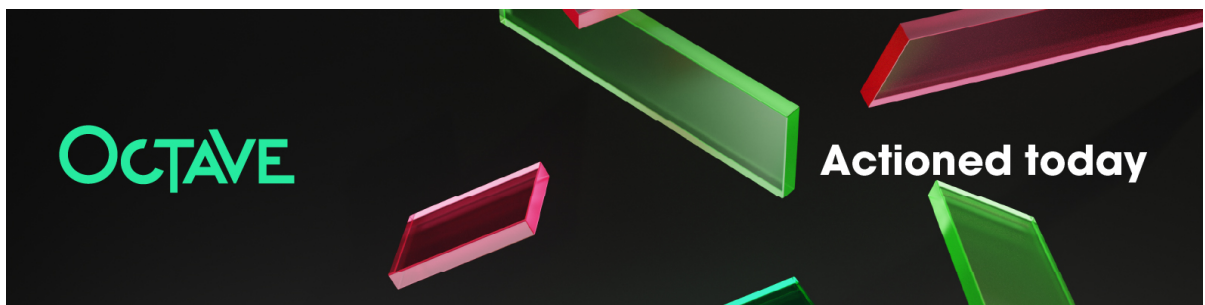


Mini Hackathon 2023

Preliminary Round

OCTAVE &
Department of Statistics
University of Colombo
DECEMBER 1, 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.

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 1 shows two unique SKUs produced by Company X.



(a) Choco Heaven Bars



(b) Gummy Carnival Assortment

Figure 1: 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

The Confectionery Company uses ARIMA (Auto-Regressive Integrated Moving Average) to estimate the sales of each item category across stores, which is used to create a distribution plan of the items. However, recently the company has identified that the methods in use were not efficient enough to realize the full potential of the outlets, and therefore has decided to improve the forecasting through advanced analytics techniques. Their aim is to utilize these

techniques to gain insights of data to **forecast the sales for the next week**. The company considers the standard week for operations to start on Monday.

3 Data set and variable description

You are provided with two types of data sets for your analysis. They are as follows.

1. train.csv – Contains outlet, week wise sales data.
 - outlet_code: Unique code to identify the outlet.
 - week_start_date: Start date of the week in which the transaction took occurred.
 - expected_rainfall: The expected rainfall for the week of the transaction.
 - freezer_status: Specify if the outlet has a freezer or not.
 - outlet_region: The geographical regions of the outlets.
 - sales_quantity: Quantity of items sold.
2. test.csv – You are required to use this data set to test your solution.
 - outlet_code: Unique code to identify the outlet.
 - week_start_date: Start date of the week in which the transaction took occurred.
 - expected_rainfall: The expected rainfall for the week of the transaction.
 - freezer_status: Specify if the outlet has a freezer or not.
 - outlet_region: The geographical regions of the outlets.
 - sales_quantity: Quantity of items sold.

4 Deliverables and Evaluation Metrics

1. **Predictions for next week (60 marks):** You are required to provide the forecasting for the sales for the next week. You must upload a file containing the outlet_code, and the forecasts for next week in the below format.

outlet_code	predicted_sales
4379	XXX
4368	XXX
4456	XXX
4234	XXX
4232	XXX
...	...

You may use the Mean Absolute Percentage Error (MAPE) for evaluation. MAPE is calculated as follows:

$$\text{MAPE} = \frac{\text{Sum}(|\text{Predicted Sales (I)} - \text{Actual Sales(I)}|)}{\text{Sum}(\text{Actual Sales(I)})}$$

where I is for each outlet.

2. **Code Solution (10 marks):** You can use any programming language to code the solution.

- Include comments in your code to explain key steps and decisions.
- Participants can use Python and any relevant libraries for this task.

3. **Report (30 marks):** A comprehensive report that covers your approach and methods used in solving the problem, assumptions made, feature engineering techniques. In addition, provide answers for the following questions.

- What synthetic features are you able to generate from the provided data sets? Compile a table with the name, a brief description, and the data type of the synthetic features. (4 marks)
- Conduct EDA on the datasets. You need to generate the summary statistics for all the synthetic feature columns as well. Include 4 key observations from the EDA. (8 marks)

- (c) What is the selected target variable? How did you engineer it? (4 marks)
- (d) What was your approach for forecasting the sales? Justify your methodology. (4 marks)
- (e) Confectionery Manufacturer Company X has instructed your team that they are looking for data insights and they have specifically asked for the following questions. (10 marks)
- What are the average weekly sales volumes for each outlet_region?
 - How do you assess the impact of rainfall on the weekly sales volumes? Is there any correlation between rainfall and the total weekly sales?
 - Visualize the weekly sales trends across different outlet regions. Are there any noticeable patterns in the sales data?