

Report on

# Sales Analysis on supermarket grocery dataset

CodeClause Data Science Internship – September 2022

**Author name: Akash Mahajan**

## **Objectives:**

### **1. Investigate the correlation of various quantitative parameters in dataset**

This activity reduces the number of columns being analyzed to avoid extra data being taken for testing and analysis. The outcome of this activity is the list of correlated columns or data.

### **2. To list out the popularity of products by amount of total sales**

Each sales earns an overhead profit. When the sales amount and profit is summed up, we get an amount which represents total sales. In this activity, sub-category level products are listed out by their total sales to show top 5 popular products in inventory

### **3. To investigate the cities where top 5 products are sold**

In the last activity, products by their total sales are listed. In this activity, the cities where the top 5 selling products are sold the most are investigated.

### **4. Investigate demand of each top product over the years**

The demand of each top product by its total sales amount is investigated. This activity shows the increasing or decreasing demand of a product from 2015 to 2018.

### **5. Providing suggestions to improve sales in the store based on past investigations**

This activity sums up all the investigations and provides a list of practical approaches for sales department to produce more sales.

**Program files:** A separate Python program file is created for each new objective. The program files have the objective commented on top to identify the purpose of program file.

## Pre-analysis activity 1: Dataset report

The dataset used for experiments is totally fictional and is NOT owned by and collected from any original establishment of business. The dataset represents the region-wise sales of grocery products in Tamil Nādu state, India. An individual sale is separated by a unique order ID where each sale is categorized in major category and later in subcategory. The dataset has quantified columns – sales, discount, and profit. Sales column represents the amount of sales, discount column represents the amount of discount, and profit is the overhead amount earned as a profit per sale. An associated date of sale is also recorded.

### Dataset cleaning policy:

- Order ID and customer name are not considered for exploratory data analysis as well as for further procedures because it has no apparent impact on sales. Hence, removed from dataset.
- The whole data is collected for only single State – Tamil Nādu. Hence, the column ‘state’ is removed because all rows contain only single variation of state names.
- The quantified columns like sales amount, discount, and profit amount are only considered till 2 decimal places. No further precision is considered.

Following is the systematic report on the grocery sales dataset.

## Pre-analysis activity 2: Dataset Exploration

Similarly, you can run `basic_information_utility.py` to print following information on console

### Basic information of the dataset

| Total records in dataset: 9994 |               |               |         |
|--------------------------------|---------------|---------------|---------|
| Column name                    | Minimum value | Maximum value | Average |
| Sales                          | 500.00        | 2500.00       | 1496.59 |
| Discount                       | 0.10          | 0.35          | 0.23    |
| Profit                         | 25.25         | 1120.95       | 374.93  |

### Meta information of the dataset

#### Unique subcategories of sales

| Category          | Subcategory products |                   |                   |              |
|-------------------|----------------------|-------------------|-------------------|--------------|
|                   | 1                    | 2                 | 3                 | 4            |
| Eggs, Meat & Fish | Eggs                 | Mutton            | Chicken           | Fish         |
| Snacks            | Chocolate            | Noodles           | Cookies           |              |
| Beverages         | Health Drinks        | Soft Drinks       |                   |              |
| Food Grains       | Organic Staples      | Atta & Flour      | Dals & Pulses     | Rice         |
| Oil & Masala      | Masala               | Edible Oil & Ghee | Spices            |              |
| Bakery            | Breads & Buns        | Biscuits          | Cakes             |              |
| Fruits & Veggies  | Fresh Vegetable      | Organic Fruits    | Organic Vegetable | Fresh Fruits |

