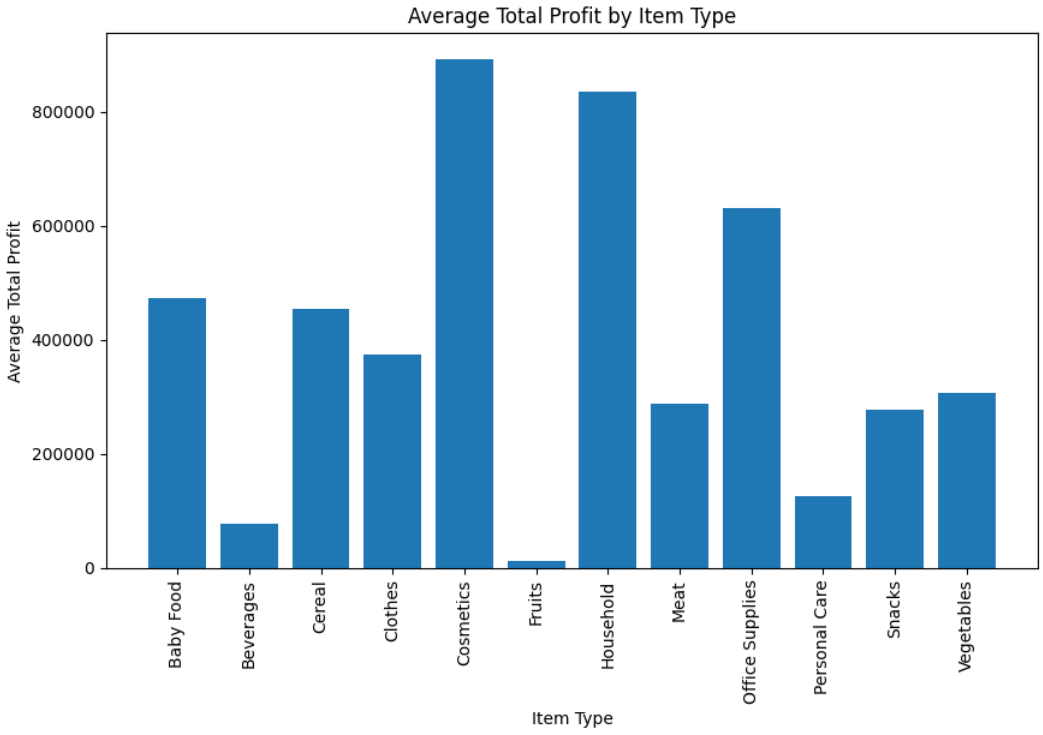

Figure 2. Average Total Profit by Item Type

2.3 Variable 3

This table represents the sales volume and total profit of each product. The table below indicates that while the number of sales is similar across items, there is a significant difference in the amounts.

Item Type	Units Sold	Total Profit
Baby Food	5,523,479	529,480,700
Beverages	5,551,424	86,935,300
Cereal	5,746,503	509,082,700
Clothes	5,782,684	424,680,300
Cosmetics	5,733,725	996,922,800
Fruits	5,580,799	13,449,700
Household	5,621,254	931,610,400
Meat	5,657,637	323,616,800
Office Supplies	5,692,848	718,722,100
Personal Care	5,574,489	139,696,700
Snacks	5,629,228	310,395,600
Vegetables	5,471,144	345,393,300

3. Multivariate analysis

3.1 Correlation

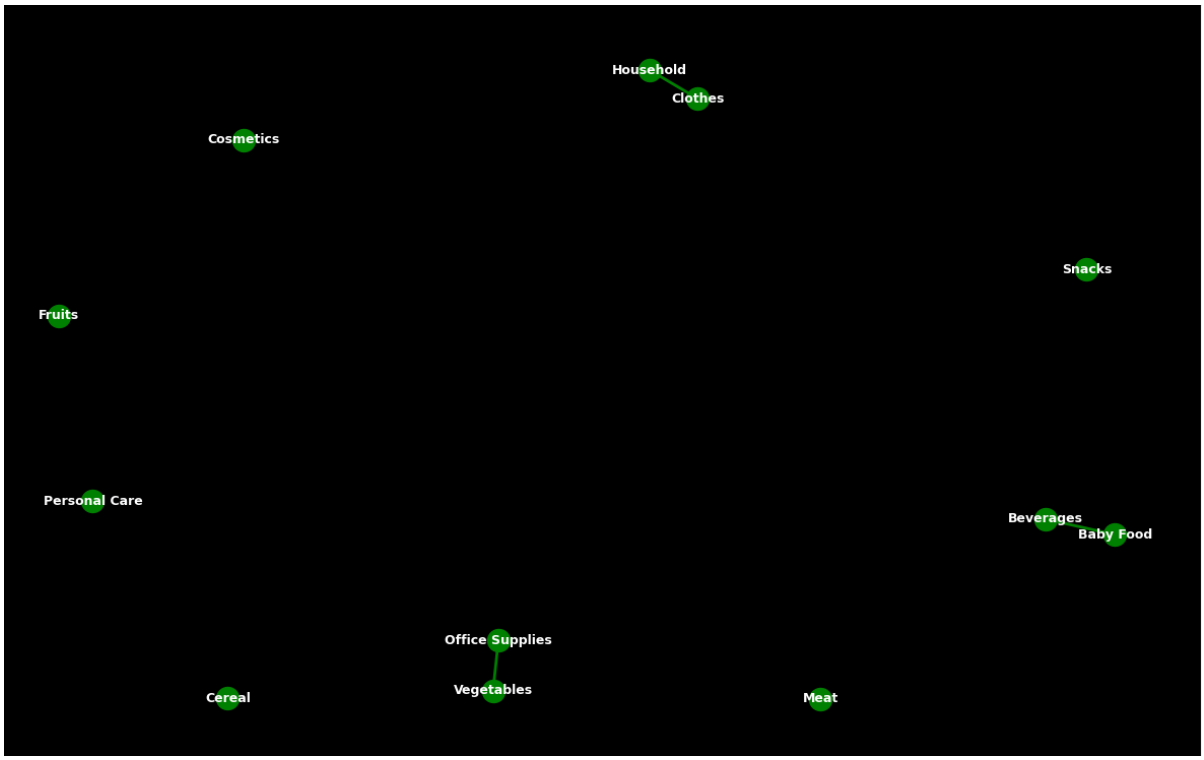

Figure 3. Offline Item Type Correlation

The following pairs are deemed to have a high level of correlation, as their correlation coefficients are all above 0.7:

Baby Food - Beverages = 0.7134679880900501
Clothes - Household = 0.8094924402667149
Office Supplies - Vegetables = 0.717379932023296

4. Suggestion

I would like to suggest arranging the following items together when positioning them in a large mart:

1. Baby Food - Beverages
   These two items are related to consumers' meals. When considering the target audience as infants, there is a high likelihood that they would enjoy beverages. Therefore, it supports the hypothesis that households purchasing baby food would often also purchase beverages.

2. Clothes - Household
   These two items are closely related to consumers' daily lives. Additionally, there are items that need to be stocked seasonally. Therefore, consumers purchasing household items often have a high likelihood of also purchasing personal hygiene products and clothing.

3. Office Supplies - Vegetables
   The two items belong to completely different categories. However, they need to be regularly supplied as consumables. Since it's unlikely that the same individual would purchase them simultaneously, additional analysis and judgment are needed to determine whether they should be placed together.

Displaying items with similar sales trends together can reduce the path of shoppers as their purchase timing would likely be similar. It can also potentially increase sales and allow for coordinated sale events.