Mini Hackathon 2023

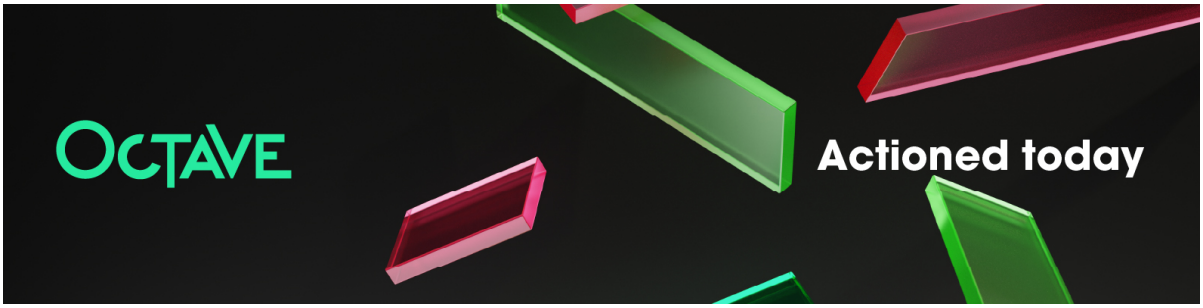
Final Round

OCTAVE &

Department of Statistics

University of Colombo

December 16, 2023



1 Background to the business

Company X is a major confectionery manufacturer operating in Sri Lanka. Their manufacturing facilities are in the Western Province. They manufacture a wide range of confectionery products such as chocolates, biscuits, and candies. The confectionery company does not sell directly to end consumers. They distribute their products through a large network of distributors. These distributors buy large quantities of goods at a price called PTD. The distributors then sell these products to different outlets such as grocery shops, supermarkets, pharmacies, restaurants, etc. This is sold at a price known as the PTR. Consequently the outlets sell these products to end consumers at a price known as MRP.

A group of DSRs are employed by each distributor. They typically visit outlets and promote sale of confectionery products. DSRs are allocated such that they cover the entire island. Outlet visits are done based on a route assigned to each DSR. The route gives directions as to what outlets a DSR should visit within a day

Under each distributor, there are DSRs whose job is to visit each outlet to promote and sell the products of the confectionery company. These sales representatives are allocated to a local region where they should operate within. There are a set of outlets allocated to each sales representative, which they should visit and make sales. These outlets are tagged to routes and each DSR followed the allocated route to make a sale of an SKU to an outlet.

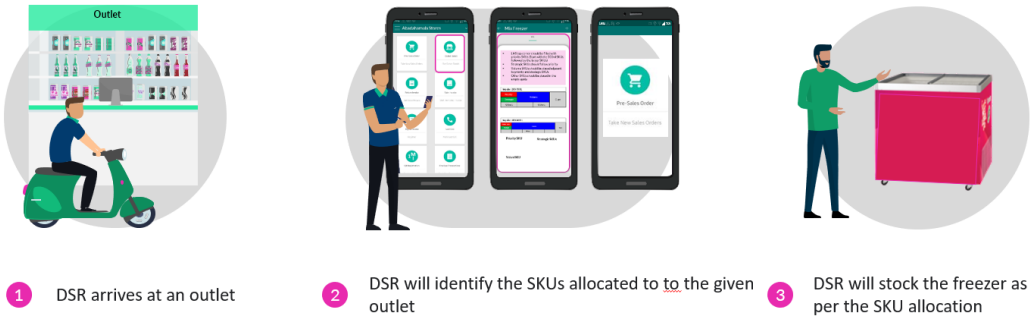


Figure 1: Process of a DSR selling SKUs to outlets

1.1 Definitions of important technical terms

Certain technical terms that will be essential in understanding the business further have been explained below.

- Price to Distributor (PTD): Price at which the products are sold by confectionery manufacturer to the distributor
- Price to Retailer (PTR): Price at which the products are sold by distributors to the outlets
- Maximum Retail Price (MRP): Price at which the products are sold by outlets to the end consumers
- Stock Keeping Unit (SKU): A unique code that identifies a confectionery product. Figure 2 shows two unique SKUs produced by comapany X.

(a) SKU002(check): Choco Heaven Bars

(b) SKU001(check): Gummy Carnival Assortment

Figure 2: Sample SKUs produced by Company X

- Distributor Sales Representative (DSR): An employee of the distributor who visits outlets

to promote and sell products

2 Background to the problem

2.1 Part 1

The company forecasts the monthly total sales of the products that they produce and sets an overall company target. This target is then divided among the distributors. The company has identified that the current distributor target setting approach is inefficient. Therefore, they have approached you to develop an analytical approach for dividing the overall company target among its distributors.

The current approach for setting distributor sales targets is to equally divide the company sales target among all the distributors. As this method doesn't take into account historical sales of distributors, it doesn't provide customized targets for each distributor to match their sales potential.

Historical sales data has been provided for 5 months starting from 2023-01-01 to 2023-05-31. The company has forecasted 547,575 of sales quantity for the month of June 2023. You are required to come up with a logic to provide the distributor wise sale quantity targets for the month of June 2023.

You must submit a file containing the distributor_id, and the distributor wise sales quantity targets for June 2023 in the below format as a .csv file.

Furthermore, clearly explain the logic used to determine the distributor wise sales quantity targets and document the assumptions that you made when calculating these targets and explain your approach to solving this problem.

distributor_id	sales_target
distributor_1	XXX
distributor_2	XXX
...	...
distributor_24	XXX
distributor_25	XXX

2.2 Part 2

The management of the confectionery company believes that there are groups of outlets that have similar sales potential throughout the country. But due to inefficiencies in the distribution process the full potential of the outlets is not being realized. Using the historical sales data and outlet related data that is provided, you are required to identify groups of similar outlets.

You are required to answer the following questions.

1. Prepare a single table (master table) combining all relevant tables at a suitable granularity in order to perform the clustering.

2. Explain the features you chose for the clustering task. How did you determine their relevance to the problem?

3. Which clustering algorithms did you consider for this problem, and why did you choose the final algorithm?

4. Describe the methods you used to determine the optimal number of clusters for your clustering problem. Did you employ any specific techniques or metrics? If so, explain the rationale behind your choice and how it contributed to the final decision on the number of clusters.

5. Provide an interpretation for each outlet cluster and describe interventions that can be done for these outlet clusters in order to increase sales.

3 Dataset and variable description

You are provided with three types of datasets for your analysis. They are as follows.

1. Sales – Contains outlet, SKU wise sales data (Use when answering Part 1 and Part 2)
    - outlet_id: Unique code to identify the outlet.
    - transaction_time: Timestamp of the transaction
    - expected_rainfall: Expected rainfall for the week
    - freezer_status: Specify if the outlet has a freezer or not
    - sales_quantity: Quantity of items sold
2. Distributor Outlet Mapping – Contains the mapping of which outlets belong to which distributor. (Use when answering Part 1)
    - outlet_id: Unique code to identify the outlet.
    - distributor_id: Unique code to identify the distributor.
3. Outlet Info – Contains outlet related information (Use when answering Part 2)
    - outlet_id: Unique code to identify the outlet.
    - region: Outlet's region based on geographic location. They can be categorized into one of the below three types
        - 1: western,
        - 2: outstation
        - 3: upcountry
    - outlet_size: Square area of the outlet in square feet
    - population_density: Population density within 1 km radius of the outlet
    - number_of_skus_sold: Number of unique products sold at the outlet
    - average_household_income: Average household income within 1 km radius of the outlet