### Cities where products are sold

There are 24 unique cities where the products are sold in outlets or through orders

|             |             |             |        |                |            |
|-------------|-------------|-------------|--------|----------------|------------|
| Nagercoil   | Tirunelveli | Ooty        | Theni  | Ramanadhapuram | Perambalur |
| Bodi        | Madurai     | Chennai     | Cumbum | Virudhunagar   | Tenkasi    |
| Kanyakumari | Salem       | Krishnagiri | Trichy | Vellore        | Namakkal   |

|            |       |            |            |            |          |
|------------|-------|------------|------------|------------|----------|
| Coimbatore | Karur | Viluppuram | Dharmapuri | Pudukottai | Dindigul |
|------------|-------|------------|------------|------------|----------|

### Other information

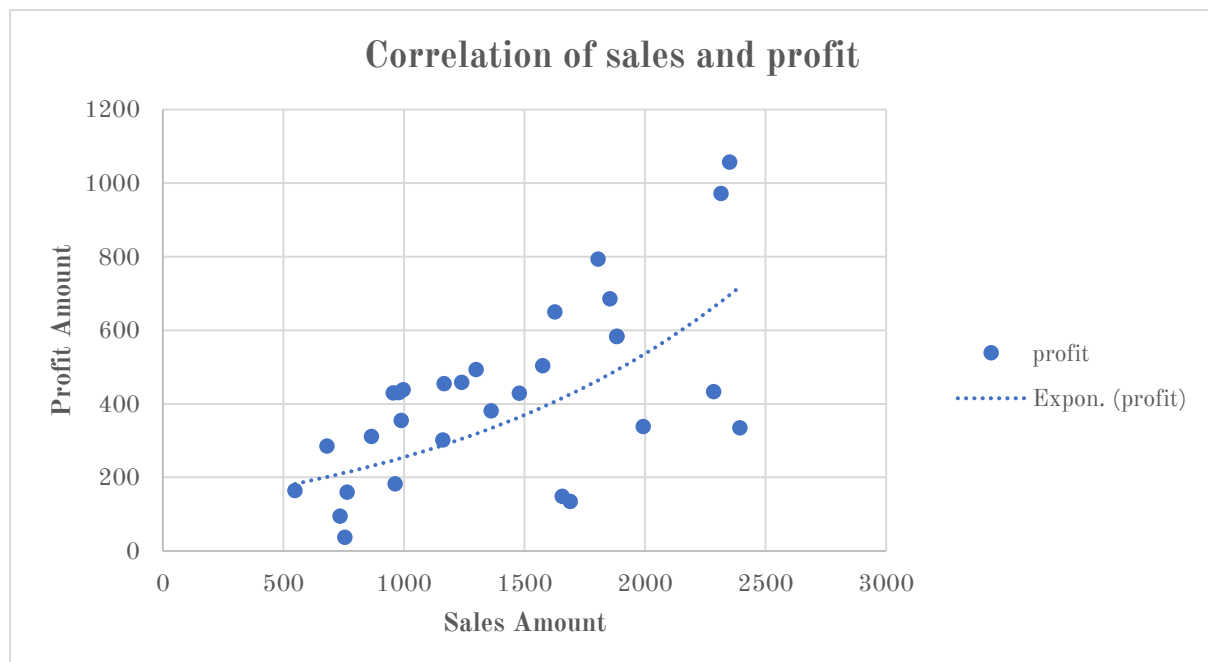
|   |                          |
|---|--------------------------|
| <b>Total number of records:</b>                               | 9994 (No Null values)    |
| <b>The range of dates between which all sales took place:</b> | 03/01/2015 to 30/12/2018 |
| <b>The maximum price of sales:</b>                            | 2500.00                  |
| <b>The minimum price of sales:</b>                            | 500.00                   |
| <b>The minimum discount:</b>                                  | 10%                      |
| <b>The maximum discount:</b>                                  | 35%                      |
| <b>The minimum profit:</b>                                    | 25.00                    |
| <b>The maximum profit:</b>                                    | 1120.00                  |

## Objective 1: Investigate the correlation of various quantitative parameters in dataset

The quantitative fields in the dataset are sales amount, discount, and profit. Following is the brief on experiments to investigate correlation among these variables

| Parameter 1 | Parameter 2 | Correlation Coefficient |
|-------------|-------------|-------------------------|
| Sales       | Discount    | -0.006                  |
| Sales       | Profit      | 0.605                   |
| Profit      | Discount    | 0.000                   |

This experiment highlights the stronger relation between sales amount and profit. i.e., if one increases, other increases and vice versa. This can also be seen in the following graph. The first 10 records are visualized.



