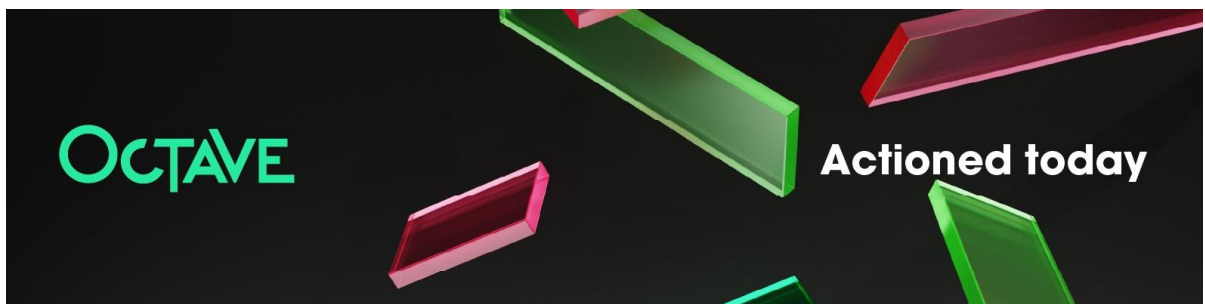


Data Storm 4.0

Semi Final - Case Crack

OCTAVE &
Rotaract Club of University of Moratuwa,
MAY 4th, 2023



1 Utilization of Freezer Assets

1.1 Business Problem

On the previous round of Data Storm 4.0, you provided an analytics solution to Beverages Company XYZ to perform store profiling and enhance their decision-making process. Since you successfully provided an analytical solution during the previous round, Company XYZ now expects your support to enhance their asserts allocation process through advanced analytics.

Company XYZ currently produces both beverages and ice cream items and has expanded its distribution to more than 1000 stores. There are stores that specialize in selling only impulse items or bulk items, and there are also stores that sell both types of products. Company XYZ allocates freezers of different volumes and power consumption rates to each store. The current freezer allocation process is mainly overseen by Area Distributor Managers (ADM). Stores could request a freezer from Company XYZ directly or through their distributors. ADM would visit the specific store to assess:

1. Outlet Size
2. Outlet Space Availability
3. Outlet Sales
4. Outlet Location

Based on these features, the ADM would determine the freezer type. But, given they have expanded their distribution to over 1000 stores, they want to optimally allocate freezers to stores, which would improve Company XYZ sales against the cost invested in freezers. For this they want to,

- 1) Perform a store segmentation to identify the stores with similar characteristics. Assumption here is stores within the same segment are potential candidates to share the same freezer type due to the similar nature & behavior of the stores.
- 2) Recommend a suitable freezer type for each identified outlet segment. Few metrics you can considers are,

$$\text{Return of Investment} = \frac{\text{Ice cream sales in LKR}}{\text{Freezer cost + maintenance \& power consumption cost}}$$
$$\text{Item Sales Ratio} = \frac{\text{Ice cream sales volume}}{\text{Freezer capacity}}$$

Explore the given dataset and you can come up with new metrics/ features to recommend a most suitable freezer type to maximize the ROI and sales of freezers. Your analytical solution can be developed with simple rule base system to machine learning algorithms. Details about the available freezer types are provided in the *freezer_data.csv* file.

1.2 Data Sources

You are provided with the following data source files to develop a data analytics solution.

1. **Sales Data Set** – Historical product sales data set (distributed product volumes for each store in a given week) consists of 81000 records at Outlet-Week-Item level. Please refer to the Data-dictionary.xlsx for more details about the attributions. (*sales_data.csv*)
2. **Outlet Data Set** – Consists of 988 records of stores with shops space (outlet area in square feet). The granularity of the data set is at outlet level. Please refer to the Data-dictionary.xlsx for more details about the attributions. (*outlet_data.csv*)
3. **Product Data Set** – Consists of product wise data such as product name, volume, and price. The granularity of the data set is at product level. Please refer to the Data-dictionary.xlsx for more details about the attributions. (*product_data.csv*)

-
4. **Week Data Set** – Consists of week start date and end date for the given week id. The granularity of the data set is at week level. Please refer to the Data-dictionary.xlsx for more details about the attributions. (*week_data.csv*)
 5. **Freezer Data Set** - Consists of freezer asset wise data such as freezer type, volume, and power consumption, etc. The granularity of the data set is at freezer asset level. Please refer to the Data-dictionary.xlsx for more details about the attributions. (*freezer_data.csv*)

1.3 Deliverables & Evaluation Metrics

In this competition, you are required to submit the following.

1. **Analytical Solution** – In this competition, you are required to create analytical model(s) to:

- (a) Perform a store segmentation.
- (b) Recommend freezer type for each identified segment.

You are required to submit a notebook file (.ipynb) that includes well-commented code with clear steps for performing both tasks. Your results must be displayed, and you must evaluate the performance of your segmentation model against the following metrics:

- **Inertia** - Inertia is the sum of squared distances of samples to their closest cluster center. A good model is one with low inertia AND a low number of clusters (K).
- **Silhouette Coefficient** - Silhouette score for a set of sample data points is used to measure how dense and well-separated the clusters are. The silhouette score of 1 means that the clusters are very dense and nicely separated. The score of 0 means that clusters are overlapping. The score of less than 0 means that data belonging to clusters may be wrong/incorrect.
- **Davies-Bouldin Index** - Davies Bouldin index is based on the principle of within-cluster and between cluster distances. It is commonly used for deciding the number of clusters in which the data points should be labeled. It is different from

the other two as the value of this index should be small. So, the main motive is to decrease the DB index.

- **Calinski-Harabasz Index** - The Calinski Harabasz index is based on the principle of variance ratio. This ratio is calculated between two parameters within-cluster diffusion and between cluster dispersion. The higher the index the better it is clustering.

Furthermore, clearly show the steps to recommend the most suitable freezer type to maximize the ROI and sales of freezers and display the final ROI and sales ratio values for each identified segment – **40 Points**.

2. Technical Report (60 Points) – Report of your solution with clearly defined steps to:

1. Project Description: The clarity and completeness of the project description, including the problem statement and objectives. - **5 Points**.
2. Exploratory Data Analysis and Feature Engineering steps: The insights obtained from the EDA and the effectiveness of the feature engineering steps performed. - **10 Points**.
3. Clustering/ Segmentation Technique: The appropriateness of the clustering technique selected for the project, the effectiveness of the algorithm implementation, and the selection and use of appropriate evaluation metrics to assess the quality of the clustering results. - **10 Points**
4. Identify and differentiate the characteristics of each segment (include any supporting visualizations, summary tables, etc.). – **15 Points**.
5. Analytical approach to allocate freezers to stores – The effectiveness of the analytical approach taken, correctness of the ROI and sales ratio calculations. – **15 Points**.
6. Conclusion and Intervention strategies - The effectiveness of the conclusions drawn from the project, including recommendations for the Company XYZ to maximize their sales against the cost invested in freezers. – **5 Points**.