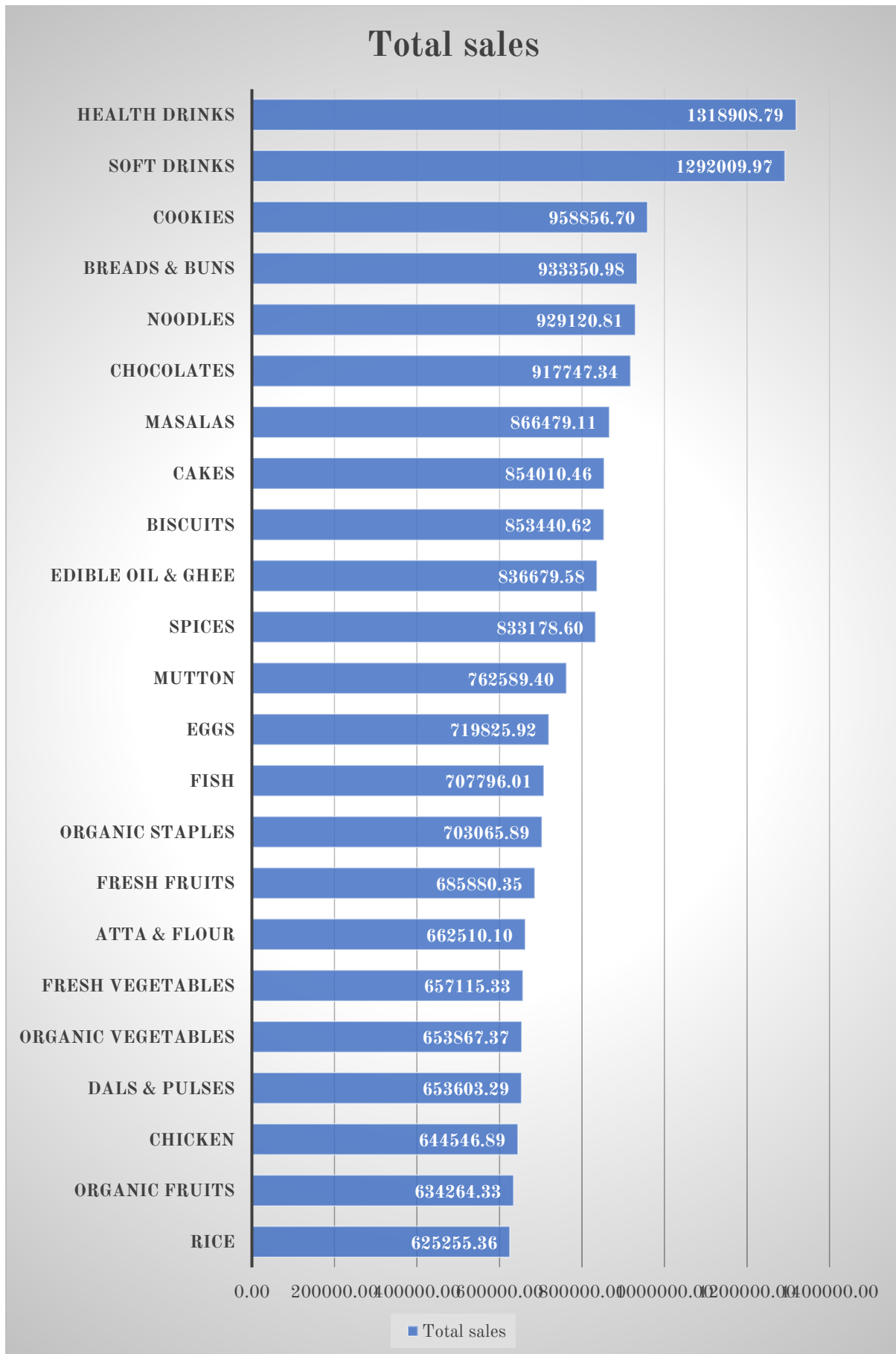
### Remarks:

This experiment aimed at finding the correlation of quantitative parameters in dataset. It is found that profit and sales amount are highly correlated variables. However, this information is vague. Proceeding experiments investigate this relation on product level.

## Objective 2: To list out the popularity of products by amount of total sales

To find out the popular products, the sales and profit columns are summed up into new column total. Based on total amount, the products are listed.

| Product name       | Total sales |
|--------------------|-------------|
| Health Drinks      | 1318908.79  |
| Soft Drinks        | 1292009.97  |
| Cookies            | 958856.70   |
| Breads & Buns      | 933350.98   |
| Noodles            | 929120.81   |
| Chocolates         | 917747.34   |
| Masalas            | 866479.11   |
| Cakes              | 854010.46   |
| Biscuits           | 853440.62   |
| Edible Oil & Ghee  | 836679.58   |
| Spices             | 833178.60   |
| Mutton             | 762589.40   |
| Eggs               | 719825.92   |
| Fish               | 707796.01   |
| Organic Staples    | 703065.89   |
| Fresh Fruits       | 685880.35   |
| Atta & Flour       | 662510.10   |
| Fresh Vegetables   | 657115.33   |
| Organic Vegetables | 653867.37   |
| Dals & Pulses      | 653603.29   |
| Chicken            | 644546.89   |
| Organic Fruits     | 634264.33   |
| Rice               | 625255.36   |





### Objective 3: To investigate the cities where top 5 products are sold

This experiment takes the top 5 selling products from the former experiment and finds out their sales by city. In this way, a vendor of these product can target the cities to promote their products and services associated with the top 5 selling products.

| City           | Health Drinks | Soft Drinks | Cookies  | Breads & Buns | Noodles  |
|----------------|---------------|-------------|----------|---------------|----------|
| Nagercoil      | 9052.86       | 8047.87     | 6942.33  | 6409.05       | 7641.20  |
| Tirunelveli    | 12425.93      | 13785.69    | 11054.44 | 7593.75       | 8079.60  |
| Ooty           | 12316.03      | 9017.22     | 7650.61  | 7986.45       | 6660.11  |
| Theni          | 6806.67       | 12853.38    | 6699.46  | 8632.15       | 7785.09  |
| Ramanadhapuram | 14286.74      | 11422.75    | 10489.00 | 8478.50       | 4682.90  |
| Perambalur     | 10375.79      | 16840.46    | 9144.91  | 10042.75      | 11006.40 |
| Bodi           | 10897.24      | 11083.22    | 8858.67  | 9928.68       | 8951.75  |
| Madurai        | 13501.33      | 14447.46    | 4514.02  | 8170.41       | 7525.17  |
| Chennai        | 14951.10      | 11656.28    | 9790.92  | 4016.45       | 9504.04  |
| Cumbum         | 10155.01      | 10639.12    | 8939.05  | 10257.02      | 6746.33  |
| Virudhunagar   | 11982.68      | 5674.05     | 8663.60  | 8433.34       | 11340.09 |
| Tenkasi        | 14226.99      | 9869.76     | 12069.20 | 8973.99       | 5459.10  |
| Kanyakumari    | 11803.91      | 7942.22     | 9001.74  | 6067.34       | 10310.63 |
| Salem          | 9610.01       | 9938.96     | 8597.31  | 8942.52       | 9789.98  |
| Krishnagiri    | 9211.69       | 9866.78     | 5895.91  | 9720.84       | 10138.40 |
| Trichy         | 8397.26       | 11083.70    | 4236.12  | 5463.34       | 4259.75  |
| Vellore        | 16319.80      | 10713.11    | 6713.61  | 7067.98       | 7783.40  |
| Namakkal       | 8518.20       | 13565.66    | 4000.28  | 9426.85       | 7570.12  |
| Coimbatore     | 13520.79      | 11967.26    | 5482.31  | 7605.72       | 7557.67  |
| Karur          | 8149.46       | 11702.32    | 9835.96  | 11536.08      | 8024.35  |

|                   |          |         |          |         |          |
|-------------------|----------|---------|----------|---------|----------|
| <b>Viluppuram</b> | 12840.60 | 9659.68 | 4499.38  | 6235.49 | 8001.62  |
| <b>Dharmapuri</b> | 7792.49  | 6927.34 | 11998.55 | 4989.35 | 6443.00  |
| <b>Pudukottai</b> | 11163.53 | 9858.27 | 7386.60  | 9567.58 | 11253.30 |
| <b>Dindigul</b>   | 9163.68  | 9573.41 | 8179.72  | 5219.35 | 7171.81  |

### Category and their top 5 selling cities

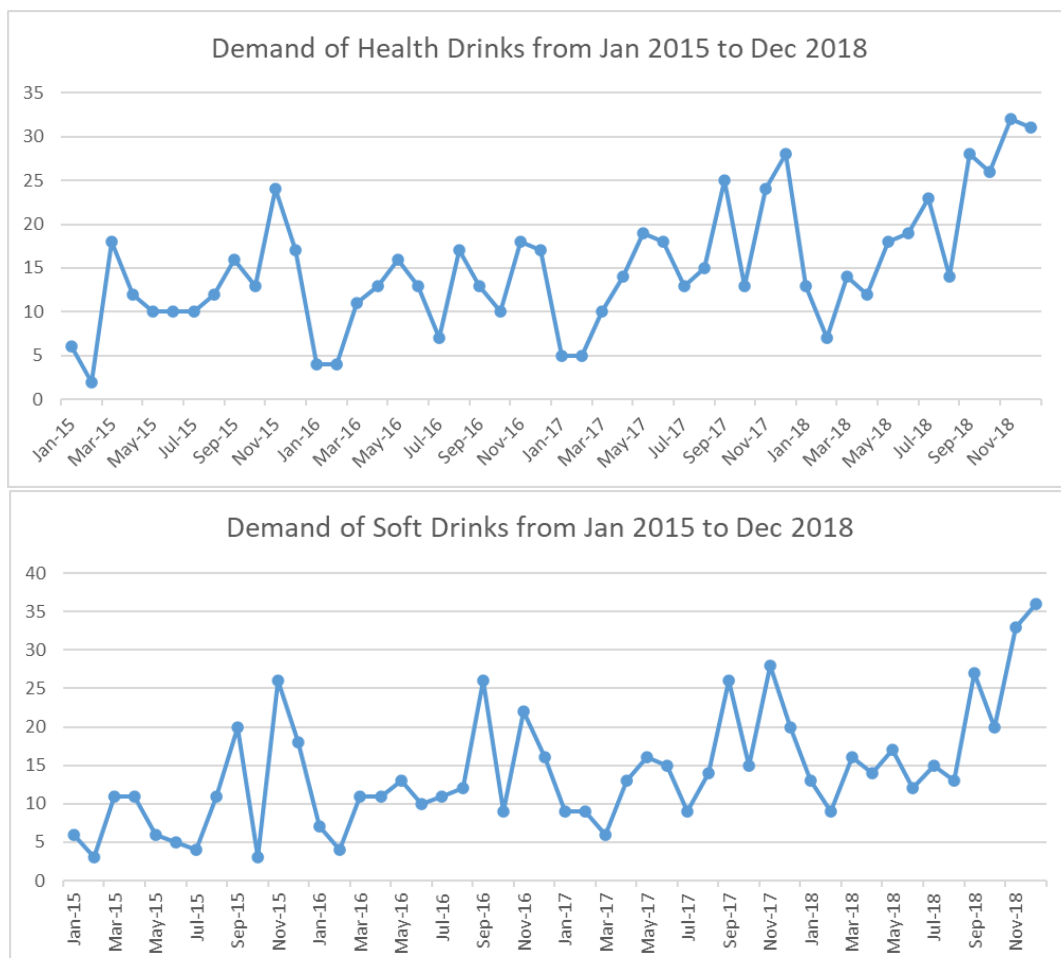
| <b>Health Drinks</b> | <b>Soft Drinks</b> | <b>Cookies</b> | <b>Breads &amp; Buns</b> | <b>Noodles</b> |
|----------------------|--------------------|----------------|--------------------------|----------------|
| Vellore              | Perambalur         | Tenkasi        | Karur                    | Virudhunagar   |
| Chennai              | Madurai            | Dharmapuri     | Cumbum                   | Pudukottai     |
| Ramanadhapuram       | Tirunelveli        | Tirunelveli    | Perambalur               | Perambalur     |
| Tenkasi              | Namakkal           | Ramanadhapuram | Bodi                     | Kanyakumari    |
| Coimbatore           | Theni              | Karur          | Krishnagiri              | Krishnagiri    |

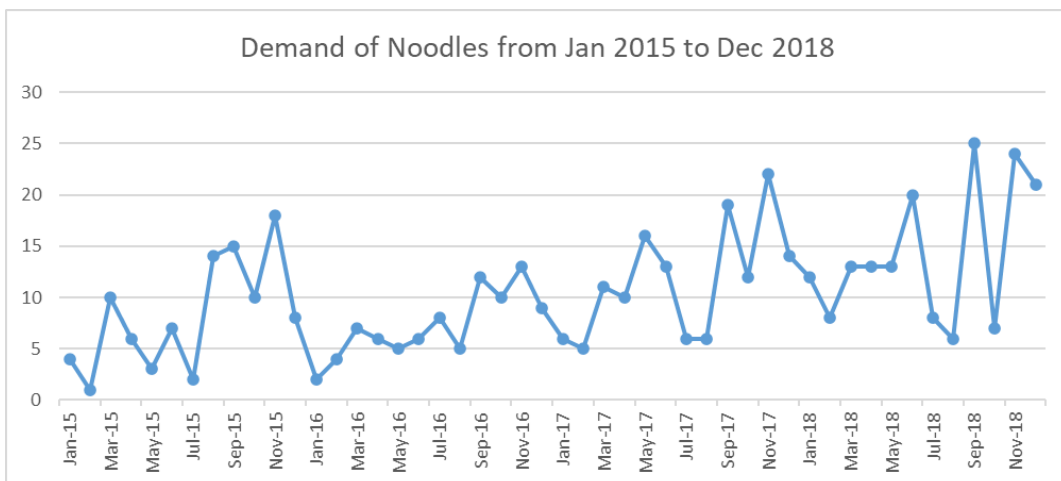
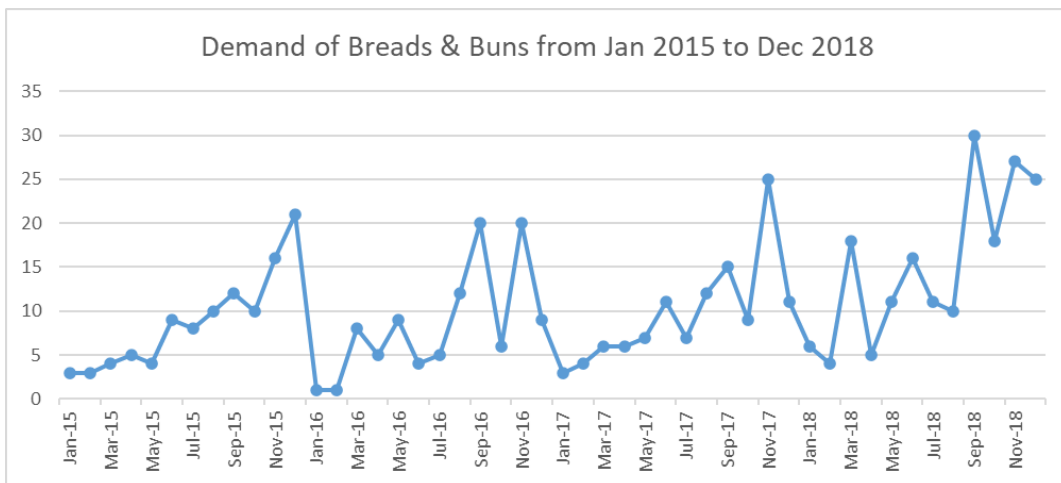
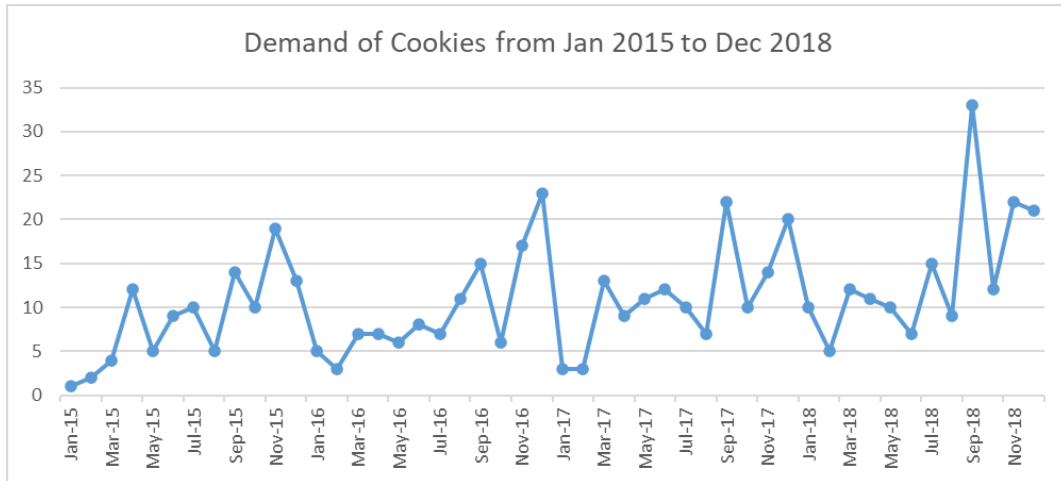
## Objective 4: Investigate demand of each top product over the years

In this activity, the top 5 products which have higher income value are tested for demand over the years. The demand for each month from 2015 to 2018 is visualized. In an inventory, demand of a product in a month can be defined as total number of transactions for that product in a month. For example, if a product is sold 50 times in a month, then the demand value for that month is 50. The demand is irrespective of total sales price, location of outlet, and profit.

**Following is the procedure to calculate demand:**

1. Take a slice from dataset for the target product. For example, Health Drinks
2. From this slice, take a slice for each month of each year. For example, December 2018
3. Calculate the length of output, which is the total transactions in the month for this product.





## **Objective 5: Providing suggestions to improve sales in the store based on past investigations**

Top 5 selling products are : Health Drinks, Soft Drinks, Cookies, Breads & Buns, and Noodles. It can easily be inferred that most of the population is purchasing drinks and fast-food products. The next question arises is whether which population is consuming these products at its highest. The experiment following can be useful to find out the answer to this question.

To produce higher income from Health Drinks, vendors should expand their business or take a special care of demand in Vellore, Chennai, Ramanadhapuram, Tenkasi, and Coimbatore outlets. Similar actions could be taken for other types of products and their cities.

The top 5 selling products are not profiting, but they are being popular among the citizens as years passes. There exists a predictable trend in demand of each product. This information is useful for marketing strategist and outlet owners to anticipate the demand of these products in future. Therefore, benefitting the outlets in terms of profit and maximized sales.